



Title	META-ANALYSIS METHODS BASED ON SIMULATED INDIVIDUAL PATIENT DATA
Author(s)	Yamaguchi, Yusuke
Citation	大阪大学, 2013, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/27506
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

**META-ANALYSIS METHODS BASED ON
SIMULATED INDIVIDUAL PATIENT DATA**

YUSUKE YAMAGUCHI

MARCH 2013

META-ANALYSIS METHODS BASED ON SIMULATED INDIVIDUAL PATIENT DATA

A dissertation submitted to
THE GRADUATE SCHOOL OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN ENGINEERING

BY
YUSUKE YAMAGUCHI

MARCH 2013

Abstract

In clinical evaluation processes, meta-analysis is a statistical methodology to synthesise results of several trials for the purpose of quantitatively summarising the evidence expressed as a treatment effect (e.g. mean treatment difference for continuous outcomes). Especially in clinical perspectives, there is a growing interest in extending meta-analysis beyond estimating an overall treatment effect, and to produce results tailored to the individual patient or clinically relevant subgroups. Meta-regression (MR) is a technique for modeling relationships between the treatment effect and trial-level covariates, and also can be used for assessing how the patient characteristics affect the clinical effectiveness in the context of meta-analysis. However, most meta-analysis methods build models on aggregate data (AD) obtained by summarising individual patient data (IPD) which should have been measured originally in each trial, and thus a pooled treatment effect is estimated with ignoring scheme of sampling IPD. For this reason, the MR model have often been criticised. The MR model incorporates covariates as summary statistics on background factors of patients, such as a mean age and a proportion of male patients in each trial. This means that the patient characteristics are evaluated expediently with trial-level covariates in place of patient-level covariates. It causes a technical issue that is referred to as ecological bias, and leads to a limitation in interpretation. In particular, it is well known that, for a treatment-covariate interaction between a clinical treatment and a patient characteristic, a test using the MR model has seriously lower statistical power than that using the IPD-based model. Note that, for the inference of the treatment-covariate interaction, the MR model yields a result of just an ‘across-trial interaction effect’ between the treatment effect estimates and mean covariate values, not a ‘within-trial interaction effect’ between individual outcomes and individual covariate values.

As alternative solutions to this problems, some meta-analysis methods using IPD have been suggested. In these methods, once the original IPD including patient-level covariates are collected from all trials involved, any flexible statistical approaches, such as multilevel models and hierarchical random effect models, are applied to the IPD. Meta-analyses based on IPD allow one to achieve much more meaningful evaluation on the treatment-covariate interaction by separating it into the across-trial and the within-trial effect; in particular, the IPD meta-analysis is an only way to assess the within-trial interaction effect. However, use of the IPD may have a disadvantage related to their resources, such as substantial time and costs

to obtain and process the IPD. And also, practitioners cannot always collect the IPD from all trials because the IPD might have been lost or destroyed. For this reason, it has become increasingly important to consider situations where some trials provide IPD (IPD trials) and the others provide only AD (AD trials). Some researchers have already investigated how to combine IPD and AD in meta-analysis, especially when treatment-covariate interaction is of interest.

From these backgrounds, we propose a meta-analysis method for estimating both the across-trial and the within-trial interaction. For the case that all trials provide only AD, we first assume an IPD meta-analysis model including parameters of the across-trial and the within-trial interaction effect, and then marginalise the density of IPD with respect to missing IPD. This process produces a likelihood for the AD available, and allows one to get information on the within-trial interaction by meta-analysing several AD trials. We emphasise that the within-trial interaction can be approximately estimated by using this likelihood even if only the AD are available from each trial. Actually, some simulation studies suggested that the proposed method has potential benefits to the inference of the within-trial interaction in comparison with the existing MR approach. When some trials provide IPD and the others provide only AD, the proposed method is simply extended to combine IPD and AD. There, the likelihood for parameters to be estimated is given by product of a likelihood for the IPD trials and the marginalised likelihood for the AD trials. This again allows one to get information on the within-trial interaction from the AD trials. Through simulation studies, the proposed method provided smaller biases and smaller mean-square errors for estimator of the within-trial interaction in comparison with some existing meta-analysis methods, especially when the proportion of available IPD was small. And also, simulation studies investigated how the proportion of available IPD affects the biases and the mean-square errors for estimator of the within-trial interaction obtained from the proposed method. These results could offer a useful guidance if one considers how many IPD trials should be collected to preserve a desired level of statistical power. Note that the proposed method is applicable when parameters to be estimated can be assumed as fixed effects; so that the treatment effect and the treatment-covariate interaction effects are assumed to be common across trials.

As a breakthrough of the existing meta-analysis methods, we propose a meta-analysis method based on simulated IPD (SIPD), which reconstructs the missing IPD for each trial and then applies a standard IPD meta-analysis model to each SIPD. We here discuss two types of sampling procedures for generating the SIPD: frequentist and Bayesian procedures. Since

the proposed method based on SIPD also uses the scheme of marginalising the missing IPD, any advantages mentioned above are held in this framework. When some trials provide IPD and the others provide only AD, the proposed method reconstructs the missing IPD from the AD trials and then meta-analyses each set of SIPD combined with the collected IPD. Through an illustration with 5 IPD trials in hypertension, which investigate to what extent lowering of systolic blood pressure and diastolic blood pressure contributed to cardiovascular prevention, we demonstrated that the proposed method was much superior to the existing meta-analysis methods in terms of the biases and the mean-square errors for estimator of the within-trial interaction. Using the SIPD enables one to apply any approaches for the IPD meta-analysis, and could have a huge possibility to produce novel findings (e.g. a flexible trial design) which is never provided by the existing meta-analysis methods.

)

Acknowledgements

The author has received much advice, support and friendships of many people from preparation till completion of his dissertation.

The author wishes to thank Honorary Professor Shingo Shirahata of Osaka University, Professor Hidetoshi Shimodaira of Osaka University, Professor Yutaka Kano of Osaka University, Professor Masayuki Uchida of Osaka University and Associate Professor Wataru Sakamoto of Osaka University for providing helpful and useful comments. The author would like to thank Dr. Richard D. Riley of University of Birmingham for giving the author an opportunity to research under his instruction for four month (from June 2012 to October 2012). His helpful comments made this work substantially proceed.

The author would deeply grateful to thank Dr. Masashi Goto of Biostatistical Research Association (BRA), NPO, who led the author to this theme, provided various suggestions, corrections and encouragement for him and gave useful advice and much support.

The author wishes to thank all members of BRA; especially Dr. Satoru Fukinbara, Dr. Norisuke Kawai, Dr. Masaki Fujisawa, Dr. Tomoyuki Sugimoto, Dr. Toshio Shimokawa, Mr. Takao Takase, Dr. Takaharu Yamabe, Dr. Masanori Ito, Kimitoshi Ikeda, Dr. Takashi Nagakubo, Dr. Hiroki Motogaito, Dr. Kazushi Maruo, Mr. Motoki Ohe, Mr. Masatoshi Nakamura, Dr. Naoki Isogawa, Mr. Syusuke Oyama, Mr. Takanori Yoshikawa. The author received many valuable comments from them.

Finally, the author is grateful to his parents for their support and encouragement.

Contents

Abstract	i
Acknowledgements	v
1 Introduction	1
1.1 Backgrounds	1
1.2 Motivating examples	5
1.2.1 Hypertension data	5
1.2.2 Home safety education data	8
1.3 Components of this paper	10
2 Existing methods	13
2.1 Continuous outcome and covariate	13
2.1.1 The case for mixture of IPD and AD	17
2.2 Binary outcome and covariate	19
2.2.1 The case for mixture of IPD and AD	23
3 Proposed methods	27
3.1 Meta-analysis with marginalising the missing IPD	27
3.1.1 The case for only AD	27
3.1.2 The case for mixture of IPD and AD	29
3.2 Meta-analysis based on simulated IPD	30
3.2.1 Continuous outcome and covariate	32
3.2.2 Binary outcome and covariate	45
4 Application and simulation studies	51
4.1 Introduction	51

4.2	Simulation 1: Performance of the proposed method with marginalising the missing IPD in the case for only AD	52
4.2.1	Design of Simulation 1	53
4.2.2	Results of Simulation 1	54
4.3	Simulation 2: Performance of the proposed method with marginalising the missing IPD in the case for mixture of IPD and AD	56
4.3.1	Design of Simulation 2	57
4.3.2	Results of Simulation 2	58
4.4	Application to hypertension data: Illustration of the proposed method based on simulated IPD	61
4.5	Simulation 3: Performance of the proposed method based on simulated IPD in the situation of continuous outcome and covariate	68
4.5.1	Design of Simulation 3	68
4.5.2	Results of Simulation 3	69
4.6	Simulation 4: Performance of the proposed method based on simulated IPD in the situation of binary outcome and covariate	75
4.6.1	Design of Simulation 4	75
4.6.2	Results of Simulation 4	76
5	Discussion and further developments	83
	References	91
	List of publications	99

1 Introduction

1.1 Backgrounds

In clinical evaluation processes, meta-analysis is a statistical methodology to synthesise results of several trials for the purpose of quantitatively summarising the evidence expressed as a treatment effect (e.g. mean treatment difference for continuous outcomes). The fundamental objectives of meta-analysis are to accumulate evidence from smaller trials and to increase statistical power to detect an effectiveness of a clinical treatment (Borenstein *et al.*, 2009). For example, when an investigator is looking for beneficial effects in specific subgroups of patients, a single trial may contain too few patients in the subgroup of interest to be informative. In drug development, meta-analysis is recognised as a useful tool to summarise the overall efficacy results of a drug application and to analyse less frequent outcomes in the overall safety evaluation (Jones, 2008). Sutton and Higgins (2008) reviewed highlights of recent developments in meta-analysis in medical research, and outlined how emphasis has been placed on: heterogeneity and random-effects analyses, special consideration in different areas of application, assessing bias within and across trials, extension of ideas to complex evidence synthesis.

In clinical perspectives, there is a growing interest in extending meta-analysis beyond estimating an overall treatment effect, and to produce results tailored to the individual patient or clinically relevant subgroups (Thompson and Higgins, 2005). Rubin (1990) has criticised conventional meta-analysis techniques just averaging the treatment effects from each trial, and has suggested a need to estimate the effect of treatment versus control as a function of a set of scientific factors that influence efficacy (e.g. age, race and gender). This requires meta-analysis models that assess the association (or interaction) between patient-level covariates and the statistical measure of interest. Meta-regression (MR) is a technique for modelling relationships between the treatment effect and trial-level covariates, and also can be used for assessing how the patient characteristics affect the clinical effectiveness in the context

of meta-analysis. The MR approach has been successfully applied with trial-level variables. Berkey *et al.* (1995) showed that the efficacy of BCG vaccine for tuberculosis increased with distance of the trial site from the equator. Thompson (1993) demonstrated that cholesterol-lowering drugs were more effective in reducing ischemic heart disease in trials in which the treatment groups achieved greater average reductions in serum cholesterol levels relative to their respective control groups. However, most meta-analysis methods build models on aggregate data (AD) obtained by summarising individual patient data (IPD) which should have been measured originally in each trial, and thus a pooled treatment effect is estimated with ignoring scheme of sampling IPD. For this reason, the MR model have often been criticised (Thompson and Higgins, 2002; Riley *et al.*, 2010). The MR model incorporates covariates as summary statistics on background factors of patients, such as a mean age and a proportion of male patients in each trial. This means that the patient characteristics are evaluated expediently with trial-level covariates in place of patient-level covariates. It causes a technical issue that is referred to as ecological bias (Morgenstern, 1982), and leads to a limitation in interpretation (Thompson and Higgins, 2002). In particular, it is well known that, for a treatment-covariate interaction between a clinical treatment and a patient characteristic, a test using the MR model has seriously lower statistical power than that using the IPD-based model (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 5 trials in their clinical research, and showed that the meta-analysis based on the trial-level data failed to detect the treatment-covariate interaction. Thompson and Higgins (2002) advocated that the relationship described by the MR model is an observational association, so this suffers from the bias by confounding. Note that the MR model assumes the ‘across-trial interaction’ between the treatment effect estimates and mean covariate values reflects the more pertinent ‘within-trial interaction’ between individual outcomes and individual covariate values. This may not be true in practice, as across-trial associations are prone to trial-level confounding, and may truly not reflect within-trial associations (Riley and Steyerberg, 2010).

As alternative solutions to this problems, some meta-analysis methods using IPD have been suggested (Riley, Lambert and Abo-Zaid, 2010; Simmonds *et al.*, 2005). In these methods, once the original IPD including patient-level outcome and covariate values are collected from all trials involved, any flexible statistical approaches, such as multilevel models (Goldstein *et al.*, 2000) and hierarchical random effect models (Turner *et al.*, 2000; Whitehead *et al.*, 2001; Higgins *et al.*, 2001; Riley *et al.*, 2007), are applied to the IPD. This brings

many opportunities over the AD approaches in the sense of deriving desired summary results directly, checking modelling assumptions, and assessing non-linear trends (Riley, 2010). Meta-analyses based on IPD allow one to achieve much more meaningful evaluation on the treatment-covariate interaction by separating it into the across-trial and the within-trial effect; in particular, the IPD meta-analysis is an only way to assess the within-trial interaction effect. However, the use of full IPD is not always without its difficulties. In particular, this approach is resource intensive, because substantial time and costs are required to contact trial authors, to obtain their IPD, to input and clean the provided IPD, to resolve any data issues through dialog with the data providers, and to generate a consistent data format across trials (Riley *et al.*, 2010). And also, practitioners cannot always collect the IPD from all trials because the IPD might have been lost or damaged, or trial authors may not be contactable or willing to collaborate. Riley *et al.* (2010) pointed out that the possibility of collecting the IPD from all trials is not so high. If the collectability of IPD is associated with the results in each trial, a meta-analysis based only on the collected IPD may be biased (Stewart and Tierney, 2002). For this reason, 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 *et al.*, 2008; Sutton, Kendrick and Coupland, 2008; Riley and Steyerberg, 2010). Such approaches have also been developed in the context of ecological study (Jackson, Best and Richardson, 2008; Haneuse and Wakefield, 2007; 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. Jackson, Best and Richardson (2008) suggested Bayesian hierarchical related regression which uses Markov chain Monte Carlo method to simultaneously estimate IPD trials and AD trials models linked by common parameters, where the IPD supplement the aggregate information across different groups such as geographical areas.

From these backgrounds, we propose a meta-analysis method for estimating both the across-trial and the within-trial interaction effect. For the case that all trials provide only AD, we first assume an IPD meta-analysis model including parameters of the across-trial and the within-trial interaction effect, and then marginalise the density of IPD with respect to missing IPD, which requires an integration over a region restricted by observed AD. This process produces a likelihood for the AD available, and allows one to get information on

the within-trial interaction by meta-analysing several AD trials. The idea of marginalising the IPD meta-analysis model is inspired from ecological inference, in which the relationships between individual specific quantities are evaluated by using population-level data. In particular, Wakefield and Salway (2001) presented a statistical framework for ecological inference, describing parametric models for binary response data that include within-aggregation variability of covariates, which is intended to reduce the ecological bias. We emphasise that the within-trial interaction can be approximately estimated by using this likelihood even if only the AD are available from each trial. When some trials provide IPD and the others provide only AD, the proposed method is simply extended to combine IPD and AD. There, the likelihood for parameters to be estimated is given by product of a likelihood for the IPD trials and the marginal likelihood for the AD trials. This again allows one to get information on the within-trial interaction from the AD trials. Note that the proposed method is applicable when parameters to be estimated can be assumed as fixed effects; so that the treatment effect and the treatment-covariate interaction effects are assumed to be common across trials.

As a breakthrough of the existing meta-analyses, we propose a meta-analysis method based on simulated IPD (SIPD), which reconstructs the missing IPD for each trial and then applies a standard IPD meta-analysis model to each SIPD. When some trials provide IPD and the others provide only AD, the proposed method reconstructs the missing IPD for the AD trials and then meta-analyse each set of SIPD combined with the collected IPD. We show, once the SIPD are generated, how existing IPD meta-analysis approaches can be applied, and we demonstrate the benefits of incorporating the SIPD. We here consider two types of sampling procedures for generating the SIPD: frequentist and Bayesian procedures, which are inspired by multiple imputation applied in the analysis of incomplete data with missing outcomes and covariates (Rubin, 1987). In the frequentist procedure, each set of SIPD is generated from a conditional distribution of the missing IPD given the AD (and the collected IPD) and a known parameter, and then resulting estimates from each SIPD (combined with the collected IPD) are summarised by using Poor Man's Data Augmentation 2 proposed by Wei and Tanner (1990). In the Bayesian procedure, each set of SIPD is generated from a posterior predictive distribution of the missing IPD given the AD (and the collected IPD), and then resulting estimates from each SIPD (combined with the collected IPD) are summarised by using Rubin's (1987) rule. Both procedures ultimately produce a posterior distribution of parameters of interest, and thus a posterior mean and variance for frequentist inference. Since these approaches also use the scheme of marginalising the missing IPD, any advantages

Table I. Summary of the 5 trials in hypertension, included in the meta-analysis of Wang *et al.* (2005).

Trial name*	Number of patients		Age (years)		SBP (follow-up – baseline)		Estimate of mean difference (s.e.)
	Control	Treatment	Control mean (s.d.)	Treatment mean (s.d.)	Control mean (s.d.)	Treatment mean (s.d.)	
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.

provided above are held in these frameworks. Furthermore, using the SIPD enables one to apply any approaches for the IPD meta-analysis, and could have a huge possibility to produce novel findings (e.g. a flexible trial design) which is never provided by the existing meta-analysis methods. Note that the proposed method is again applicable when parameters to be estimated can be assumed as fixed effects.

1.2 Motivating examples

1.2.1 Hypertension data

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 (Liu *et al.*, 1998; Staessen *et al.*, 1997; Amery *et al.*, 1985). 10 trials were ultimately included, and these provided IPD for a total of 28,592 patients. To illustrate meta-analysis methods introduced in Chapter 2 and Chapter 3, we will carry out a meta-analysis of 5 (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 5 trials were chosen as they were conducted in populations with a similar mean age around 70. The mean change in SBP (follow-up minus baseline) for each treatment group in each trial are shown in Table I, with negative values indicating a beneficial effect. The treatment effect is shown in the rightmost column in Table I, with negative values indicating that the treatment is effective. Table I also shows the mean age, and the groups appear to be well balanced in each trial at baseline.

One of the usual way for displaying meta-analysis data is known as forest plot. Figure 1 shows the forest plot of the 5 trials in hypertension. The position of the black squares represents the findings (an estimate of the mean outcome difference between groups, and its standard error) from each individual trial. The size of the square is proportional to the precision of the trial (roughly speaking, the sample size). A horizontal line drawn on both sides of the squares for each trial denotes the 95 per cent confidence interval of the treatment effect estimate. A pooled treatment effect estimate obtained by combining all 5 trials is displayed as a diamond in the lowest part of the forest plot. We here assume that all 5 trials share a common true value of the mean difference (i.e. the fixed treatment effect), and we estimate the pooled treatment effect. Now let MD_i , $\widehat{MD}_i$ and $V(\widehat{MD}_i)$ be a true value of the mean difference from the i th trial ($i = 1, \dots, 5$), an estimate of MD_i and a variance estimate of $\widehat{MD}_i$, respectively. Assume $\widehat{MD}_i$ is normally distributed with mean MD_i and known variance $V(\widehat{MD}_i)$, and

$$MD = MD_1 = \dots = MD_5.$$

Then, the pooled mean difference is estimated by

$$\widehat{MD} = \frac{\sum_{i=1}^5 \frac{\widehat{MD}_i}{V(\widehat{MD}_i)}}{\sum_{i=1}^5 \frac{1}{V(\widehat{MD}_i)}} = -10.77$$

and the 95 per cent confidence interval for $\widehat{MD}$ is given by

$$\widehat{MD} \pm 1.96 \times \left(\sum_{i=1}^5 \frac{1}{V(\widehat{MD}_i)} \right)^{-1/2} = [-11.40, -10.14].$$

These results (seen in Figure 1) indicate that the treatment is significantly effective in reducing SBP by, on average, 10.77 mmHg more than placebo.

We also examine the extent of heterogeneity in the treatment effect across the 5 trials. It is generally accepted that meta-analyses should assess heterogeneity, which may be defined

as the presence of variation in true effect sizes underlying the different studies (Higgins, 2008). Cochran's Q test is often applied in meta-analysis for determining whether there is heterogeneity in treatment effects (Cochran, 1954). For the 5 trials in hypertension, the Q test evaluates a null hypothesis $H_0 : MD = MD_1 = \dots = MD_5$ by using the following Q statistic which has a chi-square distribution with (the number of trials minus 1) degrees of freedom:

$$Q = \sum_{i=1}^5 \frac{(\widehat{MD}_i - \widehat{MD})^2}{V(\widehat{MD}_i)} = 4.849 \sim \chi_4^2.$$

which provides a p-value of 0.303, and thus there is no strong evidence of heterogeneity. Furthermore, I^2 index is the proportion of total variation in the estimates of treatment effect that is due to heterogeneity across trials, and can easily be interpreted as a percentage of heterogeneity (Higgins and Thompson, 2002). For the 5 trials in hypertension, I^2 is given by

$$I^2 = \frac{Q - 4}{Q} \times 100 = 17.50 \text{ per cent.} \quad (1.1)$$

Here, according to a tentative classification of I^2 values with the purpose of helping to interpret its magnitude by Higgins and Thompson (2002), the percentages of around 25 per cent ($I^2 = 25$), 50 per cent ($I^2 = 50$), and 75 per cent ($I^2 = 75$) would mean low, medium,

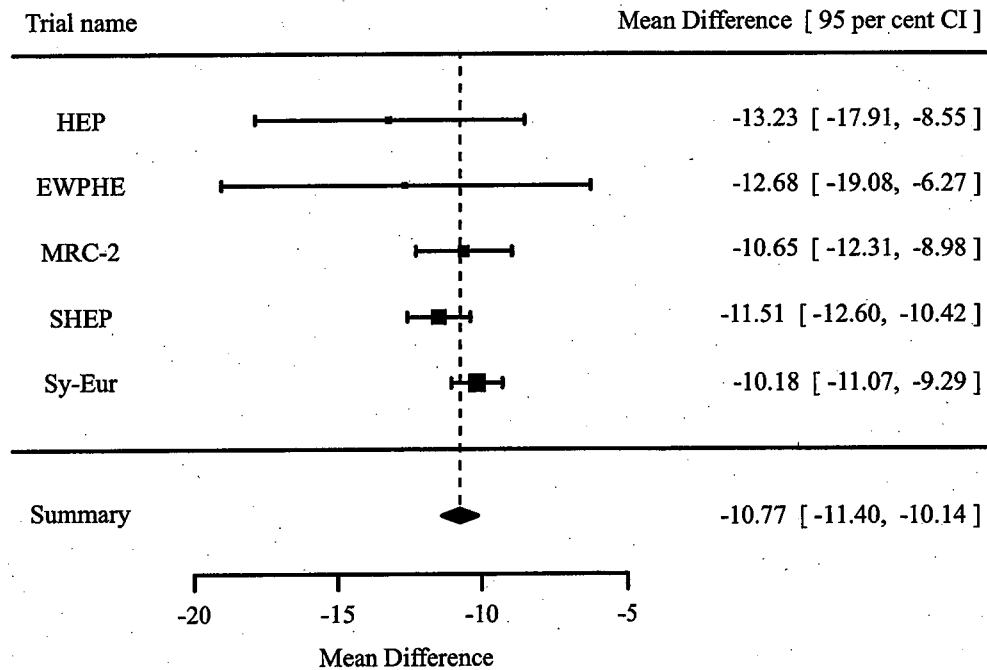


Figure 1. Forest plot of 5 trials in hypertension.

Table II. Summary of the 5 trials in home safety education, included in the meta-analysis of Sutton, Kendrick and Coupland (2008).

No. of trial*	Intervention group	Control group	Log odds ratio (s.e.)	Proportion of male participants
	Number of participants with a stair gate / total	Number of participants with a stair gate / total		
Trial 1	28 / 49	25 / 47	0.16 (0.41)	0.51
Trial 2	76 / 85	70 / 89	0.83 (0.44)	0.48
Trial 3	25 / 57	16 / 55	0.64 (0.40)	0.48
Trial 4	408 / 742	328 / 718	0.37 (0.11)	0.52
Trial 5	250 / 363	239 / 360	0.11 (0.16)	0.52

s.e.: standard error.

*Each trial is arbitrarily numbered, differently from Sutton, Kendrick and Coupland (2008)

and high heterogeneity, respectively. Thus, the I^2 value (1.1) again indicates potential low heterogeneity across the 5 trials. Therefore, we concluded that any models with fixed treatment effects, which are introduced in Chapter 2 and Chapter 3, could be appropriate for the 5 trials in hypertension.

It is also clinically important to assess how the age adjusts the treatment effect. The hypertension data will be used in this paper to demonstrate and critically assess the methods developed; those interested in more clinical conclusions are referred elsewhere (Wang *et al.*, 2005).

1.2.2 Home safety education data

Sutton, Kendrick and Coupland (2008) performed a systematic review to investigate the effectiveness of home safety education on the provision of a safety equipment. In particular, they meta-analysed 8 trials (Clamp and Kendrick, 1998; Nansel *et al.*, 2002; McDonald *et al.*, 2005; Watson *et al.*, 2005; Kendrick *et al.*, 2005; Kendrick *et al.*, 1999; Sznajder *et al.*, 2003; Gielen *et al.*, 2002), which have inspected whether an educational intervention increases the ownership of stair gates installed for the prevention of falls in children. As a participant-level socioeconomic characteristic affecting the intervention effectiveness, they were interested in whether the family is a single or two-parent household; so that they assessed how the number of families with a fitted stair gate in the intervention group is different between these two participant subgroups. The review involved 6 trials (3,447 participants) with IPD and 2 other

trials (193 participants) with AD. And also, participants in 3 trials were allocated to some clusters nested within the trials. To illustrate meta-analysis methods introduced in Chapter 2 and Chapter 3, we here carry out a meta-analysis for 5 (2,565 participants) of 8 trials, which are sufficiently homogeneous across trials with respect to a treatment effect and a trial-level covariate. Although participants in one trial are allocated to 37 clusters, we ignore this participant-clustering to avoid further complexity. We focus on gender of children as a participant-level covariate, and mainly assess how gender affects the intervention effectiveness. The number of participants in each group, and the number of participants with a stair gate in each group are shown in Table II. The intervention effects (log odds ratio between two treatment groups) from each trial are shown in Table II, with positive values indicating a beneficial effect. Table II also shows the proportion of male participants in each trial. There is small variation in the proportion of male participants across trials.

Figure 2 shows the forest plot of the 5 trials in home safety education. The position of the black squares represents the findings of the log odds ratios from each individual trial. Now let $\log OR_i$, $\log \widehat{OR}_i$ and $V(\log \widehat{OR}_i)$ be a true value of the log odds ratio from the i th trial ($i = 1, \dots, 5$), an estimate of $\log OR_i$ observed and a variance estimate of $\log \widehat{OR}_i$, respectively. Assume that $\log \widehat{OR}_i$ is normally distributed with mean $\log OR_i$ and known

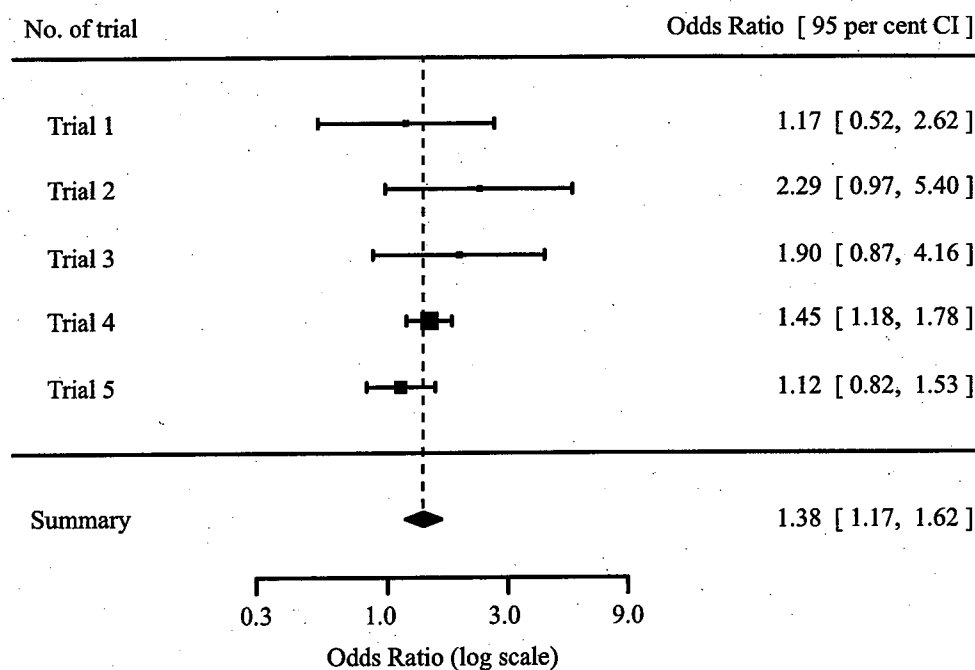


Figure 2. Forest plot of the 5 trials in home safety education.

variance $V(\log \widehat{OR}_i)$, and

$$OR = OR_1 = \dots = OR_5.$$

Then, the pooled odds ratio is estimated by

$$\widehat{OR} = \exp \left\{ \frac{\sum_{i=1}^5 \frac{\log \widehat{OR}_i}{V(\log \widehat{OR}_i)}}{\sum_{i=1}^5 \frac{1}{V(\log \widehat{OR}_i)}} \right\} = 1.38$$

and the 95 per cent confidence interval for $\widehat{OR}$ is given by

$$\exp \left\{ \log \widehat{OR} \pm 1.96 \times \left(\sum_{i=1}^5 \frac{1}{V(\log \widehat{OR}_i)} \right)^{-1/2} \right\} = [1.17, 1.62].$$

These results (seen in Figure 2) indicates that the the intervention significantly increases the probability of ownership of the stair gate more than control on average.

We also examine the extent of heterogeneity in the treatment effect across the 5 trials. The Q test gave a Q statistic

$$Q = \sum_{i=1}^5 \frac{(\log \widehat{OR}_i - \log \widehat{OR})^2}{V(\log \widehat{OR}_i)} = 4.067 \sim \chi_4^2.$$

and its p-value of 0.397, and thus there is no strong evidence of heterogeneity. Further, I^2 index computed in the same way for (1.1) was 1.65 per cent. This again indicates potential low heterogeneity across the 5 trials. Therefore, we concluded that fixed treatment effect models introduced in Chapter 2 and Chapter 3 could be appropriate for the 5 trials in home safety education.

It is also clinically important to assess how gender of children adjusts the treatment effect. The home safety education data will be used in this paper to demonstrate and critically assess the methods developed; those interested in more clinical conclusions are referred elsewhere (Sutton, Kendrick and Coupland, 2008).

1.3 Components of this paper

In Chapter 2, we introduce IPD and AD in two situations where: (i) a single continuous outcome and a single continuous covariate are observed from each patient, (ii) a single binary outcome and a single binary covariate are observed from each patient, and describe IPD

meta-analysis models and the MR models for each situation, respectively. We also discuss the differences between the within-trial and the across-trial interaction. Through applications to the hypertension data and the home safety education data, the methods are illustrated. Furthermore, for the case that some trials provide IPD and the others provide only AD, we describe existing models for combining IPD and AD. In Chapter 3, we describe our new meta-analysis method with marginalising the missing IPD for the situation (i). The method is extended to combine IPD and AD. We also describe meta-analysis methods based on simulated IPD. The methods are explained by frequentist and Bayesian perspectives, and applied to both situations (i) and (ii). In Chapter 4, we conduct simulation studies to examine the performance of the proposed method with marginalising the missing IPD in comparison to existing methods. Furthermore, another simulation study and an application to the hypertension data are conducted to assess the benefit of using simulated IPD. Finally, in Chapter 5, we conclude this paper with some discussion.

2 Existing methods

Consider a meta-analysis of N trials in which patients are assigned to either a treatment group (T) or a control group (C). Let n_i be the number of patients in the i th trial ($i = 1, \dots, N$), n_{iT} and n_{iC} be the numbers of patients for the treatment and the control group, respectively. Here, let y_{ij} and z_{ij} be a patient-level outcome and covariate observed from the j th patient ($j = 1, \dots, n_i$) in the i th trial, and let x_{ij} be coded 0/1 to denote control/treatment group. We here describe some existing meta-analysis methods for two data situations (continuous outcome and covariate, and binary outcome and covariate). We also consider the case for mixture IPD and AD.

2.1 Continuous outcome and covariate

Let y_{ij} and z_{ij} denote a continuous outcome and a continuous covariate value from the j th patient in the i th trial. If just meta-analysing the IPD; i.e. (y_{ij}, x_{ij}, z_{ij}) for $i = 1, \dots, N$ and $j = 1, \dots, n_i$, Riley *et al.* (2008) proposed the following one-stage model that accounts for the clustering of patients within trials by a trial-specific intercept (ϕ_i), and estimates a pooled treatment-covariate interaction (γ_W) based on within-trial information separated from the across-trial interaction (γ_A):

$$\begin{aligned} y_{ij} &= \phi_i + \theta x_{ij} + \mu z_{ij} + \gamma_A x_{ij} \bar{z}_i + \gamma_W x_{ij} (z_{ij} - \bar{z}_i) + \epsilon_{ij}, \\ \epsilon_{ij} &\sim N(0, \sigma_y^2), \\ j &= 1, \dots, n_i; \quad i = 1, \dots, N. \end{aligned} \tag{2.1}$$

Here, ϕ_i is the fixed intercept for the i th 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 outcome for a one-unit increase in z_{ij} , γ_A and γ_W are the across-trial and the within-trial effect of treatment-covariate interaction, respectively, and $\bar{z}_i = \sum_{j=1}^{n_i} z_{ij}/n_i$ denotes a mean covariate in the i th trial. Note that, θ , μ , γ_A and γ_W

are treated as fixed effects here, and σ_y^2 is assumed to be common across trials. In contrast to Riley *et al.* (2008), model (2.1) includes a common μ and σ_y^2 across trials rather than a trial-specific μ_i and σ_{yi}^2 for the i th trial. These assumptions are necessary to build the proposed method (see Chapter 3). According to a recommendation by Riley *et al.* (2008), the treatment-covariate interaction is separated into the across-trial and the within-trial effect. This separation is clinically important to avoid making a wrong conclusion about the treatment-covariate interaction, which might occur if wrongly amalgamating the across-trial and the within-trial effects (Riley and Steyerberg, 2010). One should pay attention to the fact that the across-trial relationships can be very different from the within-trial relationships due to ecological bias and/or trial-level confounding (Riley, Lambert and Abo-Zaid, 2010). This modelling framework is also discussed in the context of regression analysis of clustered data (Neuhaus and Kalbfleisch, 1998; Begg and Parides, 2003).

In general meta-analytic situations, we observe only the AD from each trial, rather than the IPD from each patient. Here, the AD consist of sample means and sample variances for individual observations in each group and trial; i.e. $(\bar{y}_{iT}, s_{yiT}^2, \bar{z}_{iT}, s_{ziT}^2, \bar{y}_{iC}, s_{yiC}^2, \bar{z}_{iC}, s_{ziC}^2)$ for $i = 1, \dots, N$, where

$$\begin{aligned}\bar{y}_{iT} &= \frac{\sum_{j \in T} y_{ij}}{n_{iT}}, \quad s_{yiT}^2 = \frac{\sum_{j \in T} (y_{ij} - \bar{y}_{iT})^2}{n_{iT} - 1}, \quad \bar{z}_{iT} = \frac{\sum_{j \in T} z_{ij}}{n_{iT}}, \quad s_{ziT}^2 = \frac{\sum_{j \in T} (z_{ij} - \bar{z}_{iT})^2}{n_{iT} - 1}, \\ \bar{y}_{iC} &= \frac{\sum_{j \in C} y_{ij}}{n_{iC}}, \quad s_{yiC}^2 = \frac{\sum_{j \in C} (y_{ij} - \bar{y}_{iC})^2}{n_{iC} - 1}, \quad \bar{z}_{iC} = \frac{\sum_{j \in C} z_{ij}}{n_{iC}}, \quad s_{ziC}^2 = \frac{\sum_{j \in C} (z_{ij} - \bar{z}_{iC})^2}{n_{iC} - 1}.\end{aligned}$$

Then, an MR model can be applied to the AD as follows:

$$d_i = \alpha + \beta \bar{z}_i + \epsilon_i, \quad (2.2)$$

$$\epsilon_i \sim N(0, \sigma_{di}^2),$$

$$i = 1, \dots, N.$$

Here, $d_i = \bar{y}_{iT} - \bar{y}_{iC}$ denotes a mean outcome difference between groups from the i th trial, and the error variance is assumed to be known as

$$\sigma_{di}^2 = V(d_i) = \frac{n_{iT} + n_{iC}}{n_{iT}n_{iC}} s_{yi}^2$$

where

$$s_{yi}^2 = \frac{(n_{iT} - 1)s_{yiT}^2 + (n_{iC} - 1)s_{yiC}^2}{n_{iT} + n_{iC} - 2}.$$

If we assume $z_{ij} = \bar{z}_i$ in model (2.1) with trial-specific error variances, we have

$$y_{ij} = \phi_i + \theta x_{ij} + \mu \bar{z}_i + \gamma_A x_{ij} \bar{z}_i + \epsilon_{ij}, \quad (2.3)$$

$$\epsilon_{ij} \sim N(0, \sigma_{yi}^2),$$

$$j = 1, \dots, n_i; i = 1, \dots, N.$$

When first taking an average of both sides in model (2.3) for each group and trial, and subtracting the average of the control group from that of the treatment group for each trial, a model for the mean outcome differences between groups can be derived as follows:

$$\bar{y}_{iT} - \bar{y}_{iC} = \theta + \gamma_A \bar{z}_i + \bar{\epsilon}_i,$$

$$\bar{\epsilon}_i \sim N\left(0, \frac{n_{iT} + n_{iC}}{n_{iT}n_{iC}} \sigma_{yi}^2\right)$$

$$i = 1, \dots, N$$

where σ_{yi}^2 is estimated by $\hat{\sigma}_{yi}^2 = s_{yi}^2$. This model has the same form as the representation of the MR model (2.2); so that, under an assumption of $z_{ij} = \bar{z}_i$, θ and γ_A in model (2.1) are equivalent to α and β in the MR model (2.2), respectively. This means that the MR model (2.2) gives a restrictive result about the across-trial relationships under a condition that variation of the covariate is equal to zero. However, we have more interest in the parameter of γ_W , which represents an increase in the treatment effect according to one-unit increase in the patient-level covariate z_{ij} . Thus, if we intend to estimate γ_W by using the estimate of β , this might lead to an incorrect conclusion for the treatment-covariate interaction. When the number of trials and variation of $\bar{z}_i$ across trials are small, due to the ecological bias, the statistical power of β becomes much lower than those of γ_W (Simmons and Higgins, 2007).

Application to hypertension data

Consider the hypertension data, and we now demonstrate how age modify the treatment effect on change in SBP (follow-up minus baseline). Fitting model (2.1) to the IPD from the 5 trials, estimates of each parameter in model (2.1) were $\hat{\theta} = 35.95$ (s.e. = 32.83 and p-value = 0.273), $\hat{\mu} = 0.035$ (s.e. = 0.039 and p-value = 0.370), $\hat{\gamma}_A = -0.662$ (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. Fitting the MR model (2.2) to the AD from the 5 trials, estimates of each parameter in model (2.2) were $\hat{\alpha} = 43.13$ (s.e. = 32.94 and p-value = 0.188) and $\hat{\beta} = -0.766$ (s.e. = 0.466 and p-value = 0.100), respectively. The across-trial relationships ($\hat{\theta}$ and $\hat{\gamma}_A$) obtained by fitting model

(2.1) were similar to those ($\hat{\alpha}$ and $\hat{\beta}$) obtained by fitting the MR model (2.2). A slight difference between $\hat{\theta}$ and $\hat{\alpha}$ (or $\hat{\gamma}_A$ and $\hat{\beta}$) is due to model assumptions; so that the MR model (2.2) assumes known error variances ($\sigma_{\hat{\alpha}i}^2$) for each trial while model (2.1) assumes a common error variance (σ_y^2) across trials, and $\hat{\theta}$ and $\hat{\gamma}_A$ are affected by $\hat{\mu}$ and $\hat{\gamma}_W$. The across-trial relationships had much larger standard errors in comparison with those of the within-trial relationships ($\hat{\mu}$ and $\hat{\gamma}_W$). As for the treatment-covariate interaction, the across-trial effect was substantially different from the within-trial effect on the point estimates. This shows the importance of separating the within-trial interaction from the across-trial interaction,

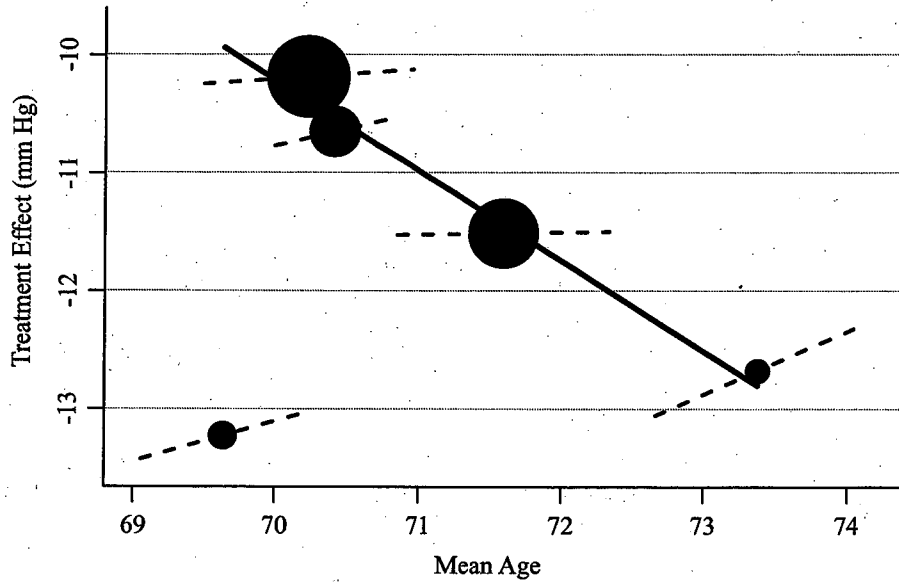


Figure 3. Scatter plot for the 5 trials in hypertension with across-trial and within-trial interaction effect estimates, in which:

- A solid line represents the across-trial interaction ($\hat{\gamma}_A$) between mean age ($\bar{z}_i$) and treatment effect estimated by fitting model (2.2).
- Dashed lines represents the within-trial interaction ($\hat{\gamma}_W$) between age and treatment effect estimated separately within each trial using IPD and model (2.1) without γ_A .
- The gradient of each dashed line indicates the change in treatment effect for a one 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 centered at $\bar{z}_i$ in each trial; the circle size is proportional to the sample size in each trial.

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. 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 5 trials. There was no observed between-trial heterogeneity in the within-trial interaction ($I^2 = 0\%$), and thus the fixed effect assumption is also plausible for this parameter. Figure 3 also shows this difference between $\hat{\gamma}_A$ and $\hat{\gamma}_W$; the within-trial interaction (dashed lines) have almost flat gradients, especially in the larger trials, while the across-trial interaction (solid line) has a steep negative gradient. It highlights pitfall of using $\hat{\gamma}_A$ to make inferences about $\hat{\gamma}_W$, that is, ecological bias and confounding.

2.1.1 The case for mixture of IPD and AD

Consider a meta-analysis of N' trials which consist of N trials providing AD (AD trials) and $N' - N$ trials providing IPD (IPD trials). The IPD trials provide the patient-specific observations; i.e. (y_{ij}, x_{ij}, z_{ij}) for $i = N + 1, \dots, N'$ and $j = 1, \dots, n_i$. When a mixture of IPD and AD trials are available, model (2.1) must be modified to combined IPD and AD. The simplest solution is to reduce the collected IPD to AD and treat all the data as AD, 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 combining IPD and AD, which simultaneously estimates the within-trial relationships (using just the IPD trials) and the across-trial relationships (using both IPD and AD trials). All these approaches are now described.

Meta-regression model that uses only AD from all trials

Once the IPD for trials $i = N + 1, \dots, N'$ are summarised to the AD, the MR model (2.2) can be applied to the AD for all trials $i = 1, \dots, N'$.

Model that uses only IPD trials available

If one uses only the collected IPD, model (2.1) can be applied to the IPD from trials $i = N + 1, \dots, N'$. When the number of IPD trials is one (i.e. $N' = N + 1$), model (2.1) is

modified as follows:

$$\begin{aligned} y_{ij} &= \phi_i + \theta x_{ij} + \mu z_{ij} + \gamma_W x_{ij}(z_{ij} - \bar{z}_i) + \epsilon_{ij}, \\ \epsilon_{ij} &\sim N(0, \sigma_y^2), \\ j &= 1, \dots, n_i; i = N. \end{aligned} \quad (2.4)$$

This is because one cannot estimate the across-trial interaction γ_A with a single trial.

Model that uses both IPD and AD trials

The model for combining IPD and AD proposed by Riley *et al.* (2008) is as follows:

$$\begin{aligned} y_{ij}^* &= D_i \phi_i + \theta x_{ij} + D_i \mu z_{ij}^* + \gamma_A x_{ij} \bar{z}_i + D_i \gamma_W x_{ij}(z_{ij}^* - \bar{z}_i) + \epsilon_{ij}^*, \\ \epsilon_{ij}^* &\sim N(0, V_i^*) \end{aligned} \quad (2.5)$$

where D_i is a dummy variable to distinguish IPD trials from AD trials. For the i th IPD trial ($i = N + 1, \dots, N'$), $D_i = 1$, $y_{ij}^* = y_{ij}$, $V_i^* = \sigma_y^2$ and $z_{ij}^* = z_{ij}$. For the i th AD trial ($i = 1, \dots, N$), there is only one outcome ($j = 1$) and $D_i = 0$, $x_{i1} = 1$, $y_{i1}^* = d_i$, $V_i^* = V(d_i)$ assumed known, and $z_{i1}^* = \bar{z}_i$. Model (2.5) ensures that the AD from trials $i = 1, \dots, N$ help to estimate only the across-trial relationships (θ and γ_A), whereas the IPD from trials $i = N + 1, \dots, N'$ help to estimate all the parameters. That is, only the collected IPD contributes to the estimation of the within-trial relationships (μ and γ_W). As in model (2.1), we again assume that θ , μ , γ_A and γ_W are fixed effects, and σ_y^2 is common across trials.

Application to hypertension data

Consider again the hypertension data, and we now demonstrate how age modify the treatment effect on change in SBP (follow-up minus baseline) in the case that some trials provide IPD and the others provide only AD. To imitate situations involving IPD for some trials and only AD for others, we generated scenarios where only a limited number of trials (from 1 to 4 of the 5 trials) provided IPD, and the other trials just provided AD as presented in Table I, which is typical of the AD available to meta-analysts in practice. In each scenario, we carried out analyses by: (i) fitting the MR model (2.2) to AD from all 5 trials, (ii) fitting model (2.1) or (2.4) to IPD from only IPD trials available, (iii) fitting model (2.5) to the mixture of IPD and AD from all 5 trials. In both parts (ii) and (iii), the analyses were run for each possible combination of IPD and AD trials. For example, in the scenario that 2 trials provide IPD (i.e. 2 IPD trials and 3 AD trials), we performed 10 analyses, one for each combination of

which 2 trials provide IPD and 3 provide AD. In each scenario, we compared the results with those from a meta-analysis of IPD from all 5 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 by each method are shown in Table III. Fitting the MR model (2.2) to the AD for all 5 trials naturally provided results only of the across-trial relationships (α and β), whose estimates and standard errors were close to those of θ and γ_A from the full IPD analysis. Fitting model (2.1) or (2.4) to only available IPD provided both of the results for the within-trial and across-trial relationships. Estimates of γ_W and their standard errors got close to those from the full IPD analysis rapidly as the proportion of trials providing IPD increased; however, estimates of γ_A differed seriously from those from the full IPD analysis with huge standard errors, especially in the case of small proportion of IPD trials.

The strategy of combining IPD and AD by fitting model (2.5) allowed us to not only get more accurate results for the across-trial relationships but also evaluate on the within-trial relationships with a certain degree of precision. Including AD trials remarkably improved the precision of estimates for the across-trial relationships in comparison with analyses by using only the collected IPD. It was also confirmed that model (2.5) correctly allowed only the IPD trials to estimate μ and γ_W . This explains why the standard errors of γ_W increase as the proportion of IPD trials decreases and emphasises why it is better to obtain IPD from all trials.

2.2 Binary outcome and covariate

We now suppose that a single binary outcome variable Y and a single binary covariate Z are observed for each patient in each trial. Let y_{ij} and z_{ij} be a binary outcome and a binary covariate value for the j th patient ($j = 1, \dots, n_i$) in the i th trial.

If just meta-analysing the IPD; i.e. (y_{ij}, x_{ij}, z_{ij}) for $i = 1, \dots, N$ and $j = 1, \dots, n_i$, one can use the following one-stage model that accounts for the clustering of patients by a trial-specific effect (ϕ_i), and estimates a pooled treatment-covariate interaction (γ_W) based on within-trial information separated from the across-trial interaction (γ_A):

$$y_{ij} \sim \text{Bernoulli}(q_{ij}), \tag{2.6}$$

$$\log \frac{q_{ij}}{1 - q_{ij}} = \phi_0 + \phi_i + \theta x_{ij} + \mu z_{ij} + \gamma_A x_{ij} \bar{z}_i + \gamma_W x_{ij} (z_{ij} - \bar{z}_i),$$

$$j = 1, \dots, n_i; \quad i = 1, \dots, N.$$

Table III. Average of estimates and their standard errors for each parameter when analysing change in SBP (follow-up minus baseline) from the hypertension data, where estimates are averaged across all combinations of IPD trials.

	Number of trials providing IPD [‡]	Average of estimate		Average of standard error	
		Only-IPD	Model (2.5)	Only-IPD	Model (2.5)
Hypothetical treatment effect θ	5 / 5*	35.95		32.83	
	4 / 5	29.40	43.39	41.77	33.10
	3 / 5	3.29	43.34	65.05	32.75
	2 / 5	-31.55	43.56	129.2	32.89
	1 / 5	-11.86	44.45	1.480	32.85
	0 / 5 [†]	-	43.41	-	32.94
Covariate effect μ	5 / 5*	0.035		0.039	
	4 / 5	0.027	0.027	0.045	0.045
	3 / 5	-0.017	-0.018	0.059	0.059
	2 / 5	-0.117	-0.117	0.091	0.091
	1 / 5	-0.311	-0.303	0.176	0.175
	0 / 5 [†]	-	-0.303	-	0.175
Across-trial interaction effect γ_A	5 / 5*	-0.662		0.464	
	4 / 5	-0.569	-0.766	0.590	0.468
	3 / 5	-0.199	-0.766	0.920	0.463
	2 / 5	0.293	-0.768	1.831	0.464
	1 / 5	NA	-0.781	NA	0.464
	0 / 5 [†]	-	-0.766	-	0.466
Within-trial interaction effect γ_W	5 / 5*	0.087		0.055	
	4 / 5	0.092	0.091	0.063	0.063
	3 / 5	0.116	0.117	0.084	0.084
	2 / 5	0.165	0.166	0.131	0.130
	1 / 5	0.252	0.244	0.259	0.258

Only-IPD: Fit model (2.1) or (2.4) to only the collected IPD.

Model (2.5): Fit model (2.5) to the mixture of IPD and AD.

*Results by fitting model (2.1) to the full IPD from all 5 trials.

[†]Results by fitting the MR model (2.2) to the AD from all 5 trials.

[‡]The numbers of combinations of trials providing IPD are 5, 10, 10 and 5 in the scenarios of 1, 2, 3 and 4 IPD trials, respectively.

NA: Not available

Here, $\text{Bernoulli}(q_{ij})$ represents a random variable following Bernoulli distribution with probability q_{ij} , ϕ_0 is a fixed intercept, ϕ_i is the fixed effect for the i th trial, θ is a fixed hypothetical treatment effect in a trial with $\bar{z}_i = 0$, μ is a log odds ratio between covariate subgroups in control group, γ_A and γ_W are the across-trial and the within-trial interaction effect respectively. This modelling framework were proposed by Riley and Steyerberg (2010) in the case of binary outcomes with a single group. Note that, as in model (2.1), θ , μ , γ_A and γ_W are treated as fixed effects here.

In general meta-analytic situations, we observe only the AD from each trial, rather than the IPD from each patient. Here, the AD consist of the grouped forms of outcome and covariate; i.e. $(n_{iT}, m_{iT}, n_{iC}, m_{iC}, \bar{z}_i)$ where n_{iT} (or n_{iC}) is the number of patients assigned to treatment (or control) group in the i th trial, m_{iT} (or m_{iC}) is the number of patients with $Y = 1$ in n_{iT} (or n_{iC}) patients, and $\bar{z}_i = \sum_{j=1}^{n_i} z_{ij}/n_i$ is the proportion of patients with $Z = 1$ in the i th trial. Then, an MR model which has the same form as model (2.2) can be applied to the AD, where

$$d_i = \log \frac{m_{iT}(n_{iC} - m_{iC})}{m_{iC}(n_{iT} - m_{iT})}$$

and

$$\sigma_{di}^2 = V(d_i) = \frac{1}{m_{iT}} + \frac{1}{n_{iT} - m_{iT}} + \frac{1}{m_{iC}} + \frac{1}{n_{iC} - m_{iT}}.$$

As another approach for meta-analysing the AD, we here consider to partially recreate a binary data form of IPD from the grouped form of AD, where patients with $Y = 1$ or $Y = 0$ in each group are represented by a series of ones or zeros (Riley, Simmonds and Look, 2007). In particular, (n_{ik}, m_{ik}) for $i = 1, \dots, N$ and $k \in \{T, C\}$, where k is a group indicator that takes a value of T for treatment or C for control, are rewritten as (y_{ij}, x_{ij}) for $i = 1, \dots, N$ and $j = 1, \dots, n_i$. Since it is impossible to recreate the patient-level covariates directly from the AD, the information of covariate is limited to $\bar{z}_i$ for each trial. Then, one can use the following model:

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(q_{ij}), \\ \log \frac{q_{ij}}{1 - q_{ij}} &= \phi_0 + \phi_i + \theta x_{ij} + \gamma_A x_{ij} \bar{z}_i, \\ i &= 1, \dots, N. \end{aligned} \tag{2.7}$$

Thompson, Turner and Warn (2001) suggested that this direct modelling would be appropriate in comparison to using the ordinary MR model where the log odds ratio estimated from

each trial are assumed to be normally distributed with known variances, especially when the observed event probabilities in a particular trial are close to 0 or 1, and where the sample size in each trial is small.

Application to home safety education data

Consider the home safety education data, and we now demonstrate how gender of children modifies the intervention effect on the provision of a stair gate. Fitting model (2.6) to the IPD from the 5 trials, estimates of each parameter in model (2.6) were $\hat{\theta} = 7.184$ (s.e. = 4.300 and p-value = 0.095), $\hat{\mu} = 0.071$ (s.e. = 0.116 and p-value = 0.541), $\hat{\gamma}_A = -13.31$ (s.e. = 8.343 and

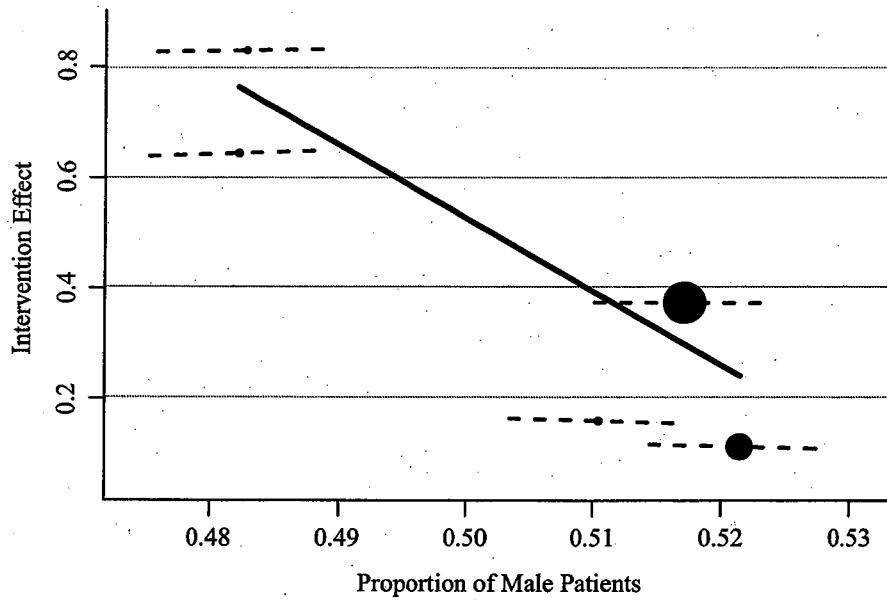


Figure 4. Scatter plot for the 5 trials in home safety education with across-trial and within-trial interaction effect estimates, in which:

- A solid line represents the across-trial interaction ($\hat{\gamma}_A$) between the proportion of male patients ($\bar{z}_i$) and intervention effect estimated by model (2.7).
- Dashed lines represent the within-trial interaction ($\hat{\gamma}_W$) between gender of children and intervention effect estimated separately within each trial using IPD and model (2.6) without γ_A .
- The gradient of each dashed line indicates the change in intervention effect from females to males within each trial; the length of the dashed lines is unimportant and is kept the same for each simply to aid clarity.
- Each circle represents a trial and is centered at $\bar{z}_i$ in each trial; the circle size is proportional to the sample size in each trial.

p-value = 0.111) and $\hat{\gamma}_W = -0.212$ (s.e. = 0.165 and p-value = 0.199), respectively. Fitting model (2.7) to the AD from the 5 trials, estimates of each parameter in model (2.7) were $\hat{\theta} = 7.176$ (s.e. = 4.299 and p-value = 0.095) and $\hat{\gamma}_A = -13.30$ (s.e. = 8.341 and p-value = 0.111), respectively. The across-trial interaction was substantially different from the within-trial interaction on the point estimates. As suggested in the application to the hypertension data, this shows the importance of separating the treatment-covariate interaction. 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 intervention and gender. 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 proportion of male participants were fairly homogeneous across the 5 trials. Figure 4 also shows this difference between $\hat{\gamma}_A$ and $\hat{\gamma}_W$; the within-trial interaction (dashed lines) have almost flat gradients, where the across-trial interaction (solid line) has a steep negative gradient. It highlights the pitfall of using $\hat{\gamma}_A$ to make inferences about $\hat{\gamma}_W$, that is, ecological bias and confounding.

2.2.1 The case for mixture of IPD and AD

Consider the same case supposed in Chapter 2.1.1; i.e. a meta-analysis of the mixture of N AD trials and $N' - N$ IPD trials. When a mixture of IPD and AD trials are available, model (2.6) must be modified to combine IPD and AD. As mentioned in Chapter 2.1.1, simple solutions are to deal all the data as AD, or to use only the collected IPD. In contrast, Riley and Steyerberg (2010) proposed a model for combining IPD and AD. All these approaches are now described.

Model that uses only AD trials

Once the IPD for trials $i = N + 1, \dots, N'$ are summarised to the AD, model (2.7) can be applied to the AD for all trials $i = 1, \dots, N'$.

Model that uses only IPD trials available

If one uses only the collected IPD, model (2.6) can be applied to the IPD from trials $i = N + 1, \dots, N'$. When the number of IPD trials is one (i.e. $N' = N + 1$), model (2.6) is

modified as follow:

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(q_{ij}), \\ \log \frac{q_{ij}}{1 - q_{ij}} &= \phi_0 + \phi_i + \theta x_{ij} + \mu z_{ij} + \gamma_W x_{ij}(z_{ij} - \bar{z}_i) \\ j &= 1, \dots, n_i; i = N. \end{aligned} \tag{2.8}$$

Model that uses both IPD and AD trials

The model for combining IPD and AD proposed by Riley and Steyerberg (2010) is as follows:

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(q_{ij}), \\ \log \frac{q_{ij}}{1 - q_{ij}} &= \phi_0 + \phi_i + \theta x_{ij} + D_i \mu z_{ij} + \gamma_A x_{ij} \bar{z}_i + D_i \gamma_W x_{ij}(z_{ij} - \bar{z}_i) \end{aligned} \tag{2.9}$$

where D_i is a dummy variable to distinguish IPD trials ($D_i = 1$) from AD trials ($D_i = 0$), and (y_{ij}, x_{ij}) for the AD trials $i = 1, \dots, N$ are recreated from (n_{ik}, m_{ik}) for $k \in \{T, C\}$ by the same way described above. As in model (2.5), Model (2.9) ensures that the AD from trials $i = 1, \dots, N$ help to estimate only the across-trial relationships (θ and γ_A), whereas the IPD from trials $i = N + 1, \dots, N'$ help to estimate all the parameters. We again assume that θ , μ , γ_A and γ_W are fixed effects.

Application to home education data

Consider again the home safety education data, and we now demonstrate how gender of children modifies the intervention effect on the provision of the stair gate in the case that some trials provide IPD and the others provide only AD. To imitate situations involving IPD for some trials and only AD for others, we generated scenarios in the same manner as the application to the hypertension data; i.e. we assumed that only a limited number of trials (from 1 to 4 of the 5 trials) provided IPD and the other trials just provided AD as presented in Table II. In each scenario, we carried out analyses by: (i) fitting model (2.7) to AD from all 5 trials, (ii) fitting model (2.6) or (2.8) to IPD from only IPD trials available, (iii) fitting model (2.9) to the mixture of IPD and AD from all 5 trials. In both parts (ii) to (iii), the analyses were run for each possible combination of IPD and AD trials. In each scenario, we compared the results with those from a meta-analysis of IPD from all 5 trials.

The results by each method are shown in Table IV. Fitting model (2.7) to the AD for all 5 trials naturally provided results only of the across-trial relationships. Fitting model (2.6) or (2.8) to only available IPD provided both of the results for the within-trial and across-trial

relationships. By the similar trend shown Table III, estimates of γ_W and their standard errors got close to those from the full IPD analysis rapidly as the proportion of trials providing IPD increased; however, the estimates of γ_A differed seriously from those from the full IPD analysis with huge standard errors especially in the case of small proportion of IPD trials.

As in the application to the hypertension data, the strategy of combining IPD and AD by fitting model (2.9) improved the precision of estimates for the across-trial relationships in comparison with analyses by using only the collected IPD. It was also confirmed that model (2.9) correctly allowed only the IPD trials to estimate μ and γ_W .

Table IV. Average of estimates and their standard errors for each parameter when analysing the home safety education data, where estimates are averaged across all combinations of IPD trials.

	Number of trials providing IPD [‡]	Average of estimate		Average of standard error	
		Only-IPD	Model (2.9)	Only-IPD	Model (2.9)
Hypothetical treatment effect θ	5 / 5*	7.184		4.300	
	4 / 5	7.321	7.179	5.010	4.300
	3 / 5	8.253	7.171	7.179	4.301
	2 / 5	-7.953	7.155	58.90	4.302
	1 / 5	0.420	7.223	0.307	4.315
	0 / 5 [†]	-	7.176	-	4.299
Covariate effect μ	5 / 5*	0.071		0.116	
	4 / 5	0.082	0.081	0.135	0.135
	3 / 5	0.076	0.075	0.170	0.169
	2 / 5	0.045	0.041	0.239	0.239
	1 / 5	-0.042	-0.056	0.424	0.424
Across-trial interaction effect γ_A	5 / 5*	-13.31		8.343	
	4 / 5	-13.60	-13.30	9.726	8.344
	3 / 5	-15.46	-13.29	13.99	8.345
	2 / 5	17.96	-13.26	120.5	8.347
	1 / 5	NA	-13.39	NA	8.371
	0 / 5 [†]	-	-13.31	-	8.343
Within-trial interaction effect γ_W	5 / 5*	-0.212		0.165	
	4 / 5	-0.225	-0.225	0.192	0.192
	3 / 5	-0.218	-0.218	0.242	0.242
	2 / 5	-0.178	-0.176	0.343	0.344
	1 / 5	-0.048	-0.041	0.615	0.614

Only-IPD: Fit model (2.6) or (2.8) to only the collected IPD.

Model (2.9): Fit model (2.9) to the mixture of IPD and AD.

*Results by fitting model (2.6) to the full IPD from all 5 trials.

[†]Results by fitting model (2.7) to the AD from all 5 trials.

[‡]The numbers of combinations of trials providing IPD are 5, 10, 10 and 5 in the scenarios of 1, 2, 3 and 4 IPD trials, respectively.

NA: Not available

3 Proposed methods

3.1 Meta-analysis with marginalising the missing IPD

A structural limitation of the MR model (2.2) and model (2.7) is that their inferential objectives are restricted to the across-trial relationships (the hypothetical treatment effect and the across-trial interaction effect). As illustrated in Chapter 2, the across-trial relationships are prone to trial-level confounding and often suffer from large standard error, in comparison with the within-trial relationships (the covariate effect and the within-trial interaction effect). We now introduce a meta-analysis method for estimating not only the across-trial relationships but also the within-trial relationships when all trials provide only AD. The proposed method is simply extended to the case that some trials provide IPD and the others provide only AD (i.e. to combine IPD and AD). We here suppose the situation where a single continuous outcome and a single continuous covariate are observed from each patient in each trial.

3.1.1 The case for only AD

Consider a meta-analysis of N trials which provide only AD. Original IPD which have been observed in each trial can be regarded as missing data. We first assume the following IPD meta-analysis model to the missing IPD:

$$\begin{aligned} y_{ij} &= \phi_i + \theta x_{ij} + \mu z_{ij} + \gamma_A x_{ij} \bar{z}_i + \gamma_W x_{ij} (z_{ij} - \bar{z}_i) + \epsilon_{ij}, \\ \epsilon_{ij} &\sim N(0, \sigma_y^2), \\ j &= 1, \dots, n_i; \quad i = 1, \dots, N \end{aligned} \tag{3.1}$$

where $(\bar{y}_{iT}, s_{yiT}^2, \bar{z}_{iT}, s_{z iT}^2, \bar{y}_{iC}, s_{y iC}^2, \bar{z}_{iC}, s_{z iC}^2)$ are available for the i th trial, in place of the patient-specific observations (y_{ij}, x_{ij}, z_{ij}) for $j = 1, \dots, n_i$. Then, we cannot obtain maximum likelihood estimates (MLEs) of parameters of interest directly from (3.1) for the reason that

each trial provides only AD. Indeed, sufficient statistics for model (3.1) are given by

$$\begin{aligned} \sum_{j \in T} y_{ij}, \quad \sum_{j \in T} y_{ij}^2, \quad \sum_{j \in T} z_{ij}, \quad \sum_{j \in T} z_{ij}^2, \quad \sum_{j \in T} y_{ij} z_{ij}, \\ \sum_{j \in C} y_{ij}, \quad \sum_{j \in C} y_{ij}^2, \quad \sum_{j \in C} z_{ij}, \quad \sum_{j \in C} z_{ij}^2, \quad \sum_{j \in C} y_{ij} z_{ij} \end{aligned}$$

for $i = 1, \dots, N$; while $\sum_{j \in T} y_{ij} z_{ij}$ and $\sum_{j \in C} y_{ij} z_{ij}$ are not available.

To estimate the parameters included in model (3.1) by using only AD, the covariates of patients assigned to the treatment (or control) group in the i th trial are assumed to be independent and identically distributed normal random variables with mean m_{ziT} (or m_{ziC}) and variance σ_{ziT}^2 (or σ_{ziC}^2); i.e.

$$z_{ij} \sim \begin{cases} N(m_{ziT}, \sigma_{ziT}^2), & j \in T \\ N(m_{ziC}, \sigma_{ziC}^2), & j \in C \end{cases} \quad (3.2)$$

Here, $\bar{z}_i = \sum_{j=1}^{n_i} z_{ij}/n_i$ is considered to be constant, and parameters included in the covariate distribution (3.2) are estimated by

$$\hat{m}_{ziT} = \bar{z}_{iT}, \quad \hat{\sigma}_{ziT}^2 = s_{ziT}^2, \quad \hat{m}_{ziC} = \bar{z}_{iC}, \quad \hat{\sigma}_{ziC}^2 = s_{ziC}^2 \quad (3.3)$$

for $i = 1, \dots, N$. If we also assume that z_{ij} and ϵ_{ij} are independent of each other, we have the following conditional distribution of y_{ij} given z_{ij} :

$$y_{ij}|z_{ij} \sim \begin{cases} N(\phi_i + \theta + (\mu + \gamma_W)z_{ij} + (\gamma_A - \gamma_W)\bar{z}_i, \sigma_y^2), & j \in T \\ N(\phi_i + \mu z_{ij}, \sigma_y^2), & j \in C \end{cases} \quad (3.4)$$

And then, marginalising the joint distribution of (y_{ij}, z_{ij}) with respect to z_{ij} , we have the following marginal distribution of y_{ij} :

$$y_{ij} \sim \begin{cases} N(m_{yiT}, \sigma_{yiT}^2), & j \in T \\ N(m_{yiC}, \sigma_{yiC}^2), & j \in C \end{cases} \quad (3.5)$$

where

$$m_{yiT} = \phi_i + \theta + \mu \hat{m}_{ziT} + \gamma_A \bar{z}_i + \gamma_W (\hat{m}_{ziT} - \bar{z}_i), \quad (3.6)$$

$$\sigma_{yiT}^2 = (\mu + \gamma_W)^2 \hat{\sigma}_{ziT}^2 + \sigma_y^2$$

and

$$m_{yiC} = \phi_i + \mu \hat{m}_{ziC}, \quad (3.7)$$

$$\sigma_{yiC}^2 = \mu^2 \hat{\sigma}_{ziC}^2 + \sigma_y^2.$$

Therefore, a log-likelihood function for the parameters $(\phi_1, \dots, \phi_N, \theta, \mu, \gamma_A, \gamma_W, \sigma_y^2)$ included in model (3.1) is given by

$$l_{AD} = \frac{1}{2} \sum_{i=1}^N \left[\frac{n_{iT} \left\{ (\bar{y}_{iT} - \phi_i - \theta - \mu \hat{m}_{ziT} - \gamma_A \bar{z}_i - \gamma_W (\hat{m}_{ziT} - \bar{z}_i))^2 + s_{yiT}^2 \right\}}{(\mu + \gamma_W)^2 \hat{\sigma}_{ziT}^2 + \sigma_y^2} \right. \\ \left. - \frac{n_{iC} \left\{ (\bar{y}_{iC} - \phi_i - \mu \hat{m}_{ziC})^2 + s_{yiC}^2 \right\}}{\mu^2 \hat{\sigma}_{ziC}^2 + \sigma_y^2} \right. \\ \left. - n_{iT} \log((\mu + \gamma_W)^2 \hat{\sigma}_{ziT}^2 + \sigma_y^2) - n_{iC} \log(\mu^2 \hat{\sigma}_{ziC}^2 + \sigma_y^2) \right]. \quad (3.8)$$

A remarkable aspect of using the log-likelihood (3.8) is that the correlation between y_{ij} and z_{ij} are replaced with the correlation between $\bar{y}_{iT}$ and $\hat{m}_{ziT}$, or $\bar{y}_{iC}$ and $\hat{m}_{ziC}$. Since we cannot compute MLEs of the parameters from the log-likelihood (3.8) in a closed-form, it is necessary to use an iterative numerical computing algorithm such as Newton-Raphson method.

3.1.2 The case for mixture of IPD and AD

Consider the meta-analysis of N' trials which consist of N AD trials and $N' - N$ IPD trials. As in Chapter 3.1.1, we first assume the IPD meta-analysis model to the collected IPD and the missing IPD; i.e.

$$y_{ij} = \phi_i + \theta x_{ij} + \mu z_{ij} + \gamma_A x_{ij} \bar{z}_i + \gamma_W x_{ij} (z_{ij} - \bar{z}_i) + \epsilon_{ij}, \quad (3.9) \\ \epsilon_{ij} \sim N(0, \sigma_y^2), \\ j = 1, \dots, n_i; \quad i = 1, \dots, N'$$

where $(\bar{y}_{iT}, s_{yiT}^2, \bar{z}_{iT}, s_{ziT}^2, \bar{y}_{iC}, s_{yiC}^2, \bar{z}_{iC}, s_{ziC}^2)$ for $i = 1, \dots, N$ and (y_{ij}, x_{ij}, z_{ij}) for $i = N + 1, \dots, N'$ and $j = 1, \dots, n_i$ are available. Because all trials are independent of each other, a log-likelihood for the parameters included in model (3.9) can be derived as summation of the log-likelihood for the AD trials and that for the IPD trials. The former is already given by (3.8), and the latter is simply given by using the normal densities as follows:

$$l_{IPD} = \frac{1}{2} \sum_{i=N+1}^{N'} \sum_{j=1}^{n_i} \left\{ -\log \sigma_y^2 - \sigma_y^{-2} (y_{ij} - \phi_i - \theta x_{ij} - \mu z_{ij} - \gamma_A x_{ij} \bar{z}_i - \gamma_W x_{ij} (z_{ij} - \bar{z}_i))^2 \right\}.$$

Then, we can estimate the parameters $(\phi_1, \dots, \phi_{N'}, \theta, \mu, \gamma_A, \gamma_W, \sigma_y^2)$ included in model (3.9) by maximising the log-likelihood

$$l_{AD} + l_{IPD}$$

with respect to the parameters.

3.2 Meta-analysis based on simulated IPD

Subsequently, we introduce meta-analysis methods based on simulated IPD (SIPD), in which the missing IPD are reconstructed by using the scheme of marginalising the missing IPD. For the case that all trials provide only AD, the proposed method takes the following procedures for inference of parameters.

For the case that all trials provide only AD:

- (1) Generate multiple sets of SIPD for each trial.
- (2) Fit an IPD meta-analysis model to each set of SIPD.
- (3) Suitably summarise resulting estimates from the set of meta-analyses from Step (2).

We refer to these whole estimating processes as SIPD method. Figure 5 shows a flow diagram of the SIPD method. Furthermore, for the case that some trials provide IPD and the others provide only AD, the proposed method is extended to combine IPD and AD.

For the case that some trials provide IPD and the others provide only AD:

- (1) Generate multiple sets of SIPD for each trial providing only AD.
- (2) Fit an IPD 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).

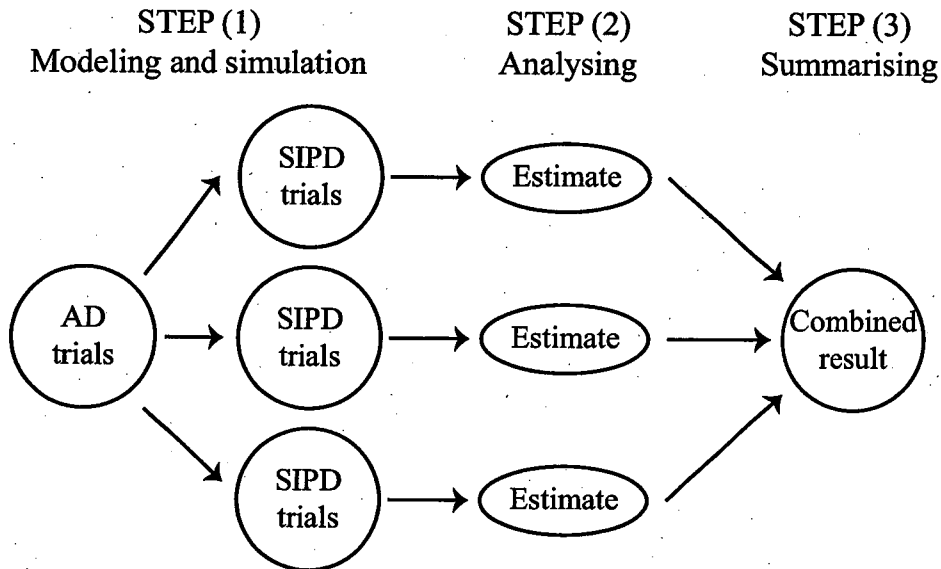


Figure 5. Flow diagram of SIPD method.

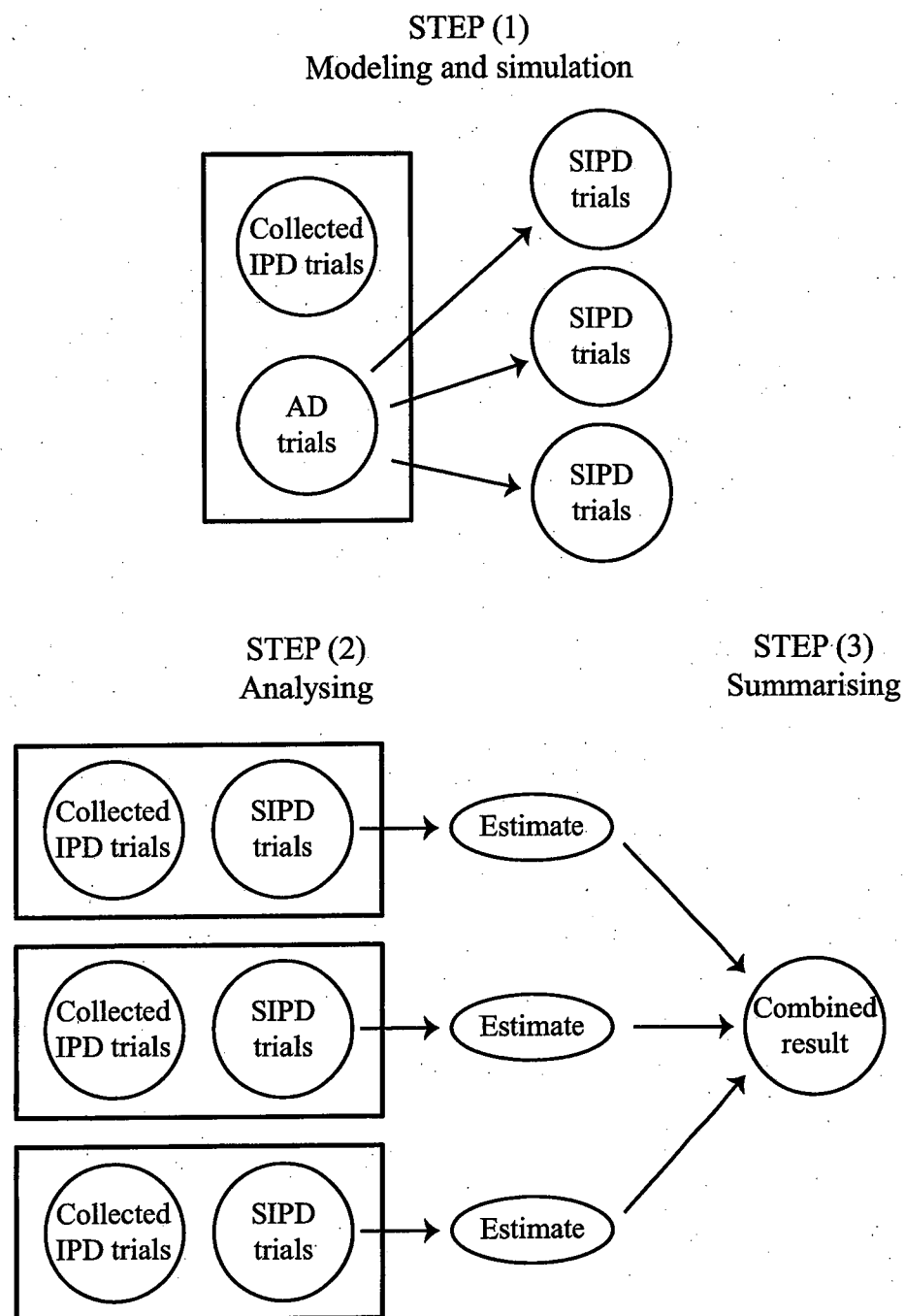


Figure 6. Flow diagram of SIPD method with combined the collected IPD.

Figure 6 shows a flow diagram of the SIPD method with combined the collected IPD. In this chapter, each step in the SIPD method is described in detail for two situations of meta-analysis data (continuous outcome and covariate, and binary outcome and covariate).

3.2.1 Continuous outcome and covariate

Now, let

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

where $Y_{\text{miss-IPD}}$ is the uncollected IPD for trials $i = 1, \dots, N$ and Y_{AD} is the AD summarised from them. Suppose that Y_{AD} can be written by a function $h(Y_{\text{miss-IPD}})$; i.e. $Y_{\text{AD}} = h(Y_{\text{miss-IPD}})$. Now, the function h transforms the patient-specific observations into the sample mean and the sample variance for each group in each trial. Then, we again consider the IPD meta-analysis model (3.1) to $Y_{\text{miss-IPD}}$, and assume that the covariates of patients for each group in the i th trial follow the normal distribution (3.2). If we again assume that z_{ij} and ϵ_{ij} are independent of each other, we have the conditional distribution of y_{ij} given z_{ij} as (3.4). Here, parameters in the covariate distributions can be estimated as (3.3), and now let

$$\hat{\xi} = \{(\hat{m}_{ziT}, \hat{\sigma}_{ziT}^2, \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2) : i = 1, \dots, N\}. \quad (3.10)$$

And also, let

$$\eta = (\phi_1, \dots, \phi_N, \theta, \mu, \gamma_A, \gamma_W, \sigma_y^2) \quad (3.11)$$

be parameter to be estimated.

Combining IPD and AD For the case that some trials provide IPD and the others provide only AD trials, let

$$Y_{\text{IPD}} = \{(y_{ij}, x_{ij}, z_{ij}) : j = 1, \dots, n_i; i = N + 1, \dots, N'\}$$

be the collected IPD for trials $i = N + 1, \dots, N'$. As in the case that all trials provide only AD, we consider the IPD meta-analysis model (3.9) to $Y_{\text{miss-IPD}}$ and Y_{IPD} , and assume that the covariates of patients for each group in the i th trial follow the normal distribution (3.2). Here, parameters in the covariate distributions can be estimated as (3.3), and let

$$\hat{\xi} = \{(\hat{m}_{ziT}, \hat{\sigma}_{ziT}^2, \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2) : i = 1, \dots, N'\}. \quad (3.12)$$

And also, let

$$\eta = (\phi_1, \dots, \phi_{N'}, \theta, \mu, \gamma_A, \gamma_W, \sigma_y^2) \quad (3.13)$$

be parameter to be estimated. For both cases, each step in the SIPD method is now described from frequentist and Bayesian perspectives.

SIPD method via frequentist procedure

Step (1): Generating SIPD

Let $f(Y_{\text{miss-IPD}}; \hat{\xi}, \eta)$ be a density function of the missing IPD with parameter η in (3.11), which is equivalent to product of bivariate normal densities from

$$\begin{bmatrix} y_{ij} \\ z_{ij} \end{bmatrix} \sim \begin{cases} N_2 \left(\begin{bmatrix} m_{yiT} \\ \hat{m}_{ziT} \end{bmatrix}, \begin{bmatrix} \sigma_{yiT}^2 & (\mu + \gamma_W)\hat{\sigma}_{ziT}^2 \\ (\mu + \gamma_W)\hat{\sigma}_{ziT}^2 & \hat{\sigma}_{ziT}^2 \end{bmatrix} \right), & j \in T \\ N_2 \left(\begin{bmatrix} m_{yiC} \\ \hat{m}_{ziC} \end{bmatrix}, \begin{bmatrix} \sigma_{yiC}^2 & \mu\hat{\sigma}_{ziC}^2 \\ \mu\hat{\sigma}_{ziC}^2 & \hat{\sigma}_{ziC}^2 \end{bmatrix} \right), & j \in C \end{cases} \quad (3.14)$$

for $i = 1, \dots, N$ and $j = 1, \dots, n_i$. Here, m_{yiT} , σ_{yiT}^2 , m_{yiC} and σ_{yiC}^2 are given as (3.6) and (3.7). Then, we draw the SIPD (say $Y_{\text{miss-IPD}}^*$) from the conditional distribution of $Y_{\text{miss-IPD}}$ given Y_{AD} and a known parameter value $\tilde{\eta}$; i.e

$$Y_{\text{miss-IPD}}^* \sim f(Y_{\text{miss-IPD}} | Y_{\text{AD}}; \hat{\xi}, \tilde{\eta}) \quad (3.15)$$

where $\tilde{\eta}$ is computed by maximising the following likelihood under an assumption known as CAR (Coarsening at Random) by Heitjan and Rubin (1991):

$$L_{\text{AD}}(\eta) = f(Y_{\text{AD}}; \hat{\xi}, \eta) = \int_{Y_{\text{AD}}=h(Y_{\text{miss-IPD}})} f(Y_{\text{miss-IPD}}; \hat{\xi}, \eta) dY_{\text{miss-IPD}}. \quad (3.16)$$

This means that

$$\tilde{\eta} = \arg \max_{\eta} L_{\text{AD}}(\eta).$$

Once obtaining the parameter estimate, $\tilde{\eta}$, we can get R sets of SIPD by repeated drawing of (3.15); say

$$Y_{\text{miss-IPD}}^{[r]} = \{(y_{ij}^{[r]}, x_{ij}^{[r]}, z_{ij}^{[r]}) : j = 1, \dots, n_i; i = 1, \dots, N\} \quad (3.17)$$

for $r = 1, \dots, R$. To use (3.15) for generating the SIPD, it is necessary to calculate (3.16) in an explicit form and then draw from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}; \hat{\xi}, \tilde{\eta})$. Below, we describe how to calculate $L_{\text{AD}}(\eta)$ and draw from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}; \hat{\xi}, \tilde{\eta})$.

Calculating $L_{AD}(\eta)$: The calculation of the likelihood (3.16) requires the integration with respect to $Y_{\text{miss-IPD}}$ over the region that satisfies $Y_{\text{miss-IPD}} = h(Y_{AD})$. In particular, by the derivation in Appendix A, this is given by

$$L_{AD}(\eta) = K \times \prod_{i=1}^N \left[(\sigma_{yiT}^2)^{-n_{iT}/2} \exp \left\{ -\frac{n_{iT}}{2\sigma_{yiT}^2} ((\bar{y}_{iT} - m_{yiT})^2 + s_{yiT}^2) \right\} \right. \\ \left. \times (\sigma_{yiC}^2)^{-n_{iC}/2} \exp \left\{ -\frac{n_{iC}}{2\sigma_{yiC}^2} ((\bar{y}_{iC} - m_{yiC})^2 + s_{yiC}^2) \right\} \right] \quad (3.18)$$

where K is a constant term unrelated to η . Taking the logarithm of (3.18), we can derive a log-likelihood which has the same form as (3.8).

Drawing from $f(Y_{\text{miss-IPD}}|Y_{AD}; \hat{\xi}, \tilde{\eta})$: We can easily derive $f(Y_{\text{miss-IPD}}; \hat{\xi}, \tilde{\eta})$ from (3.14) with the known parameter $\tilde{\eta}$, while $f(Y_{\text{miss-IPD}}|Y_{AD}; \hat{\xi}, \tilde{\eta})$ is difficult to derive exactly because its sample space is defined on the region that satisfies $Y_{AD} = h(Y_{\text{miss-IPD}})$. 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 drawn from $f(Y_{\text{miss-IPD}}|Y_{AD}; \hat{\xi}, \tilde{\eta})$, must be equivalent to the corresponding sample means and sample variances in Y_{AD} . We here describe how to achieve the drawing from $f(Y_{\text{miss-IPD}}|Y_{AD}; \hat{\xi}, \tilde{\eta})$ by using a sampling technique proposed by Lindqvist and Taraldsen (2005).

Now, we represent the observation vectors for each group in the i th trial as follows:

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

Recall that $Y_{\text{miss-IPD}}$ denotes the uncollected IPD; i.e. (y_{ij}, x_{ij}, z_{ij}) for $i = 1, \dots, N$ and $j = 1, \dots, n_i$. Because of between-trial and between-group independence, we have the density functions of $Y_{\text{miss-IPD}}$ given Y_{AD} and $\tilde{\eta}$ as follows:

$$f(Y_{\text{miss-IPD}}|Y_{AD}; \hat{\xi}, \tilde{\eta}) = \\ \prod_{i=1}^N f(\mathbf{z}_{iT} | \bar{z}_{iT}, s_{ziT}^2; \hat{m}_{ziT}, \hat{\sigma}_{ziT}^2) f(\mathbf{y}_{iT} | \bar{y}_{iT}, s_{yiT}^2, \mathbf{z}_{iT}; \tilde{\eta}) \\ \times f(\mathbf{z}_{iC} | \bar{z}_{iC}, s_{ziC}^2; \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2) f(\mathbf{y}_{iC} | \bar{y}_{iC}, s_{yiC}^2, \mathbf{z}_{iC}; \tilde{\eta}). \quad (3.19)$$

Then, the r th set of SIPD for the i th trial (say $\mathbf{y}_{iT}^{[r]}$, $\mathbf{z}_{iT}^{[r]}$, $\mathbf{y}_{iC}^{[r]}$ and $\mathbf{z}_{iC}^{[r]}$) are generated as random samples drawn from the corresponding conditional distribution in (3.19), that indicates the following sequential sampling procedure:

$$\mathbf{z}_{iT}^{[r]} \sim f(\mathbf{z}_{iT} | \bar{z}_{iT}, s_{ziT}^2; \hat{m}_{ziT}, \hat{\sigma}_{ziT}^2), \\ \mathbf{y}_{iT}^{[r]} | \mathbf{z}_{iT}^{[r]} \sim f(\mathbf{y}_{iT} | \bar{y}_{iT}, s_{yiT}^2, \mathbf{z}_{iT}^{[r]}; \tilde{\eta})$$

and

$$\begin{aligned} \mathbf{z}_{iC}^{[r]} &\sim f(\mathbf{z}_{iC} | \bar{z}_{iC}, s_{ziC}^2; \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2), \\ \mathbf{y}_{iC}^{[r]} | \mathbf{z}_{iC}^{[r]} &\sim f(\mathbf{y}_{iC} | \bar{y}_{iC}, s_{yiC}^2, \mathbf{z}_{iC}^{[r]}; \tilde{\eta}). \end{aligned}$$

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

Here, $\mathbf{z}_{iT}^{[r]}$ represents samples from the conditional normal distribution given sample mean $\bar{z}_{iT}$ and sample variance s_{ziT}^2 . A result by Lindqvist and Taraldsen (2005) described in Appendix B allows one to achieve this drawing as follows:

$$\mathbf{z}_{iT}^{[r]} = \left\{ \bar{z}_{iT} + \frac{u_{ij} - \bar{u}_i}{s_{ui}} s_{ziT} : j \in T \right\} \quad (3.20)$$

where $\{u_{ij} : j \in T\}$ denotes n_{iT} random samples from the standard normal distribution, $\bar{u}_i$ and s_{ui}^2 are a sample mean and a sample variance summarised from them respectively. Furthermore, letting $\tilde{\mu}$ and $\tilde{\gamma}_W$ be the corresponding components in $\tilde{\eta}$, we can draw $\mathbf{y}_{iT}^{[r]}$ in a similar way to (3.20) as follows:

$$\mathbf{y}_{iT}^{[r]} = \{ \bar{y}_{iT} + (\tilde{\mu} + \tilde{\gamma}_W)(z_{ij}^{[r]} - \bar{z}_{iT}) + \delta_i(v_{ij} - \bar{v}_i) : j \in T \} \quad (3.21)$$

where $\{v_{ij} : j \in T\}$ denotes n_{iT} random samples from the standard normal distribution, $\bar{v}_i$ and s_{vi}^2 are a sample mean and a sample variance summarised from them respectively. And also

$$\begin{aligned} \delta_i &= \frac{-(\tilde{\mu} + \tilde{\gamma}_W)s_{zi,vi} + \sqrt{(\tilde{\mu} + \tilde{\gamma}_W)^2 s_{zi,vi}^2 - s_{vi}^2 s_{ziT}^2 + s_{vi}^2 s_{yiT}^2}}{s_{vi}^2}, \\ s_{zi,vi} &= \frac{1}{n_{iT} - 1} \sum_{j \in T} (z_{ij}^{[r]} - \bar{z}_{iT})(v_{ij} - \bar{v}_i). \end{aligned}$$

For $j \in C$, (3.20) and (3.21) can be used in a similar manner, except that $(\tilde{\mu} + \tilde{\gamma}_W)$ in (3.21) is replaced by $\tilde{\mu}$.

Combining IPD and AD For the case that some trials provide IPD and the others provide only AD, the collected IPD are essentially utilised for supplementing to the computation of $\tilde{\eta}$. Because all trials are independent of each other, we compute $\tilde{\eta}$ by maximising product of the likelihood for the IPD trials and that for the AD trials; i.e.

$$\tilde{\eta} = \arg \max_{\eta} L_{\text{IPD}}(\eta) \times L_{\text{AD}}(\eta) \quad (3.22)$$

where $L_{AD}(\eta)$ is already given by (3.18), and $L_{IPD}(\eta)$ is given by

$$L_{IPD}(\eta) = f(Y_{IPD}; \hat{\xi}, \eta) = K \times \prod_{i=N+1}^{N'} \prod_{j=1}^{n_i} \frac{1}{\sigma_y} \exp \left\{ -\frac{1}{2\sigma_y^2} (y_{ij} - \phi_i - \theta x_{ij} - \mu z_{ij} - \gamma_A x_{ij} \bar{z}_i - \gamma_W x_{ij} (z_{ij} - \bar{z}_i))^2 \right\} \quad (3.23)$$

where K is a constant term unrelated to η . Once obtaining $\tilde{\eta}$ from (3.22), we can get R sets of SIPD in the same way as (3.20) and (3.21).

The fact that we can use the IPD from a part of trials offers another solution to compute $\tilde{\eta}$. The IPD trials partly provide information of $\sum_{j \in T} y_{ij} z_{ij}$ and $\sum_{j \in C} y_{ij} z_{ij}$, indicating that EM (Expectation Maximisation) algorithm by Dempster, Laird and Rubin (1977) can be applied for the computation of $\tilde{\eta}$.

EM algorithm: Because of between-trial and between-group independency, we have the density function of $Y_{\text{miss-IPD}}$ and Y_{IPD} as follows:

$$f(Y_{\text{miss-IPD}}, Y_{IPD}; \hat{\xi}, \eta) = \prod_{i=1}^{N'} f(\mathbf{y}_{iT}, \mathbf{z}_{iT}, \mathbf{y}_{iC}, \mathbf{z}_{iC}; \hat{m}_{ziT}, \hat{\sigma}_{ziT}^2, \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2, \eta).$$

The EM algorithm repeats two steps of calculation referred to as E-step and M-step. In E-step, given (Y_{AD}, Y_{IPD}) and a current parameter value $\tilde{\eta}^{[t]}$, we calculate the following conditional expectation:

$$\begin{aligned} Q(\eta, \tilde{\eta}^{[t]}) &= E_{\tilde{\eta}^{[t]}} \left[\sum_{i=1}^{N'} \log f(\mathbf{y}_{iT}, \mathbf{z}_{iT}, \mathbf{y}_{iC}, \mathbf{z}_{iC}; \hat{m}_{ziT}, \hat{\sigma}_{ziT}^2, \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2, \eta) \middle| Y_{AD}, Y_{IPD} \right] \\ &= E_{\tilde{\eta}^{[t]}} \left[\sum_{i=1}^N \log f(\mathbf{y}_{iT}, \mathbf{z}_{iT}, \mathbf{y}_{iC}, \mathbf{z}_{iC}; \hat{m}_{ziT}, \hat{\sigma}_{ziT}^2, \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2, \eta) \middle| Y_{AD} \right] \\ &\quad + \sum_{i=N+1}^{N'} \log f(\mathbf{y}_{iT}, \mathbf{z}_{iT}, \mathbf{y}_{iC}, \mathbf{z}_{iC}; \hat{m}_{ziT}, \hat{\sigma}_{ziT}^2, \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2, \eta) \end{aligned}$$

where the second term can be derived as the bivariate normal densities from (3.14). In the first term, we need to calculate the conditional expectations for the sufficient statistics of

$$\begin{aligned} &\sum_{j \in T} y_{ij}, \quad \sum_{j \in T} y_{ij}^2, \quad \sum_{j \in T} z_{ij}, \quad \sum_{j \in T} z_{ij}^2, \quad \sum_{j \in T} y_{ij} z_{ij}, \\ &\sum_{j \in C} y_{ij}, \quad \sum_{j \in C} y_{ij}^2, \quad \sum_{j \in C} z_{ij}, \quad \sum_{j \in C} z_{ij}^2, \quad \sum_{j \in C} y_{ij} z_{ij} \end{aligned}$$

for $i = 1, \dots, N$. Among them, we have

$$\begin{aligned} E_{\tilde{\eta}^{(t)}} \left[\sum_{j \in T} y_{ij} \middle| \bar{y}_{iT} \right] &= n_{iT} \bar{y}_{iT}, & E_{\tilde{\eta}^{(t)}} \left[\sum_{j \in T} y_{ij}^2 \middle| \bar{y}_{iT}, s_{yiT}^2 \right] &= n_{iT} \bar{y}_{iT}^2 + \frac{n_{iT}}{n_{iT} - 1} s_{yiT}^2, \\ E_{\tilde{\eta}^{(t)}} \left[\sum_{j \in T} z_{ij} \middle| \bar{z}_{iT} \right] &= n_{iT} \bar{z}_{iT}, & E_{\tilde{\eta}^{(t)}} \left[\sum_{j \in T} z_{ij}^2 \middle| \bar{z}_{iT}, s_{ziT}^2 \right] &= n_{iT} \bar{z}_{iT}^2 + \frac{n_{iT}}{n_{iT} - 1} s_{ziT}^2 \end{aligned}$$

and the same calculations are applied to those for $j \in C$. For the conditional expectations of $\sum_{j \in T} y_{ij} z_{ij}$ and $\sum_{j \in C} y_{ij} z_{ij}$, we here use Monte Carlo approximation. An algorithm in which the integration calculation in E-step is replaced by Monte Carlo approximation is known as Monte Carlo EM (MCEM) algorithm by Wei and Tanner (1990). The MCEM algorithm first draws B sets of $Y_{\text{miss-IPD}}$ (say $Y_{\text{miss-IPD}}^{[b]}$ for $b = 1, \dots, B$) from the conditional distributions $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}; \hat{\xi}, \tilde{\eta}^{[t]})$, where the sampling technique by Lindqvist and Taraldsen (2005) can be used in the same way as (3.20) and (3.21). Then, we have

$$\begin{aligned} & E_{\tilde{\eta}^{[t]}} \left[\sum_{j \in T} y_{ij} z_{ij} \middle| \bar{y}_{iT}, s_{yiT}^2, \bar{z}_{iT}, s_{ziT}^2 \right] \\ &= \iint \left(\sum_{j \in T} y_{ij} z_{ij} \right) f(\mathbf{y}_{iT}, \mathbf{z}_{iT} | \bar{y}_{iT}, s_{yiT}^2, \bar{z}_{iT}, s_{ziT}^2; \hat{m}_{ziT}, \hat{\sigma}_{ziT}^2, \tilde{\eta}^{[t]}) d\mathbf{y}_{iT} d\mathbf{z}_{iT} \\ &\approx \frac{1}{B} \sum_{b=1}^B \left(\sum_{j \in T} y_{ij}^{[b]} z_{ij}^{[b]} \right) \end{aligned}$$

and

$$\begin{aligned} & E_{\tilde{\eta}^{[t]}} \left[\sum_{j \in C} y_{ij} z_{ij} \middle| \bar{y}_{iC}, s_{yiC}^2, \bar{z}_{iC}, s_{ziC}^2 \right] \\ &= \iint \left(\sum_{j \in C} y_{ij} z_{ij} \right) f(\mathbf{y}_{iC}, \mathbf{z}_{iC} | \bar{y}_{iC}, s_{yiC}^2, \bar{z}_{iC}, s_{ziC}^2; \hat{m}_{ziC}, \hat{\sigma}_{ziC}^2, \tilde{\eta}^{[t]}) d\mathbf{y}_{iC} d\mathbf{z}_{iC} \\ &\approx \frac{1}{B} \sum_{b=1}^B \left(\sum_{j \in C} y_{ij}^{[b]} z_{ij}^{[b]} \right). \end{aligned}$$

In M-step, the current parameter value $\tilde{\eta}^{[t]}$ is updated by

$$\tilde{\eta}^{(t+1)} = \arg \max_{\eta} Q(\eta, \tilde{\eta}^{[t]}).$$

These iterative of E-step and M-step are repeatedly implemented until a convergence condition holds, and the final parameter value is regarded as $\tilde{\eta}$.

Step (2): Fitting IPD meta-analysis model

Step (1) produces R sets of SIPD for trials $i = 1, \dots, N$; i.e. $Y_{\text{miss-IPD}}^{[r]}$ for $r = 1, \dots, R$ in (3.17). We can now fit the IPD meta-analysis model (2.1) to each of SIPD (or each of SIPD combined with the collected IPD). This produces R sets of MLEs for parameters of interest and their variance estimates; for instance, the within-trial treatment-covariate interaction effect, $(\hat{\gamma}_W^{[r]}, V(\hat{\gamma}_W^{[r]}))$ for $r = 1, \dots, R$.

Step (3): Summarising estimates

In Step (3), resulting estimates for each set of SIPD are suitably summarised. For example, suppose that there is an interest in the posterior distribution of γ_W ; say $\pi(\gamma_W|Y_{\text{AD}})$, which is written by

$$\pi(\gamma_W|Y_{\text{AD}}) = \int \pi(\gamma_W|Y_{\text{miss-IPD}})f(Y_{\text{miss-IPD}}|Y_{\text{AD}})dY_{\text{miss-IPD}}. \quad (3.24)$$

We here consider an approximation known as Poor Man's Data Augmentation (PMDA) 2 by Wei and Tanner (1990). Given $\hat{\gamma}_W$, $(\hat{\gamma}_W^{[r]}, V(\hat{\gamma}_W^{[r]}))$ and $Y_{\text{miss-IPD}}^{[r]}$ for $r = 1, \dots, R$, the PMDA 2 approximates the posterior distribution of γ_W as follows:

$$\pi(\gamma_W|Y_{\text{AD}}) \approx \frac{\sum_{r=1}^R w_r \pi(\gamma_W|Y_{\text{miss-IPD}}^{[r]})}{\sum_{r=1}^R w_r} \quad (3.25)$$

where

$$w_r = V(\hat{\gamma}_W^{[r]})^{1/2} \frac{\pi(\hat{\gamma}_W^{[r]}|Y_{\text{miss-IPD}}^{[r]})}{\pi(\hat{\gamma}_W|Y_{\text{miss-IPD}}^{[r]})}. \quad (3.26)$$

The weights w_r for $r = 1, \dots, R$ are importance sampling weights designed to correct for the fact that one is not sampling from $f(Y_{\text{miss-IPD}}|Y_{\text{AD}})$, and PMDA 2 provides an unbiased estimate of the observed data posterior (Steele, Wang and Raftery, 2010). The derivations of (3.25) and (3.26) are detailed in Wei and Tanner (1990). To obtain point estimate (via the median of the posterior density) and 95 per cent confidence limit, one must obtain the required percentiles of the mixture distribution of (3.25). Obtaining the desired mixture percentile values (say c) requires solving

$$\sum_{r=1}^R w_r \pi(\gamma_W|Y_{\text{miss-IPD}}^{[r]}) = c$$

with respect to γ_W (Steele, Wang and Raftery, 2010). If considering $\hat{\gamma}_W = \sum_{r=1}^R \hat{\gamma}_W^{[r]}/R$ as an estimator of γ_W , the variance estimate for $\hat{\gamma}_W$ can be derived exactly, under an unrealistic assumption, by using an idea known as type B estimator by Wang and Robins (1998); this is beyond the scope of this paper.

Combining IPD and AD For the case that some trials provide IPD and the others provide only AD, the posterior distribution of γ_W ; say $\pi(\gamma_W|Y_{AD}, Y_{IPD})$ is written by

$$\pi(\gamma_W|Y_{AD}, Y_{IPD}) = \int \pi(\gamma_W|Y_{\text{miss-IPD}}, Y_{IPD}) f(Y_{\text{miss-IPD}}|Y_{AD}, Y_{IPD}) dY_{\text{miss-IPD}}. \quad (3.27)$$

In the similar way as (3.25), the PMDA 2 approximates the posterior distribution of γ_W as follows:

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

where

$$w_r = V(\hat{\gamma}_W^{[r]})^{1/2} \frac{\pi(\hat{\gamma}_W^{[r]}|Y_{\text{miss-IPD}}^{[r]}, Y_{IPD})}{\pi(\tilde{\gamma}_W|Y_{\text{miss-IPD}}^{[r]}, Y_{IPD})}.$$

SIPD method via Bayesian procedure

Step (1): Generating SIPD

Let $f(Y_{\text{miss-IPD}}|\hat{\xi}, \eta)$ be a Bayesian density function of the missing IPD given parameter η , which has the same form as $f(Y_{\text{miss-IPD}}|\hat{\xi}, \eta)$. Then, we draw the SIPD (say $Y_{\text{miss-IPD}}^*$) from the posterior predictive distribution of $Y_{\text{miss-IPD}}$ given Y_{AD} ; i.e.

$$Y_{\text{miss-IPD}}^* \sim f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}) \quad (3.28)$$

where

$$f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}) = \int f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}, \eta) f(\eta|Y_{AD}, \hat{\xi}) d\eta \quad (3.29)$$

and $f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}, \eta)$ is the density of $Y_{\text{miss-IPD}}$ given Y_{AD} and η , $f(\eta|Y_{AD}, \hat{\xi})$ is the posterior distribution of η given Y_{AD} . Because of the integration in (3.29), $f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi})$ cannot be expressed in a closed form. And also, it is difficult to draw samples from this distribution directly; however once obtaining samples of the parameter η , we can achieve the drawing (3.28) approximately. If R sets of the parameter values (say $\eta^{[r]}$ for $r = 1, \dots, R$) are drawn from the posterior distribution $f(\eta|Y_{AD}, \hat{\xi})$, then the posterior predictive distribution (3.29) can be approximated as follows:

$$f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}) \approx \frac{1}{R} \sum_{r=1}^R f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}, \eta^{[r]}) \quad (3.30)$$

This indicates that one random sample from $f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}, \eta^{[r]})$ corresponds to one random sample from $f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi})$, and the repetition of this drawing yields R sets of

SIPD (say $Y_{\text{miss-IPD}}^{[r]}$ for $r = 1, \dots, R$). The SIPD consist of the patient-specific observations given as (3.17). The approximation (3.30) requires one to draw from $f(\eta|Y_{\text{AD}}, \hat{\xi})$ and then draw from $f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, \hat{\xi}, \eta^{[r]})$. The latter can be implemented by the same procedure for $f(Y_{\text{miss-IPD}}|Y_{\text{AD}}; \hat{\xi}, \tilde{\eta})$. We now describe how to draw from $f(\eta|Y_{\text{AD}}, \hat{\xi})$.

Drawing from $f(\eta|Y_{\text{AD}}, \hat{\xi})$: Drawing samples of η from $f(\eta|Y_{\text{AD}}, \hat{\xi})$ is straightforward to achieve by Markov chain Monte Carlo (MCMC) method; in particular we use Metropolis-Hastings algorithm. The posterior distribution of η can be written as

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

where $f(\eta)$ is the density function for a prior distribution of η and we use a vague prior for this; i.e. $f(\eta) \propto \sigma_y^{-1}$. Then, $f(Y_{\text{AD}}|\hat{\xi}, \eta)$ has the same form as $f(Y_{\text{AD}}|\hat{\xi}, \eta)$ in (3.16). Therefore, for the purpose of drawing $\eta^{[r]}$ from $f(\eta|Y_{\text{AD}}, \hat{\xi})$, the Metropolis-Hastings algorithm takes the following procedures (Gelman *et al.*, 1995):

1. Set a starting value $\eta^{[0]}$, and iterate Step 2-4 for $r = 1, \dots, R$.
2. Draw a sample η^* from a proposal distribution with density function $\rho(\eta|\eta^{[r-1]})$; i.e.

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

We here assume $\rho(\eta|\eta^{[r-1]})$ is a normal density centered at $\eta^{[r-1]}$.

3. Compute

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

4. Set $\eta^{[r]} = \eta^*$ with probability a , otherwise set $\eta^{[r]} = \eta^{[r-1]}$.

For rapid convergence, we integrate out the parameter of trial-specific effects $(\phi_1, \dots, \phi_N)$ from $f(\eta|Y_{\text{AD}}, \hat{\xi})$, and then consider to draw $(\theta, \mu, \gamma_{\text{A}}, \gamma_{\text{W}})$ from their marginal posterior distribution. This is because we need only the values of parameter associated with the within-trial relationships (μ and γ_{W}) to generate the SIPD.

Combining IPD and AD For the case that some trials provide IPD and the others provide only AD, we draw the SIPD (say $Y_{\text{miss-IPD}}^*$) from the posterior predictive distribution of $Y_{\text{miss-IPD}}$ given Y_{AD} and Y_{IPD} ; i.e.

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

where

$$f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}) = \int f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, \hat{\xi}, \eta) f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}) d\eta \quad (3.31)$$

and $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ is the posterior distribution of η given Y_{AD} . By the same reason described above, it is necessary to first sample the parameter values of η from $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$, and then draw $Y_{\text{miss-IPD}}$ from $f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, \hat{\xi}, \eta^{[r]})$. If R sets of the parameter values ($\eta^{[r]}$ for $r = 1, \dots, R$) are drawn from the posterior distribution $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$, the posterior predictive distribution (3.31) can be approximated as follows:

$$f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}) \approx \frac{1}{R} \sum_{r=1}^R f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, \hat{\xi}, \eta^{[r]}) \quad (3.32)$$

This indicates that one random sample from $f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, \hat{\xi}, \eta^{[r]})$ corresponds to one random sample from $f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$, and the repetition of this drawing yields R sets of SIPD ($Y_{\text{miss-IPD}}^{[r]}$ for $r = 1, \dots, R$). The approximation (3.32) requires one to draw from $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ and then draw from $f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, \hat{\xi}, \eta^{[r]})$. The latter can be implemented by the same procedure for $f(Y_{\text{miss-IPD}}|Y_{\text{AD}}, \hat{\xi}, \tilde{\eta})$. We now describe how to draw from $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$.

Drawing from $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$: As in drawing from $f(\eta|Y_{\text{AD}}, \hat{\xi})$, drawing samples of η from $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ is straightforward to achieve by MCMC method (Metropolis-Hastings algorithm). The posterior distribution of η can be written as

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

where $f(\eta)$ is the vague prior of $f(\eta) \propto \sigma_y^{-1}$. Because of between-trial independence, we have

$$f(Y_{\text{AD}}, Y_{\text{IPD}}|\hat{\xi}, \eta) = f(Y_{\text{AD}}|\hat{\xi}, \eta) f(Y_{\text{IPD}}|\hat{\xi}, \eta).$$

where $f(Y_{\text{AD}}|\hat{\xi}, \eta)$ has the same form as $f(Y_{\text{AD}}|\hat{\xi}, \eta)$ in (3.16), and $f(Y_{\text{IPD}}|\hat{\xi}, \eta)$ is derived by the same form as $f(Y_{\text{IPD}}|\hat{\xi}, \eta)$ in (3.23). From these results, for the purpose of drawing $\eta^{[r]}$ from $f(\eta|Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$, the Metropolis-Hastings algorithm takes the following procedures:

1. Set a starting value $\eta^{[0]}$, and iterate Step 2-4 for $r = 1, \dots, R$.
2. Draw a sample η^* from a proposal distribution with density function $\rho(\eta|\eta^{[r-1]})$; i.e.

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

We here assume $\rho(\eta|\eta^{[r-1]})$ is a normal density centered at $\eta^{[r-1]}$.

3. Compute

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

4. Set $\eta^{[r]} = \eta^*$ with probability a , otherwise set $\eta^{[r]} = \eta^{[r-1]}$.

For the same reason described above, we integrate out the parameter of trial-specific effects $(\phi_1, \dots, \phi_{N'})$ from $f(\eta|Y_{AD}, Y_{IPD}, \hat{\xi})$, and then consider to draw $(\theta, \mu, \gamma_A, \gamma_W)$ from their marginal posterior distribution.

As is the case for the frequentist procedure, the fact that we can use the IPD from a part of trials offers another solution to draw from $f(Y_{\text{miss-IPD}}|Y_{AD}, Y_{IPD}, \hat{\xi})$, which is known as data augmentation by Tanner and Wong (1987).

Data augmentation The data augmentation is implemented by the following iterative steps in terms of $r = 1, \dots, R$.

1. Choose a moderate positive integer B , and create draws as follows:

$$\eta^{[r,b]} \sim f_r(\eta|Y_{AD}, Y_{IPD}, \hat{\xi}), \quad Y_{\text{miss-IPD}}^{[r,b]} \sim f(Y_{\text{miss-IPD}}|Y_{AD}, \hat{\xi}, \eta^{[r,b]}). \quad (3.33)$$

for $b = 1, \dots, B$. Here, $f_k(\eta|Y_{AD}, Y_{IPD}, \hat{\xi})$ denotes an approximate posterior distribution of η at iteration r , which is computed at iteration $r - 1$.

2. Update the approximate posterior distribution of η as follows:

$$f_{r+1}(\eta|Y_{AD}, Y_{IPD}, \hat{\xi}) = \frac{1}{B} \sum_{b=1}^B f(\eta|Y_{\text{miss-IPD}}^{[r,b]}, Y_{IPD}, \hat{\xi}).$$

The sequence of draws for $Y_{\text{miss-IPD}}$ and η from this iterative procedure is known to converge to a draw from $f(Y_{\text{miss-IPD}}, \eta|Y_{AD}, Y_{IPD}, \hat{\xi})$ (Little and Rubin, 2002). This is motivated by the fact that the approximate posterior distribution of η in (3.33), $f_r(\eta|Y_{AD}, Y_{IPD}, \hat{\xi})$, is easier to draw from than $f(\eta|Y_{AD}, Y_{IPD}, \hat{\xi})$. Indeed, each element composing $f_r(\eta|Y_{AD}, Y_{IPD}, \hat{\xi})$; i.e. $f(\eta|Y_{\text{miss-IPD}}^{[r,b]}, Y_{IPD}, \hat{\xi})$ for $b = 1, \dots, B$, can be derived exactly (Gelman *et al.* 1995).

More specifically, we rewrite the IPD meta-analysis model (3.9) by using a matrix form of

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\eta}_0 + \boldsymbol{\epsilon}, \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \sigma_y^2 \mathbf{I}). \end{aligned}$$

Letting $n = \sum_{i=1}^{N'} n_i$ be the total number of patients, $\mathbf{y}$ is an n -dimensional vector of y_{ij} for $i = 1, \dots, N'$ and $j = 1, \dots, n_i$, $\mathbf{X}$ is an $n \times (N' + 4)$ design matrix, $\mathbf{0}$ is an n -dimensional

zero vector, $\mathbf{I}$ is an $n \times n$ identity matrix and $\eta_0 = (\phi_1, \dots, \phi_{N'}, \theta, \mu, \gamma_A, \gamma_W)$. If we again assume the vague prior of $f(\eta_0, \sigma_y^2) \propto \sigma_y^{-2}$, then we have the posterior distribution of (η_0, σ_y^2) given $Y_{\text{miss-IPD}}$ and Y_{IPD} , $f(\eta_0, \sigma_y^2 | Y_{\text{miss-IPD}}, Y_{\text{IPD}}, \hat{\xi})$, as follows:

$$\begin{aligned}\eta_0 | \sigma_y^2, Y_{\text{miss-IPD}}, Y_{\text{IPD}} &\sim N(\hat{\eta}_0, \sigma_y^2), \\ \sigma_y^2 | Y_{\text{miss-IPD}}, Y_{\text{IPD}} &\sim \text{Inv-}\chi^2(n - N' - 4, s^2)\end{aligned}$$

where $\text{Inv-}\chi^2(n - N' - 4, s^2)$ denotes a random variable from a scaled inverse chi-square distribution with scale parameter s^2 and degrees of freedom $n - N' - 4$, and each parameter is given by

$$\hat{\eta}_0 = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3.34)$$

and

$$s^2 = \frac{1}{n - N' - 4} (\mathbf{y} - \mathbf{X} \hat{\eta}_0)^T (\mathbf{y} - \mathbf{X} \hat{\eta}_0). \quad (3.35)$$

In the actual iterative procedures, the current value of $Y_{\text{miss-IPD}}$ is substituted into $(\mathbf{y}, \mathbf{X})$ in (3.34) and (3.35).

Step (2): Fitting IPD meta-analysis model

Step (1) produces R sets of SIPD for trials $i = 1, \dots, N$; i.e. $Y_{\text{miss-IPD}}^{[r]}$ for $r = 1, \dots, R$ in (3.17). We can now fit the IPD meta-analysis model (2.1) to each of SIPD (or each of SIPD combined with the collected IPD). This produces R sets of MLEs for parameters of interest and their variance estimates; for instance, the within-trial treatment-covariate interaction effect, $(\hat{\gamma}_W^{[r]}, V(\hat{\gamma}_W^{[r]}))$ for $r = 1, \dots, R$.

Step (3): Summarising estimates

As is the case for the frequentist procedure, we here consider the posterior distribution of γ_W written by (3.24). From the Bayesian perspective, the posterior distribution of γ_W can be simulated by first drawing $Y_{\text{miss-IPD}}^{[r]}$ from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}})$, and then drawing γ_W from $\pi(\gamma_W | Y_{\text{miss-IPD}}^{[r]})$; i.e.

$$\pi(\gamma_W | Y_{\text{AD}}) \approx \frac{1}{R} \sum_{r=1}^R \pi(\gamma_W | Y_{\text{miss-IPD}}^{[r]}). \quad (3.36)$$

We apply a Rubin's (1987) combining rule in order to obtain a posterior mean and variance for γ_W , which is often used in multiple imputation. The rule approximates them as follows

(Little and Rubin, 2002):

$$\begin{aligned} E(\gamma_W|Y_{AD}) &= E[E(\gamma_W|Y_{\text{miss-IPD}})|Y_{AD}] \\ &\approx \frac{1}{R} \sum_{r=1}^R \hat{\gamma}_W^{[r]} \end{aligned} \quad (3.37)$$

and

$$\begin{aligned} \text{Var}(\gamma_W|Y_{AD}) &= E[\text{Var}(\gamma_W|Y_{\text{miss-IPD}})|Y_{AD}] + \text{Var}[E(\gamma_W|Y_{\text{miss-IPD}})|Y_{AD}] \\ &\approx \frac{1}{R} \sum_{r=1}^R V(\hat{\gamma}_W^{[r]}) + \frac{1+R^{-1}}{R-1} \sum_{r=1}^R (\hat{\gamma}_W^{[r]} - \hat{\gamma}_W)^2 \end{aligned} \quad (3.38)$$

where $\hat{\gamma}_W$ is given by (3.37). For frequentist inferences, we use (3.37) as an overall estimate $\hat{\gamma}_W$, and (3.38) as its variance estimate $V(\hat{\gamma}_W)$. If considering $\hat{\gamma}_W = \sum_{r=1}^R \hat{\gamma}_W^{[r]}/R$ as an estimator of γ_W , the variance estimate for $\hat{\gamma}_W$ can be derived exactly by using an idea known as type A estimator by Wang and Robins (1998).

Combining IPD and AD For the case that some trials provide IPD and the others provide only AD, the posterior distribution of γ_W is written by (3.27) and can be approximated by using $Y_{\text{miss-IPD}}^{[r]}$ for $r = 1, \dots, R$:

$$\pi(\gamma_W|Y_{AD}, Y_{IPD}) \approx \frac{1}{R} \sum_{r=1}^R \pi(\gamma_W|Y_{\text{miss-IPD}}^{[r]}, Y_{IPD}). \quad (3.39)$$

Then, as in (3.37) and (3.38), the Rubin's (1987) rule approximates a posterior mean and variance for γ_W as follows:

$$\begin{aligned} E(\gamma_W|Y_{AD}, Y_{IPD}) &= E[E(\gamma_W|Y_{\text{miss-IPD}}, Y_{IPD})|Y_{AD}, Y_{IPD}] \\ &\approx \frac{1}{R} \sum_{r=1}^R \hat{\gamma}_W^{[r]} \end{aligned} \quad (3.40)$$

and

$$\begin{aligned} \text{Var}(\gamma_W|Y_{AD}, Y_{IPD}) &= E[\text{Var}(\gamma_W|Y_{\text{miss-IPD}}, Y_{IPD})|Y_{AD}, Y_{IPD}] \\ &\quad + \text{Var}[E(\gamma_W|Y_{\text{miss-IPD}}, Y_{IPD})|Y_{AD}, Y_{IPD}] \\ &\approx \frac{1}{R} \sum_{r=1}^R V(\hat{\gamma}_W^{[r]}) + \frac{1+R^{-1}}{R-1} \sum_{r=1}^R (\hat{\gamma}_W^{[r]} - \hat{\gamma}_W)^2 \end{aligned} \quad (3.41)$$

where $\hat{\gamma}_W$ is given by (3.40). For frequentist inferences, we use (3.40) as an overall estimate $\hat{\gamma}_W$, and (3.41) as its variance estimate $V(\hat{\gamma}_W)$.

Table V. The grouped form of data from group k in the i th study.

Z	$Y = 0$	$Y = 1$
0		m_{0ik} n_{0ik}
1		m_{1ik} n_{1ik}
	$n_{ik} - m_{ik}$	m_{ik} n_{ik}

3.2.2 Binary outcome and covariate

We here consider the SIPD method only from the Bayesian perspective, for the case that some trials provide IPD and the others provide only AD.

SIPD method for combining IPD and AD via Bayesian procedure

Step (1): Generating SIPD

Assume that each AD trial is completely balanced at the patient-level covariate; so that the proportion of patients with $Z = 1$ for the treatment group is assumed to be equivalent to that for the control group. Letting $n_{1ik} = [n_{ik}\bar{z}_i + 0.5]$ ($[A]$ is the largest integer not greater than A) and $n_{0ik} = n_{ik} - n_{1ik}$ be the number of patients respectively with $Z = 1$ and $Z = 0$ for group k in the i th AD trial ($i = 1, \dots, N$), then the AD given by $(m_{ik}, n_{0ik}, n_{1ik})$ can be written as marginal totals of a 2×2 contingency table in Table III, where the internal cells m_{0ik} and m_{1ik} representing the number of patients respectively with $(Y = 1, Z = 0)$ and $(Y = 1, Z = 1)$ are not available. Now, for all trials $i = 1, \dots, N'$, we assume that m_{0ik} and m_{1ik} follow a pair of independent binomial distributions respectively with probabilities $p_{0ik} = \Pr(Y = 1|Z = 0, i, k)$ and $p_{1ik} = \Pr(Y = 1|Z = 1, i, k)$; i.e. $m_{0ik} \sim \text{Binominal}(n_{0ik}, p_{0ik})$ and $m_{1ik} \sim \text{Binominal}(n_{1ik}, p_{1ik})$. Then, we draw the SIPD for group k in the i th AD trial; i.e. m_{0ik} (and $m_{1ik} = m_{ik} - m_{0ik}$), from the following posterior predictive distribution:

$$\begin{aligned} & \Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}) \\ &= \int \Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, p_{0ik}, p_{1ik}) f(p_{0ik}, p_{1ik}|m_{ik}, n_{0ik}, n_{1ik}) dp_{0ik} dp_{1ik}. \end{aligned} \quad (3.42)$$

for $i = 1, \dots, N$ and $k \in \{T, C\}$. Here, $f(p_{0ik}, p_{1ik}|m_{ik}, n_{0ik}, n_{1ik})$ is the density function for the posterior distribution of (p_{0ik}, p_{1ik}) given m_{ik} , and is written as follows:

$$f(p_{0ik}, p_{1ik}|m_{ik}, n_{0ik}, n_{1ik}) \propto \Pr(m_{ik}|n_{0ik}, n_{1ik}, p_{0ik}, p_{1ik}) f(p_{0ik}, p_{1ik}) \quad (3.43)$$

where $f(p_{0ik}, p_{1ik})$ is a prior distribution of (p_{0ik}, p_{1ik}) . $\Pr(m_{ik}|n_{0ik}, n_{1ik}, p_{0ik}, p_{1ik})$ is the

probability of a marginal total $m_{ik} = m_{0ik} + m_{1ik}$ in Table V with unknown internal cells, which can be derived as convolution of the binomial distributions for m_{0ik} and m_{1ik} :

$$\begin{aligned} & \Pr(m_{ik}|n_{0ik}, n_{1ik}, p_{0ik}, p_{1ik}) \\ &= \sum_{m_{0ik}=l_{ik}}^{u_{ik}} \binom{n_{0ik}}{m_{0ik}} p_{0ik}^{m_{0ik}} (1-p_{0ik})^{n_{0ik}-m_{0ik}} \\ & \quad \times \binom{n_{1ik}}{m_{ik}-m_{0ik}} p_{1ik}^{m_{ik}-m_{0ik}} (1-p_{1ik})^{n_{1ik}-m_{ik}+m_{0ik}} \end{aligned} \quad (3.44)$$

where

$$l_{ik} = \max(0, m_{ik} - n_{1ik}), \quad u_{ik} = \min(n_{0ik}, m_{ik}) \quad (3.45)$$

represent the range of admissible values of m_{0ik} so that $m_{ik} = m_{0ik} + m_{1ik}$ and $n_{ik} - m_{ik} = n_{0ik} - m_{0ik} + n_{1ik} - m_{1ik}$ are satisfied. And also, $\Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, p_{0ik}, p_{1ik})$ in (3.42) is the probability of an internal cell given all the marginal totals in Table V. If the single internal cell m_{0ik} are drawn, the other internal cells can be identified uniquely as $m_{1ik} = m_{ik} - m_{0ik}$, $n_{0ik} - m_{0ik}$ and $n_{1ik} - m_{1ik}$. Letting

$$\lambda_{0ik} = p_{0ik}/(1-p_{0ik}), \quad \lambda_{1ik} = p_{1ik}/(1-p_{1ik}) \quad (3.46)$$

be odds of p_{0ik} and p_{1ik} respectively, $\Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, p_{0ik}, p_{1ik})$ can be derived as the probability mass function of Fisher's non-central hypergeometric distribution (or extended hypergeometric distribution) with parameter of odds ratio $\lambda_{ik} = \lambda_{0ik}/\lambda_{1ik}$:

$$\Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, \lambda_{ik}) = \frac{\binom{n_{0ik}}{m_{0ik}} \binom{n_{1ik}}{m_{ik}-m_{0ik}} \lambda_{ik}^{m_{0ik}}}{\sum_{m_0=l_{ik}}^{u_{ik}} \binom{n_{0ik}}{m_0} \binom{n_{1ik}}{m_{ik}-m_0} \lambda_{ik}^{m_0}} \quad (3.47)$$

where l_{ik} and u_{ik} are defined as (3.45).

Because of the integration in (3.42), the posterior predictive distribution of m_{0ik} cannot be expressed in a closed form. Also, it is difficult to draw samples from this distribution directly; however once obtaining samples of parameter (p_{0ik}, p_{1ik}) , we can draw m_{0ik} approximately from (3.42). If R sets of parameter values (say $(p_{0ik}^{[r]}, p_{1ik}^{[r]})$ for $r = 1, \dots, R$) are drawn from the posterior distribution (3.43), then the posterior predictive distribution (3.42) can be approximated as

$$\Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}) \approx \frac{1}{R} \sum_{r=1}^R \Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, p_{0ik}^{[r]}, p_{1ik}^{[r]}) \quad (3.48)$$

where the r th component in summation is the conditional probability of m_{0ik} given m_{ik} and a known parameter of $(p_{0ik}^{[r]}, p_{1ik}^{[r]})$. This indicates that a single random sample from $\Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, p_{0ik}^{[r]}, p_{1ik}^{[r]})$ corresponds to a single random sample from the posterior predictive distribution (3.42), and the repetition of this drawing yields R sets of SIPD (say $(m_{0ik}^{[r]}, m_{1ik}^{[r]})$ for $r = 1, \dots, R$). Therefore, we generate the simulated IPD for the AD trials by using the approximation (3.48), that requires us to draw from $f(p_{0ik}, p_{1ik}|m_{ik}, n_{0ik}, n_{1ik})$ and then draw from $\Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, p_{0ik}^{[r]}, p_{1ik}^{[r]})$. We now describe how to draw from these distributions in more detail.

Drawing from $f(p_{0ik}, p_{1ik}|m_{ik}, n_{0ik}, n_{1ik})$

Now, let

$$\begin{aligned} Y_{AD} &= \{(m_{ik}, n_{0ik}, n_{1ik}) : i = 1, \dots, N, k \in \{T, C\}\}, \\ Y_{IPD} &= \{(m_{0ik}, m_{1ik}, n_{0ik}, n_{1ik}) : i = N+1, \dots, N', k \in \{T, C\}\} \end{aligned}$$

where Y_{AD} is the AD summarised from the uncollected IPD for trials $i = 1, \dots, N$, and Y_{IPD} is the collected IPD for trials $i = N+1, \dots, N'$. Here, we assume the following model for the logits of p_{0ik} and p_{1ik} ; i.e. the logarithm of (3.46):

$$\begin{aligned} \log \lambda_{0iC} &= \alpha_{0C}, \\ \log \lambda_{1iC} &= \alpha_{1C}, \\ \log \lambda_{0iT} &= \alpha_{0T} + \beta \bar{z}_i, \\ \log \lambda_{1iT} &= \alpha_{1T} + \beta \bar{z}_i \end{aligned} \tag{3.49}$$

We now consider to draw parameters in model (3.49), and then produce the values of p_{0ik} and p_{1ik} by using model (3.49). Letting

$$\eta = (\alpha_{0C}, \alpha_{1C}, \alpha_{0T}, \alpha_{1T}, \beta)$$

be a parameter vector to be drawn, we have

$$\begin{aligned} f(\eta|Y_{AD}, Y_{IPD}) &\propto f(\eta)f(Y_{AD}|\eta)f(Y_{IPD}|\eta) \\ &= f(\eta) \prod_{i=1}^N \prod_{k \in \{T, C\}} \Pr(m_{ik}|n_{0ik}, n_{1ik}, \eta) \\ &\quad \times \prod_{i=N+1}^{N'} \prod_{k \in \{T, C\}} \Pr(m_{0ik}, m_{1ik}|n_{0ik}, n_{1ik}, \eta) \end{aligned}$$

where $f(\eta)$ is the density function for a prior distribution of η and we use a vague prior for this; i.e. $f(\eta) \propto 1$. $\Pr(m_{ik}|n_{0ik}, n_{1ik}, \eta)$ is given as (3.44), and $\Pr(m_{0ik}, m_{1ik}|n_{0ik}, n_{1ik}, \eta)$ can be derived by using probability mass functions of Binomial distribution for m_{0ik} and m_{1ik} ; i.e.

$$\begin{aligned} & \Pr(m_{0ik}, m_{1ik}|n_{0ik}, n_{1ik}, \eta) \\ &= \binom{n_{0ik}}{m_{0ik}} \binom{n_{1ik}}{m_{1ik}} p_{0ik}^{m_{0ik}} (1 - p_{0ik})^{n_{0ik} - m_{0ik}} p_{1ik}^{m_{1ik}} (1 - p_{1ik})^{n_{1ik} - m_{1ik}} \end{aligned} \quad (3.50)$$

where the corresponding components in η are suitably substituted into p_{0ik} , p_{1ik} , λ_{0ik} and λ_{1ik} .

Drawing from $f(\eta|Y_{AD}, Y_{IPD})$ is straightforward to achieve by MCMC method; in particular Metropolis-Hastings algorithm. By using (3.44) and (3.50), the Metropolis-Hastings algorithm allows one to draw R sets of η from $f(\eta|Y_{AD}, Y_{IPD})$; i.e. $\eta^{[r]}$ for $r = 1, \dots, R$, which takes the the similar implementing procedure described above. The r th set of parameter value, $\eta^{[r]}$, is transformed into $(\lambda_{0ik}^{[r]}, \lambda_{1ik}^{[r]})$ and then $(p_{0ik}^{[r]}, p_{1ik}^{[r]})$ uniquely.

Drawing from $\Pr(m_{0ik}|m_{ik}, n_{0ik}, n_{1ik}, p_{0ik}^{[r]}, p_{1ik}^{[r]})$: This drawing is equivalent to the drawing from the Fisher's non-central hypergeometric distribution with a known parameter $(p_{0ik}^{[r]}, p_{1ik}^{[r]})$. We can draw R sets of SIPD from this conditional distribution directly. Fog (2008) supposed a fast algorithm to draw from Fisher's non-central hypergeometric distribution. We get R sets of SIPD (say $(m_{0ik}^{[r]}, m_{1ik}^{[r]})$ for $r = 1, \dots, R$), which are transformed to the binary data form of SIPD $(y_{ij}, x_{ij}, z_{ij}^{[r]})$ for $j = 1, \dots, n_i$ and $r = 1, \dots, R$.

Step (2): Fitting IPD meta-analysis model

Step (1) produces R sets of SIPD for AD trials; i.e. $(y_{ij}, x_{ij}, z_{ij}^{[r]})$ for $i = 1, \dots, N$, $j = 1, \dots, n_i$ and $r = 1, \dots, R$. We can now fit the IPD meta-analysis model (2.6) to each of SIPD with the collected IPD. This produces R sets of MLEs for parameters of interest and their variance estimates; for instance, the within-trial treatment-covariate interaction effect, $(\hat{\gamma}_W^{[r]}, V(\hat{\gamma}_W^{[r]}))$ for $r = 1, \dots, R$.

Step (3): Summarising estimates

When there is an interest in the posterior distribution of γ_W , we can use the Rubin's (1987) rule as described above. The rule approximates a posterior mean and variance for γ_W as

follows:

$$E(\gamma_W|Y_{AD}, Y_{IPD}) \approx \frac{1}{R} \sum_{r=1}^R \hat{\gamma}_W^{[r]} \quad (3.51)$$

and

$$\text{Var}(\gamma_W|Y_{AD}, Y_{IPD}) \approx \frac{1}{R} \sum_{r=1}^R V(\hat{\gamma}_W^{[r]}) + \frac{1+R^{-1}}{R-1} \sum_{r=1}^R (\hat{\gamma}_W^{[r]} - \hat{\gamma}_W)^2. \quad (3.52)$$

where $\hat{\gamma}_W$ is given by (3.51). For frequentist inferences, we use (3.51) as an overall estimate $\hat{\gamma}_W$, and (3.52) as its variance estimate $V(\hat{\gamma}_W)$.

4 Application and simulation studies

4.1 Introduction

In this chapter, it is shown that the proposed methods described in Chapter 3 (method with marginalising the missing IPD, and method based on simulated IPD) has many benefits for inference of the treatment-covariate interaction. Especially for the within-trial relationships between individual observations, the proposed methods work substantially better than the existing approaches described in Chapter 2. Moreover, it is suggested that the proposed methods could have a huge possibility to produce novel findings. We now outline objectives and results of each experiment.

Simulation 1 In Chapter 4.2, we verify the performance of the proposed method with marginalising the missing IPD in the case for only AD. We suppose the case where all trials provide only AD, and compare estimates of the across-trial and the within-trial interaction from the proposed method with those from a full IPD analysis using the original IPD from all trials. When variation in within-trial covariate distributions is small, the proposed method provides accurate within-trial interaction effect estimates. This indicates that the proposed method has a potential advantage to inference of the within-trial interaction, that is never achieved by the existing approach.

Simulation 2 In Chapter 4.3, we verify the performance of the proposed method with marginalising the missing IPD in the case for mixture of IPD and AD. We suppose the case where some trials provide IPD and the others provide only AD, and compare estimates of the across-trial and the within-trial interaction from the proposed method (and model (2.5)) with those from a full IPD analysis using the original IPD from all trials. When the number of trials providing IPD is small (e.g. 1 or 2 of 20 trials), the proposed method provides more accurate within-trial interaction effect estimates than model (2.5). If meta-analysts consider

how many IPD trials should be collected to preserve a desired level of statistical power, the proposed method requires them to collect smaller number of IPD trials than the existing model (2.5).

Application to hypertension data In Chapter 4.4, we illustrate the SIPD method via Bayesian procedure through an application to the hypertension data. We are again interested in how age modifies the treatment effect on change in SBP (follow-up minus baseline), and demonstrate how the SIPD method produces the across-trial and the within-trial interaction effect estimates in the case for mixture of IPD and AD. The main gain from the SIPD method is to improve the standard error of the within-trial interaction effect estimate in comparison with the existing model (2.5), especially when the number of trials providing IPD is small. A potential benefit of using SIPD is also discussed in the context of a subgroup meta-analysis which is never conducted by using model (2.5).

Simulation 3 and Simulation 4 In Chapter 4.5 and Chapter 4.6, we verify the observed performance of the SIPD method via Bayesian procedure in two situations: one is the situation where a single continuous outcome and covariate are observed from each patient (Simulation 3), and the other is the situation where a single binary outcome and covariate are observed from each patient (Simulation 4). In particular, it is ensured that the SIPD method provides more accurate within-trial interaction effect estimates than the existing model (2.5) or model (2.9).

4.2 Simulation 1: Performance of the proposed method with marginalising the missing IPD in the case for only AD

We here supposed that all trials provide only AD, and focused on the treatment-covariate interaction estimated by fitting the MR model (2.2) and the method with marginalising the missing IPD described in Chapter 3.1. The MR model (2.2) can be used only for the inference of the across-trial interaction (β); whereas the proposed method allows one to estimate both the across-trial and the within-trial interaction (γ_A and γ_W). Some practical differences between γ_A and γ_W are as illustrated in Chapter 2, that highlighting the pitfall of using $\hat{\gamma}_A$ to make inference about $\hat{\gamma}_W$. In particular, we were interested in how the log-likelihood (3.8) computed by using only the AD available recovered the information on the within-trial relationships.

4.2.1 Design of Simulation 1

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

$$\begin{aligned}
 x_{ij} &= \begin{cases} 0, & j \in C \\ 1, & j \in T \end{cases}, \\
 m_{zi} &\sim N(30, 10), \\
 z_{ij}|m_{zi} &\sim N(m_{zi}, \sigma_z^2), \\
 y_{ij}|x_{ij}, z_{ij}, \bar{z}_i &\sim N(20 + 5x_{ij} + 0.05z_{ij} + 0.05x_{ij}\bar{z}_i + 0.1x_{ij}(z_{ij} - \bar{z}_i), \sigma_y^2), \\
 j &= 1, \dots, 800, \quad i = 1, \dots, 10
 \end{aligned} \tag{4.1}$$

where the numbers of patients for the treatment group in each trial were assumed to be equivalent to those for the control group; i.e. $n_{iT} = n_{iC} = 400$ for $i = 1, \dots, 10$. In the IPD meta-analysis model (2.1), inference of the within-trial interaction is mainly affected by: 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 . The standard errors of the within-trial interaction effect estimates are expected to become smaller as σ_z^2 increases and σ_y^2 decreases (Simmons and Higgins, 2007). We here controlled these parameters by the following scenarios: $\sigma_z^2 \in \{20, 40, 80\}$ and $\sigma_y^2 \in \{10, 20, 40\}$.

The implementing procedure was as follows. Firstly, we set parameters of σ_z^2 and σ_y^2 among 9 scenarios, and then according to the true model (4.1) with parameters set in the previous step, we generated 10,000 sets of meta-analysis data. More specifically, we generated mean covariate for the i th trial, m_{zi} , from $N(30, 10)$ and covariate values for patients in the i th trial from $N(m_{zi}, \sigma_z^2)$ given m_{zi} and σ_z^2 , and then outcome values for patients given covariate values. Secondly, we summarised the IPD from all 10 trials to the AD represented as sample means and sample variances of individual observations in each group and trial. Finally, we meta-analysed the AD by: (i) fitting the MR model (2.2), and (ii) applying the proposed method. In each analysis, we computed estimates of β and their root mean square errors (RMSEs) from the MR model (2.2), and those of γ_W and γ_A from the proposed method. The RMSE for $\hat{\beta}$ was computed by using $\gamma_A = 0.05$ as its true value. These results were compared with the results from the full IPD analysis.

4.2.2 Results of Simulation 1

Table VI shows RMSEs and mean biases for the across-trial interaction in each scenario. Fitting the MR model (2.2) to the AD from all trials naturally provided estimates only of the across-trial interaction, $\hat{\beta}$, whose RMSEs and mean biases were equivalent to those for $\hat{\gamma}_A$ from the full IPD analysis. However, as mentioned above, the estimates of the across-trial interaction must be interpreted differently from those of the within-trial interaction. For example, in a scenario of $\sigma_y^2 = 40$ and $\sigma_z^2 = 20$, Figure 7 shows scatter plots of Z-values for $\hat{\beta}$ from the MR model (2.2) against Z-values for $\hat{\gamma}_A$ (panel on the left side) and $\hat{\gamma}_W$ (panel on the right side) from the full IPD analysis. The vertical and horizontal lines represent Z-values of 1.69 (i.e. the division between statistical significance and non-significance of a one-side hypothesis test at 5 per cent level for $H_0 : \beta = 0$ or $H_0 : \gamma_W = 0$). Obviously from Figure 7, estimates of β and their standard errors from the MR model (2.2) were equivalent to those of γ_A from the full IPD analysis. For 10,000 sets of meta-analysis data, 99.7 per cent of the full IPD analyses provided significant results for $\hat{\gamma}_W$, while only 26.3 per cent of analyses by the MR model (2.2) were significant for $\hat{\beta}$. And also, for 73.4 per cent of analyses, $\hat{\gamma}_W$'s from the full IPD analysis were significant and $\hat{\beta}$'s from the MR model (2.2) were non-significant,

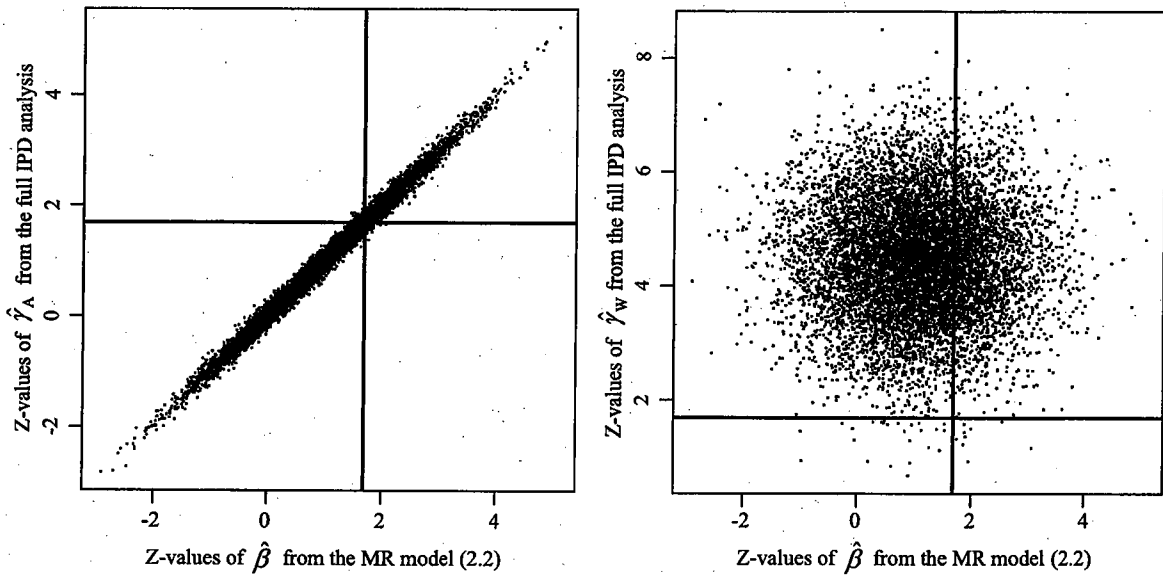


Figure 7. Scatter plots of Z-values of $\hat{\beta}$ from the MR model (2.2) against Z-values of $\hat{\gamma}_A$ (panel on the left side) and $\hat{\gamma}_W$ (panel on the right side) from meta-analyses of IPD from all 10 trials.

Table VI. Root mean square errors and mean biases for estimator of across-trial treatment-covariate interaction effect.

σ_y^2	σ_z^2	Root mean square error of $\hat{\beta}$ or $\hat{\gamma}_A$			Mean bias of $\hat{\beta}$ or $\hat{\gamma}_A$		
		MR	Proposed	(Full-IPD)*	MR	Proposed	(Full-IPD)*
40	20	0.054	0.054	(0.053)	0.001	0.001	(0.000)
	40	0.054	0.054	(0.054)	0.000	0.001	(0.000)
	80	0.054	0.054	(0.054)	0.000	0.000	(0.000)
20	20	0.038	0.038	(0.038)	0.000	0.000	(0.000)
	40	0.038	0.039	(0.038)	0.001	0.000	(0.000)
	80	0.038	0.038	(0.038)	0.000	0.000	(0.000)
10	20	0.028	0.028	(0.027)	0.000	0.000	(0.000)
	40	0.027	0.027	(0.027)	0.000	0.000	(0.000)
	80	0.028	0.027	(0.027)	0.000	0.000	(0.000)

MR: Fit the MR model (2.2) to the AD from all 10 trials.

Proposed: Apply the proposed method to the AD from all 10 trials.

*Results by fitting model (2.1) to the full IPD from all 10 trials.

σ_z^2 : Variance parameter in within-trial covariate distributions.

σ_y^2 : Variance parameter in conditional distributions of outcomes.

Table VII. Root mean square errors and mean biases for estimator of within-trial treatment-covariate interaction effect.

σ_y^2	σ_z^2	Root mean square error of $\hat{\gamma}_w$		Mean bias of $\hat{\gamma}_w$	
		Proposed	(Full-IPD)*	Proposed	(Full-IPD)*
40	20	0.220	(0.032)	-0.091	(0.000)
	40	0.206	(0.023)	-0.083	(0.000)
	80	0.202	(0.016)	-0.082	(0.000)
20	20	0.201	(0.023)	-0.090	(0.000)
	40	0.197	(0.016)	-0.081	(0.000)
	80	0.116	(0.011)	-0.017	(0.000)
10	20	0.195	(0.016)	-0.090	(0.000)
	40	0.171	(0.011)	-0.068	(0.000)
	80	0.061	(0.008)	0.005	(0.000)

Proposed: Apply the proposed method to the AD from all 10 trials.

*Results by fitting model (2.1) to the full IPD from all 10 trials.

σ_z^2 : Variance parameter in within-trial covariate distributions.

σ_y^2 : Variance parameter in conditional distributions of outcomes.

indicating that meta-analyses using the MR model (2.2) were prone to fail in detecting the treatment-covariate interaction in comparison with meta-analyses of the original IPD from all 10 trials.

In principle, the proposed method can be used for estimating not only the across-trial interaction effect but also the within-trial interaction effect. The RMSEs and the mean biases for $\hat{\gamma}_A$ from the proposed method were equivalent to those from the MR model (2.2) and the full IPD analysis (Table VI), indicating the log-likelihood (3.8) preserved information on the across-trial interaction precisely. Table VII shows RMSEs and mean biases for the within-trial interaction $\hat{\gamma}_W$ in each scenario. Note that, in the full IPD analysis, we used the IPD from all 10 trials, while in the proposed method, we used only the AD from all 10 trials. When σ_y^2 was large (e.g. $\sigma_y^2 = 40$) and σ_z^2 was small (e.g. $\sigma_z^2 = 20$), $\hat{\gamma}_W$ from the proposed method had a large RMSE with a negative bias. On the other hand, as σ_y^2 became smaller and σ_z^2 became larger, the RMSEs and the mean biases from the proposed method decreased immediately. Especially for the scenarios of $\sigma_y^2 = 10$ and $\sigma_z^2 = 80$, the proposed method offered the smallest RMSE for $\hat{\gamma}_W$ with approximately zero bias. These results suggested that the proposed method required a relatively-large variance in within-trial covariate distributions to estimate the within-trial interaction precisely, and in such situations the use of the log-likelihood (3.8) could recover information on the within-trial relationships from the AD trials.

4.3 Simulation 2: Performance of the proposed method with marginalising the missing IPD in the case for mixture of IPD and AD

We here supposed that some trials provide IPD and the others provide only AD, and focused on the across-trial and the within-trial interaction effect estimated by fitting the existing model (2.5) and the proposed method described in Chapter 3.1. Some practical benefits of combining IPD and AD are as illustrated in Chapter 2, highlighting that using only the collected IPD or reducing available IPD to AD had some disadvantages due to loss of information. We were now interested in how estimates of the within-trial interaction from the proposed method became close to those from the full IPD analysis according to the proportion of trials providing IPD. The standard errors of $\hat{\gamma}_W$ from the proposed method are expected to become smaller as the proportion of trials providing IPD. We also assessed the gains from the proposed method beyond the existing method by Riley *et al.* (2008).

4.3.1 Design of Simulation 2

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

$$\begin{aligned}
 x_{ij} &= \begin{cases} 0, & j \in C \\ 1, & j \in T \end{cases}, \\
 m_{zj} &\sim N(30, 10), \\
 z_{ij}|m_{zi} &\sim N(m_{zi}, \sigma_z^2), \\
 y_{ij}|x_{ij}, z_{ij}, \bar{z}_i &\sim N(20 + 5x_{ij} + 0.05z_{ij} + 0.05x_{ij}\bar{z}_i + 0.1x_{ij}(z_{ij} - \bar{z}_i), 40), \\
 j &= 1, \dots, 400, \quad i = 1, \dots, 20
 \end{aligned} \tag{4.2}$$

where the numbers of patients for the treatment group in each trial were assumed to be equivalent to those for the control group; i.e. $n_{iT} = n_{iC} = 200$ for $i = 1, \dots, 20$. We gave the total number of trials by 20, and controlled the number of trials providing IPD by 5 scenarios of 1, 2, 4, 8, 16 of 20 trials (correspondingly, the numbers of trials providing AD were given by $N \in \{19, 18, 16, 12, 4\}$). As for the variance parameter in within-trial covariate distributions, we considered $\sigma_z^2 = 80$ for the AD trials, and controlled that for the IPD trials by 2 scenarios of $\sigma_z^2 \in \{40, 80\}$.

The implementing procedure was as follows. Firstly, we set the number of IPD trials and σ_z^2 in the IPD trials among 10 scenarios, and then according to the true model (4.2) with parameters set in the previous step, we generated 10,000 sets of meta-analysis data. Secondly, according to the scenario of the number of IPD trials, we summarised the IPD from trials $i = 1, \dots, N$ to the AD. Finally, we meta-analysed a mixture of IPD and AD by: (i) fitting model (2.1) or model (2.4) to the collected IPD from trials $i = N + 1, \dots, 20$, (ii) summarising the collected IPD from trials $i = N + 1, \dots, 20$ to the AD and then fitting the MR model (2.2) to the AD from all 20 trials, (iii) fitting model (2.5) to the mixture of IPD and AD, (iv) applying the proposed method to the mixture of IPD and AD. In each analysis, we computed estimates and their RMSEs for β from method (ii), and those for γ_W and γ_A from method (i), (iii) and (iv). These results were compared with those obtained by fitting model (2.1) to the IPD from all 20 trials (full IPD analysis). We also computed sample mean of absolute differences between estimates of γ_W from model (2.5) or the proposed method and those from the full IPD analysis, which was intended to evaluate how far the point estimate obtained by fitting model (2.5) or the SIPD method is apart from that obtained from the full IPD analysis on average.

4.3.2 Results of Simulation 2

Table VIII shows RMSEs and mean biases for the across-trial interaction in each scenario. The MR model (2.2) again provided estimates only of the across-trial interaction, β , whose RMSEs and mean biases were equivalent to those for $\hat{\gamma}_A$ from the full IPD analysis. Fitting model (2.1) or (2.4) to only the collected IPD had seriously large RMSEs for $\hat{\gamma}_A$ when the proportion of trials providing IPD was small (e.g. 2 IPD trials and 18 AD trials). This is because the precision of the across-trial interaction effect estimate depends on the number of trials involved and between-trial heterogeneity. Table IX shows RMSEs and mean biases for the within-trial interaction in each scenario. The RMSEs and the mean biases for $\hat{\gamma}_W$ obtained by fitting model (2.1) or (2.4) to only the collected IPD got close to those from the full IPD analysis as the proportion of trials providing IPD increased.

The strategy of combining IPD and AD by fitting model (2.5) or the proposed method yielded accurate results for the across-trial interaction, as well as estimates of the within-trial interaction which got close to those from the full IPD analysis according to the proportion of trials providing IPD. Including AD trials remarkably improved the RMSEs for $\hat{\gamma}_A$ in comparison with analyses using only the collected IPD. Fitting model (2.5) and the proposed method provided similar RMSEs and mean biases for $\hat{\gamma}_A$ of each other, which were equivalent to those from the full IPD analysis. The results for the within-trial interaction from model (2.5) were equivalent to those from analyses using only the collected IPD, indicating model (2.5) correctly allowed only the IPD trials to estimate γ_W .

The main gain from the proposed method was to improve RMSEs for $\hat{\gamma}_W$. In most scenarios, the proposed method provided much smaller RMSEs than model (2.5), especially when the proportion of trials providing IPD was small (e.g. 1 IPD trials and 19 AD trials, or 2 IPD trials and 18 AD trials). The absolute differences also confirmed that $\hat{\gamma}_W$'s from the proposed method were, on average, located closer to those from the full IPD analysis than model (2.5).

Table VIII. Root mean square errors and mean biases for estimator of across-trial treatment-covariate interaction effect.

Number of trials providing IPD	σ_z^2	Root mean square error of $\hat{\beta}$ or $\hat{\gamma}_A$					Mean bias of $\hat{\beta}$ or $\hat{\gamma}_A$				
		MR	Obs- IPD	Model (2.5)	Proposed	(Full-IPD)*	MR	Obs- IPD	Model (2.5)	Proposed	(Full-IPD)*
1 / 20	40	0.049	NA	0.049	0.048	(0.048)	-0.001	NA	-0.002	-0.002	(-0.002)
	80	0.048	NA	0.048	0.048	(0.048)	0.001	NA	0.001	0.001	(0.001)
2 / 20	40	0.050	36.07	0.050	0.049	(0.049)	-0.001	1.098	-0.001	-0.001	(-0.001)
	80	0.047	13.41	0.047	0.046	(0.046)	0.001	-0.108	0.001	0.001	(0.001)
4 / 20	40	0.050	0.187	0.050	0.050	(0.050)	0.001	-0.003	0.001	0.001	(0.001)
	80	0.047	0.222	0.047	0.047	(0.047)	0.000	0.001	0.000	0.000	(0.000)
8 / 20	40	0.049	0.090	0.049	0.049	(0.048)	-0.001	0.000	-0.001	-0.001	(-0.001)
	80	0.048	0.090	0.048	0.047	(0.047)	-0.001	0.001	-0.001	-0.001	(-0.001)
16 / 20	40	0.048	0.054	0.048	0.048	(0.047)	0.001	0.000	0.001	0.001	(0.001)
	80	0.049	0.055	0.049	0.048	(0.048)	-0.001	-0.001	-0.001	-0.001	(-0.001)

MR: Fit the MR model (2.2) to the AD from all 20 trials.

Obs-IPD: Fit model (2.1) to the collected IPD.

Model (2.5): Fit model (2.5) to the mixture of IPD and AD.

Proposed: Apply the proposed method to the mixture of IPD and AD.

*Results by fitting model (2.1) to the full IPD from all 20 trials.

 σ_z^2 : Variance parameter in within-trial covariate distributions for patients in the IPD trial.

NA: Not available.

Table IX. Root mean square errors and mean biases for estimator of within-trial treatment-covariate interaction effect, and sample means of absolute differences between estimates from model (2.5) or the proposed method and those from the full IPD analysis.

Number of trials providing IPD	σ_z^2	Root mean square error of $\hat{\gamma}_w$				Mean bias of $\hat{\gamma}_w$				Mean of absolute difference	
		Obs- IPD	Model (2.5)	Proposed	(Full-IPD)*	Obs- IPD	Model (2.5)	Proposed	(Full-IPD)*	Model (2.5)	Proposed
1 / 20	40	0.101	0.101	0.082	(0.016)	0.001	0.001	0.002	(0.000)	0.079	0.060
	80	0.069	0.069	0.057	(0.016)	0.000	0.000	0.002	(0.000)	0.054	0.044
2 / 20	40	0.072	0.072	0.059	(0.016)	-0.002	-0.002	0.001	(0.000)	0.056	0.044
	80	0.050	0.050	0.044	(0.016)	-0.001	-0.001	0.000	(0.000)	0.038	0.032
4 / 20	40	0.050	0.050	0.045	(0.016)	0.001	0.001	0.002	(0.000)	0.038	0.033
	80	0.035	0.035	0.033	(0.016)	0.000	-0.001	0.000	(0.000)	0.025	0.023
8 / 20	40	0.036	0.036	0.034	(0.018)	0.000	0.000	0.001	(0.000)	0.025	0.023
	80	0.025	0.025	0.024	(0.016)	0.000	0.000	0.000	(0.000)	0.016	0.015
16 / 20	40	0.025	0.025	0.025	(0.020)	0.000	0.000	0.000	(0.000)	0.011	0.011
	80	0.018	0.018	0.018	(0.016)	0.000	0.000	0.000	(0.000)	0.006	0.006

Obs-IPD: Fit model (2.1) to the collected IPD.

Model (2.5): Fit model (2.5) to the mixture of IPD and AD, Proposed: Apply the proposed method to the mixture of IPD and AD.

*Results by fitting model (2.1) to the full IPD from all 20 trials.

σ_z^2 : Variance parameter in within-trial covariate distributions for patients in the IPD trial.

4.4 Application to hypertension data: Illustration of the proposed method based on simulated IPD

Consider the hypertension data, and we illustrate the SIPD method via Bayesian procedure described in Chapter 3.2. To imitate situations involving IPD for some trials and only AD for others, we considered scenarios as in Chapter 2.1, where only a limited number of trials (from 1 to 4 of the 5 trials) provided IPD and the other trials just provided AD. In each scenario, we carried out analyses by: (i) fitting the MR model (2.2) to AD from all 5 trials, (ii) fitting model (2.5) to the mixture of IPD and AD, and (iii) applying the SIPD method to the mixture of IPD and AD. In both parts (ii) and (iii), the analyses were run for each possible combination of IPD and AD trials. In each scenario, we compared the results with those from a meta-analysis of IPD from all 5 trials (full IPD analysis).

In the SIPD method, for the iterative process of the Metropolis-Hastings algorithm to draw $R = 500$ values of parameter, we discarded the first 5,000 samples in order to prevent dependence on the starting values. Moreover, we took a sample at only every 1,000th iteration in order to avoid autocorrelation between the samples taken. The same iterative process will be taken in Simulation 3 and Simulation 4.

The gains from the SIPD method

The results of estimates and their standard errors for the across-trial and the within-trial interaction, averaged across all possible combinations of IPD and AD trials in each scenario are shown in Table X. As for the across-trial interaction, the SIPD method produced estimates closer to the full IPD analysis compared to model (2.5) regardless of the number of IPD trials. The estimates of γ_A from model (2.5) were also close to those from the MR model (2.2). For each scenario in Table X, we also found an important difference between results for the within-trial interaction effect from model (2.5) and the SIPD method. When comparing $\hat{\gamma}_W$'s from model (2.5) with those from the full IPD analysis, model (2.5) provided point estimates located in a positive direction on average, with large standard errors. This is because model (2.5) allows only the IPD trials to estimate the within-trial interaction, and thus the estimates and their standard error for γ_W by fitting model (2.5) got close to those from the full IPD analysis as the available number of IPD trials increases. The SIPD method improved both the estimates of γ_W and their standard errors to be closer to the correct (full IPD) estimates, especially when the number of IPD trials was small. The most benefit came

Table X. Average of estimates and their standard errors for treatment-covariate interaction effect when analysing change in SBP (follow-up minus baseline) from hypertension data, where estimates are averaged across all combinations of IPD trials.

	Number of trials providing IPD [‡]	Average of estimate		Average of standard error	
		Model (2.5)	SIPD	Model (2.5)	SIPD
Across-trial interaction effect γ_A	5 / 5*	-0.662	-	0.464	-
	4 / 5	-0.766	-0.662	0.468	0.464
	3 / 5	-0.766	-0.663	0.463	0.464
	2 / 5	-0.768	-0.665	0.464	0.464
	1 / 5	-0.781	-0.667	0.464	0.463
	0 / 5 [†]	-0.766	-	0.466	-
Within-trial interaction effect γ_W	5 / 5*	0.087	-	0.055	-
	4 / 5	0.091	0.090	0.063	0.063
	3 / 5	0.117	0.096	0.084	0.076
	2 / 5	0.166	0.103	0.130	0.097
	1 / 5	0.244	0.105	0.258	0.131

Model (2.5): Fit model (2.5) to the mixture of IPD and AD.

SIPD: Apply the SIPD method to the mixture of IPD and AD.

*Results by fitting model (2.1) to the full IPD from all 5 trials.

[†]Results by fitting the MR model (2.2) to the AD from all 5 trials.

[‡]The numbers of combinations of trials providing IPD are 5, 10, 10 and 5 in the scenarios of 1, 2, 3 and 4 IPD trials, respectively.

in the scenario of 1 IPD trial, in which estimates of γ_W were $\hat{\gamma}_W = 0.244$ (s.e. = 0.258) from model (2.5) 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 using the AD trials and IPD trials to estimate the within-trial interaction, and this adjustment based on the AD trials is useful especially when the number of IPD trials is small.

Table X also shows that the difference between the estimates of γ_W (and its standard errors) from model (2.5) and those from the SIPD method became smaller when increasing the number of IPD trials. The results in the scenario of 3 IPD trials and 2 AD trials were similar to those using the full IPD, and the results in the scenario of 4 IPD trials and 1 AD trial were almost equivalent to those using the full IPD. These results suggested that model (2.5) could provide sufficiently accurate estimates of the within-trial interaction if a high proportion of IPD trials are available.

For each scenario, Figure 8 shows estimates of γ_W and their standard errors obtained by

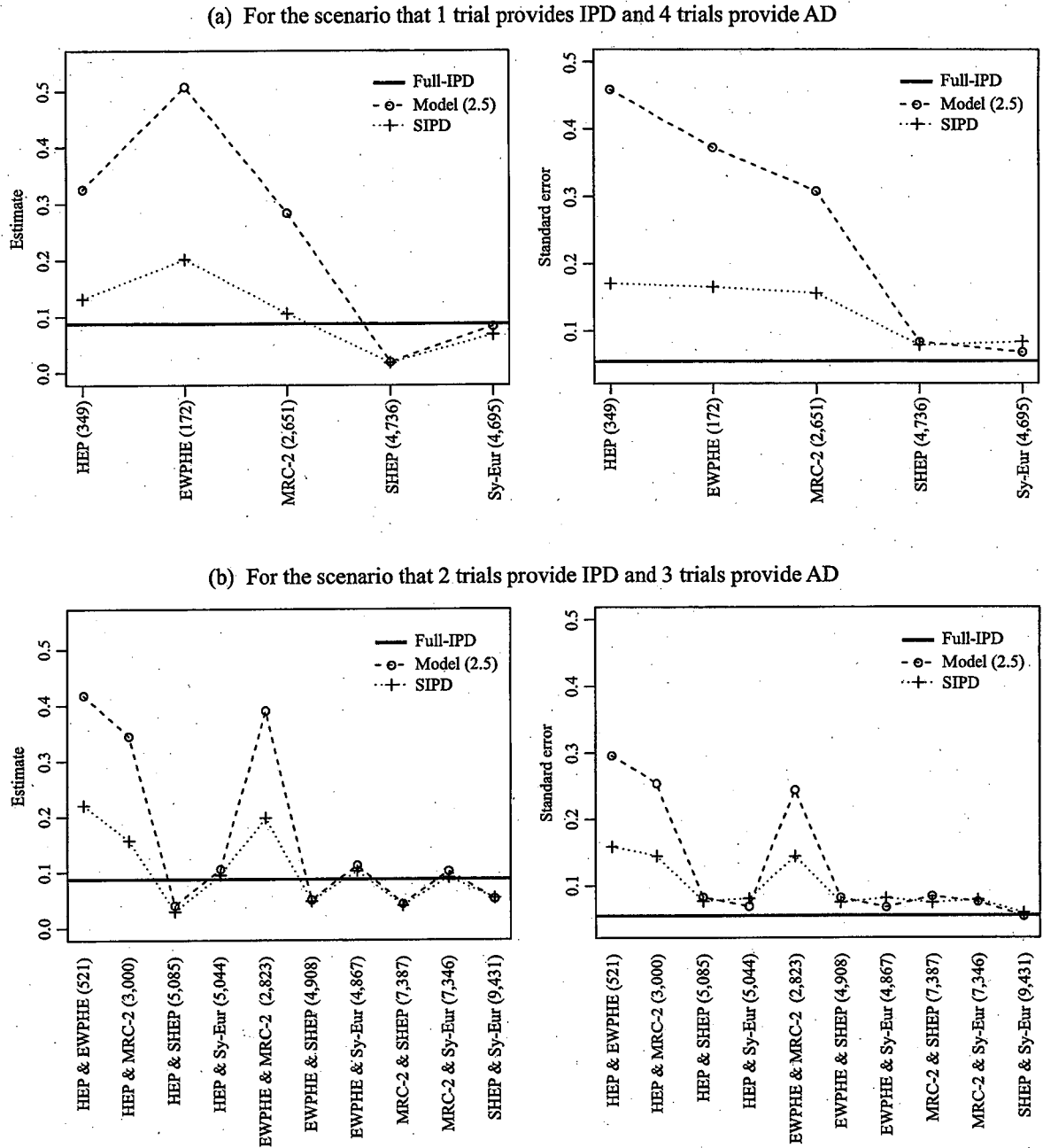
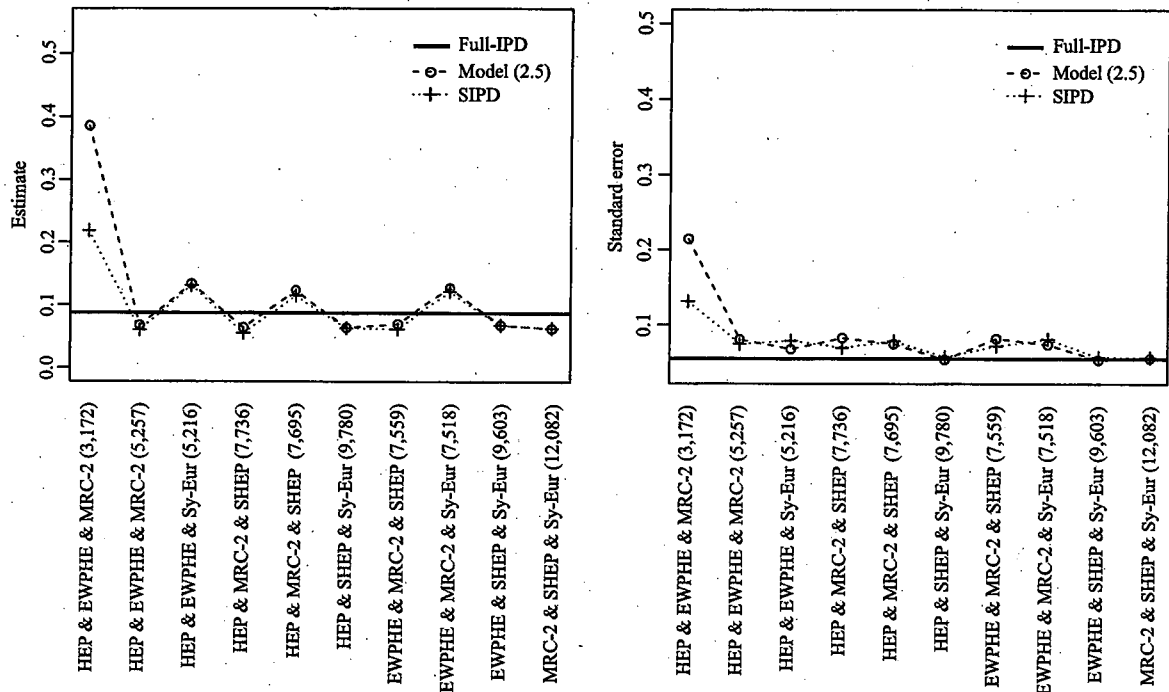


Figure 8. Estimates and their standard errors for within-trial treatment-covariate interaction effect when analysing change in SBP (follow-up minus baseline) from hypertension data in the scenarios that: (a) 1 trial provides IPD and 4 trials provide AD, (b) 2 trials provide IPD and 3 trials provide AD, (c) 3 trials provide IPD and 2 trials provide AD, (d) 4 trials provide IPD and 1 trial provides AD.

(c) For the scenario that 3 trials provide IPD and 2 trials provide AD



(d) For the scenario that 4 trials provide IPD and 1 trial provides AD (*The name of AD trials)

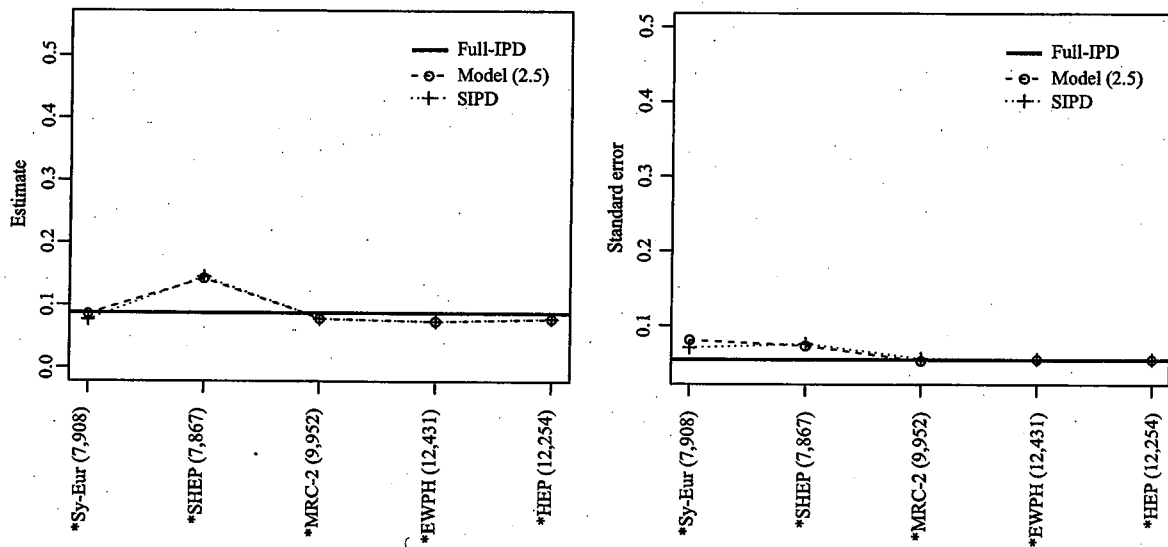


Figure 8 (continued).

fitting model (2.5) and the SIPD method for each combination, where the horizontal axis represents the name of each IPD trial with sample size in parentheses. For example, in the scenarios of 2 IPD trials and 3 AD trials, names of 2 IPD trials and sum of sample sizes from 2 IPD trials are shown, and also in the scenarios of 4 IPD trials and 1 AD trial, names of 1 AD trial and sum of sample sizes from 4 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. For almost all combinations of 1 IPD trial, the SIPD method provided estimates of γ_w and their standard errors which were located closer to those from the full IPD analysis than model (2.5). 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 were seen for the scenario of 2 IPD trials, although the results by the 2 methods were closer, and almost equivalent in the scenario of 4 IPD trials and 1 AD trial.

The difference between the results from model (2.5) 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 8 (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 X. Figure 9 shows

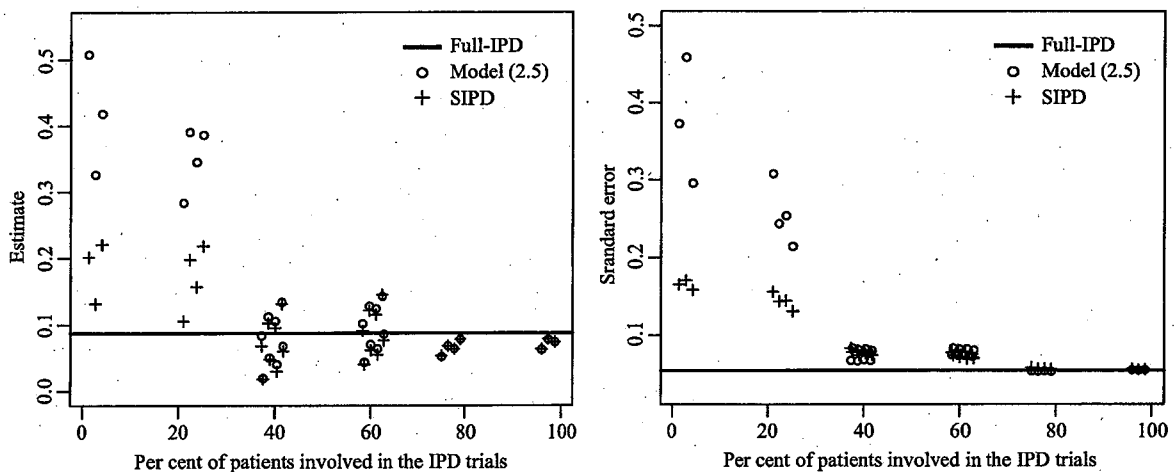


Figure 9. Estimates and their standard errors for within-trial treatment-covariate interaction effect sorted by the proportion of available IPD in all patients when analysing change in SBP (follow-up minus baseline) from hypertension data.

estimates of γ_W and their standard errors from model (2.5) 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 (2.5) 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 per cent of patients was available. Thus, the SIPD method had most notable benefits when the proportion of patients involved in the IPD trials was low.

The SIPD method could bring meta-analysts some other potential advantages, rather than just provides $\hat{\gamma}_W$ and its standard errors which are closer to those from the full IPD analysis in comparison with model (2.5). Once obtaining the SIPD, one can apply any IPD meta-analysis approaches to each set of SIPD combined with the collected IPD. We here

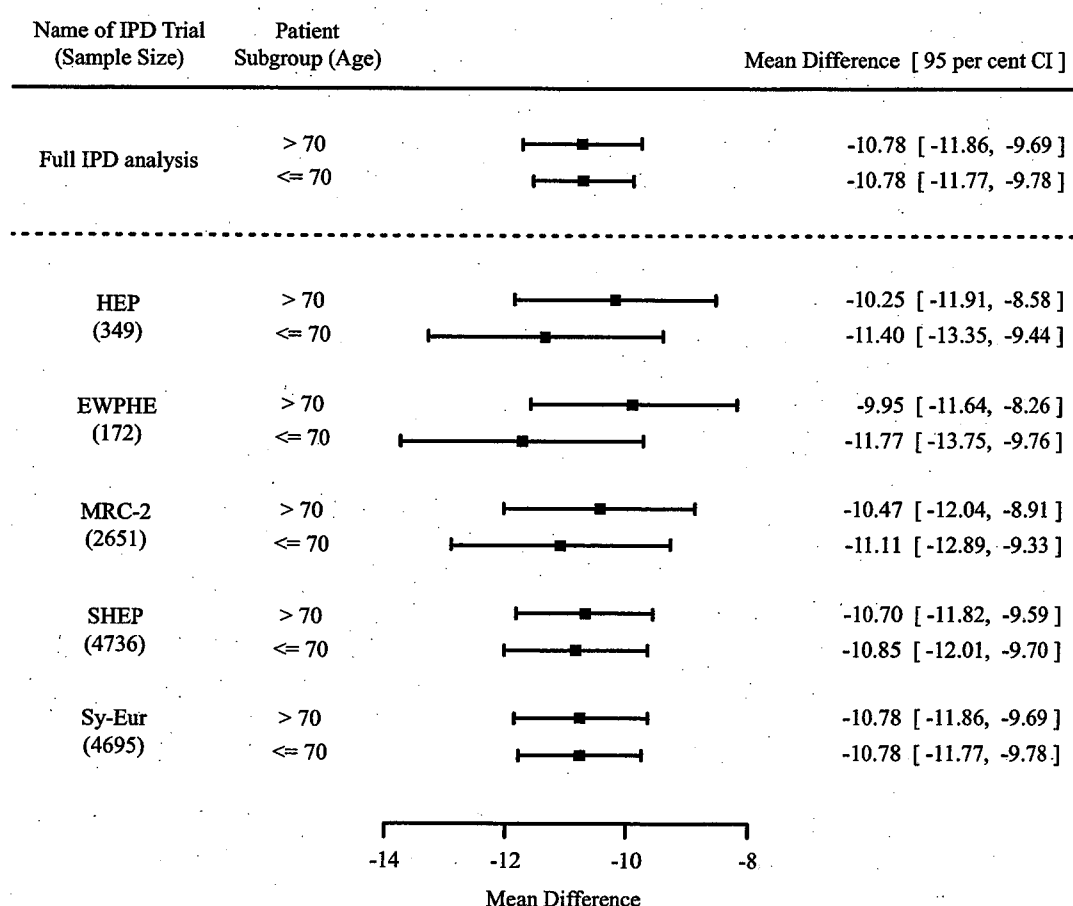


Figure 10. Pooled estimates of mean difference on change in SBP (follow-up minus baseline) between groups and their 95 per cent confidence intervals for 2 patient subgroups of age.

considered a situation of exploring beneficial effects in specific patient subgroups, and meta-analysed these subgroups within each trial in hypertension, which were identified by age of patients. Obviously, model (2.5) for combining IPD and AD cannot provide any findings for this. We allocated each patient in the IPD trials and each simulated patient in the AD trials to 2 or 3 subgroups according to the following scenarios: (i) whether age of the patient is more than 70, or not, (ii) whether age of the patient is more than 73, 67 or more to 73 less, or not. Using covariate values generated for patients in the AD trials enables one to estimate pooled treatment effects for each patient subgroup. We here considered only a situation where 1 trial provides IPD and the other 4 trials provide AD. Figure 10 shows the pooled

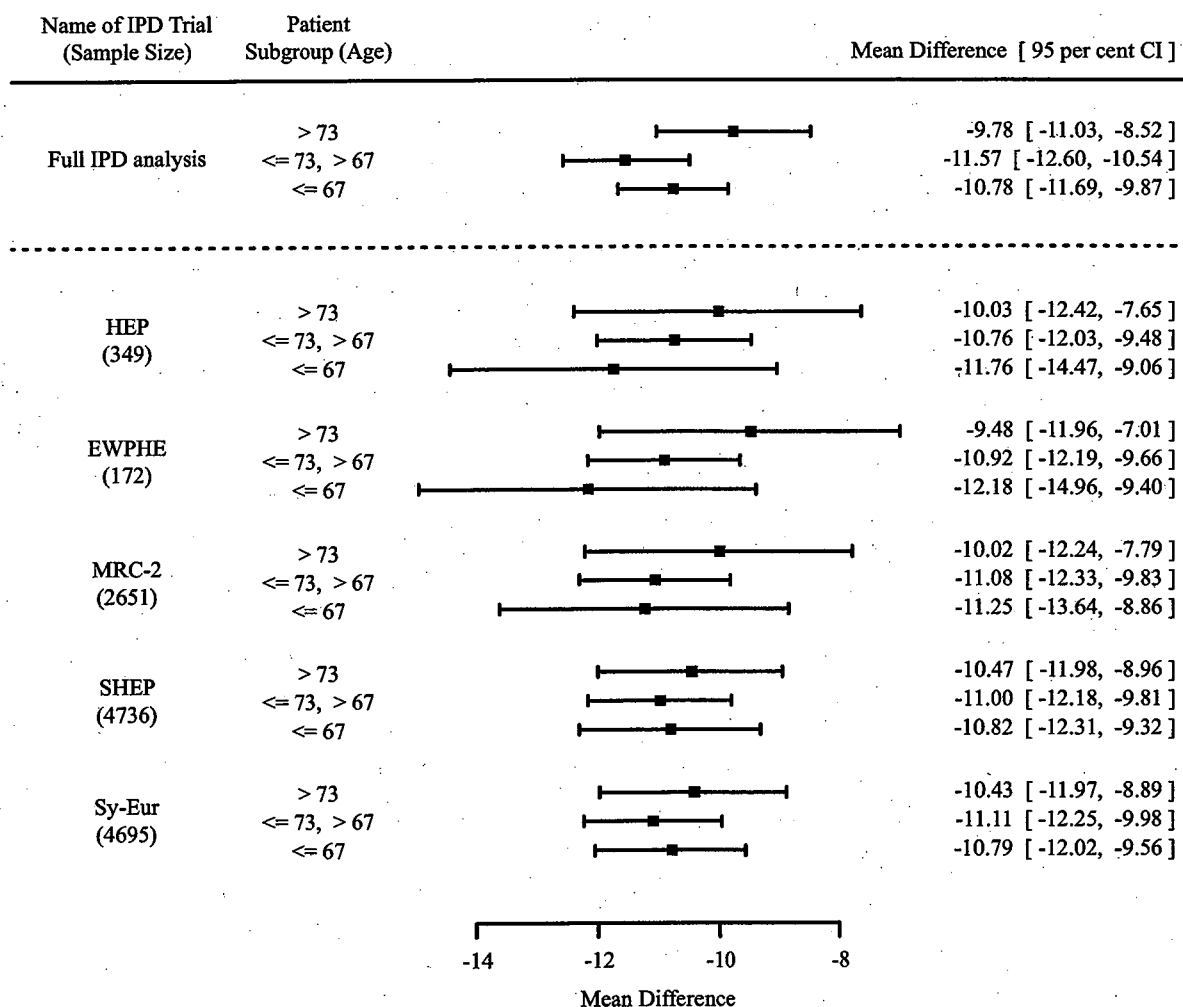


Figure 11. Pooled estimates of mean difference on change in SBP (follow-up minus baseline) between groups and their 95 per cent confidence intervals for 3 patient subgroups of age.

treatment effect estimates and their 95 per cent confidence intervals for the scenario (i) of 2 patient subgroup. The uppermost block represents results from the full IPD analysis. The names of 1 IPD trial are depicted in the left hand side, squares and horizontal lines around them denote the pooled treatment effect estimates and their 95 per cent confidence intervals respectively. When the 1 IPD trial had large sample size (e.g. SHEP and Sy-Eur), results for each patient subgroup from the SIPD method were similar to those from the full IPD analysis. Figure 11 also shows the pooled treatment effect estimates and their 95 per cent confidence intervals for the scenario (ii) of 3 patient subgroup, and the same findings were seen for the IPD trials with large sample size.

4.5 Simulation 3: Performance of the proposed method based on simulated IPD in the situation of continuous outcome and covariate

From the results in the application to the hypertension data, it was shown that the SIPD method via Bayesian procedure provided estimates of the within-trial interaction closer to those from the full IPD analysis than model (2.5), when the number of IPD trials was small and when the number of patients involved in the IPD trials was small. To check this finding, we here focused on the within-trial interaction effect, and compared some statistical properties of $\hat{\gamma}_W$ obtained by the SIPD method with those obtained by fitting model (2.5) under some settings of controlled parameters and the number of patients involved in 1 IPD trial and 9 AD trials.

4.5.1 Design of Simulation 3

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

$$\begin{aligned}
 x_{ij} &= \begin{cases} 0, & j \in C \\ 1, & j \in T \end{cases}, \\
 m_{zi} &\sim N(\hat{m}_z, \hat{\sigma}_m^2), \\
 z_{ij}|m_{zi} &\sim N(m_{zi}, \sigma_z^2), \\
 y_{ij}|x_{ij}, z_{ij}, \bar{z}_i &\sim N(\hat{\phi}_i + \hat{\theta}x_{ij} + \hat{\mu}z_{ij} + \hat{\gamma}_A x_{ij}\bar{z}_i + \gamma_W x_{ij}(z_{ij} - \bar{z}_i), \sigma_y^2), \\
 j &= 1, \dots, n_i; \quad i = 1, \dots, 10
 \end{aligned} \tag{4.3}$$

where the true parameters except for γ_W and σ_y^2 were given as estimates by fitting model (2.1) to IPD from 10 trials originally reported in Wang *et al.* (2005), with change in SBP as an outcome; e.g. $\hat{\theta} = -4.958$, $\hat{\mu} = -0.042$ and $\hat{\gamma}_A = -0.079$. The total number of patients was given by $\sum_{i=1}^{10} n_i = 6,000$, 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 the 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 1 trial provided IPD (the other 9 trials provided AD), and controlled the number of patients involved in the 1 IPD trial by six scenarios of 60, 300, 600, 1,200, 2,400 and 4,800; so that the proportions of patients with available IPD were given by 1, 5, 10, 20, 40 and 80 per cent, respectively. The 9 AD trials involved almost the same number of patients for each scenario.

We here considered $\sigma_z^2 = 100$ for the 9 AD trials, and controlled that for the 1 IPD trial by 3 scenarios of $\sigma_z^2 \in \{25, 50, 100\}$. These scenarios of σ_z^2 lead us to a situation that the 1 IPD trial provides information on the within-trial interaction 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 the within-trial interaction 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 1 IPD trial and σ_z^2 for each of 18 scenarios, and then generated 5,000 sets of meta-analysis data according to the true model (4.3) with parameters set in the previous step for each scenario. Secondly, for each set in each scenario, we summarised IPD for 9 of the 10 trials to AD. Finally, we analysed the mixture of IPD and AD by 2 methods: model (2.5) and the SIPD method. In each analysis, we computed mean-square error (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 (2.5) or the SIPD method and those obtained from the full IPD analysis. Moreover, we estimated the type I error rate and the statistical power with one-sided hypothesis test at 5 per cent level of significance for $H_0 : \gamma_W = 0$ and $H_1 : \gamma_W < 0$.

4.5.2 Results of Simulation 3

The results of MSE, mean bias and mean standard error for each scenario are shown in Table XI. In each scenario, the SIPD method provided substantially smaller MSEs and mean standard errors in comparison with model (2.5), especially when the proportion of patients with available IPD was low (e.g. 1 or 5 per cent) 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 (2.5). The difference between the results from model (2.5) and the SIPD method was the largest for the scenario of 1 per cent 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 of γ_W and its standard error 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 per cent of patients in the IPD trial and $\sigma_z^2 = 25$, the MSE was reduced by 50 per cent using the SIPD method (MSE = 0.051) rather than model (2.5) (MSE = 0.111); similarly the standard error was reduced considerably by using SIPD (mean s.e. = 0.248) rather than model (2.5) (mean s.e. = 0.328).

However, in the scenarios of smaller proportion of patients with available IPD, the estimates of γ_W from the SIPD method were more subject to a positive bias. For example, the mean bias from the SIPD method for the scenario of 1 per cent 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 SIPD method allows one to extract the information on the within-trial relationships from the AD trial by using (3.18), and thus we gain substantially smaller MSEs and mean standard errors for $\hat{\gamma}_W$ in comparison to model (2.5). On the other hand, this information from the AD trials also pull 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 per cent 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 per cent or over, and the SIPD method still has gain in MSE (up to about 40 per cent) and standard error (up to about 20 per cent) in situations between 10 and 40 per cent of patients with available IPD. Figure 12 also shows the estimates of γ_W for the scenario of $\sigma_z^2 = 100$, which are arranged in ascending order of estimates from model (2.5) and then suitably smoothed by taking an average of each 100 estimates. From Figure 12, the difference between results obtained from model (2.5) and the SIPD method were seen dynamically. In particular, for the 1 per cent of patients in 1 IPD trial in panel (a), it was confirmed that the estimates from the SIPD method were much closer to those from the full IPD analysis in

comparison with model (2.5), and suffered from a positive bias.

Figure 13 shows the type I error rates and the power for $\hat{\gamma}_W$ estimated by the 3 methods for each scenario. In the scenarios of 1 and 5 per cent 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 biased estimates of γ_W unlike when γ_W was -0.2 (results not shown in Table XI). 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 (2.5). In the scenarios of over 10 per cent of available IPD, the SIPD method had better type I error rates close to 5 per cent. The powers of model (2.5) and the SIPD method to detect the true negative interaction were very similar. The SIPD method was marginally better when 10 to 40 per cent of patients were in the IPD trial.

Table XI. 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 (2.5) or the SIPD method (for IPD from one trial and AD from nine trial) 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 (2.5)	SIPD	(Full-IPD)*	Model (2.5)	SIPD	(Full-IPD)*	Model (2.5)	SIPD	(Full-IPD)*	Model (2.5)	SIPD
1 per cent	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 per cent	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 per cent	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 per cent	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 per cent	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 per cent	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

Model (2.5): Fit model (2.5) 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 (2.1) to the full IPD from all trials.

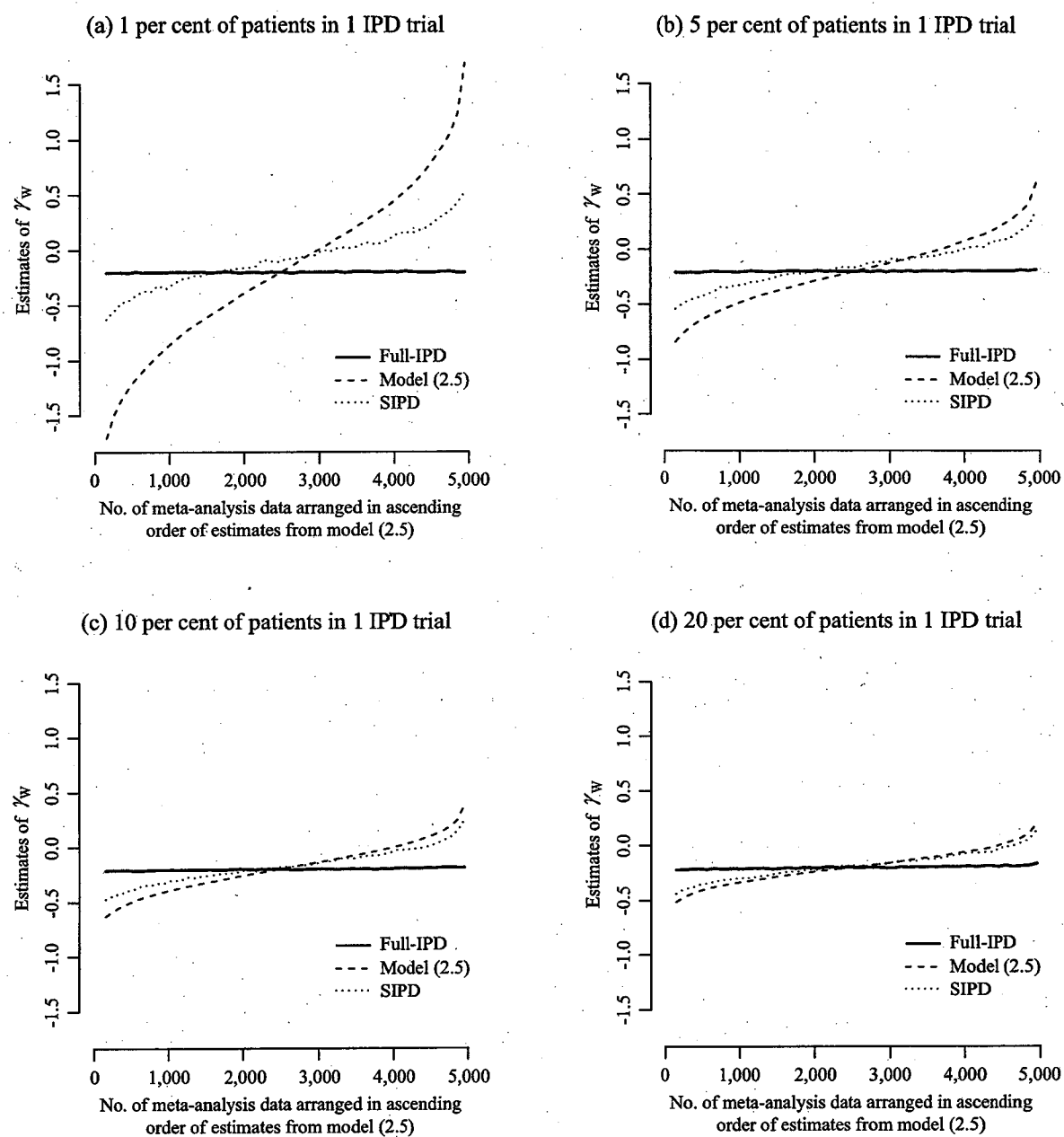


Figure 12. Estimates of within-trial treatment-covariate interaction effect for the scenario of $\sigma_z^2 = 100$.

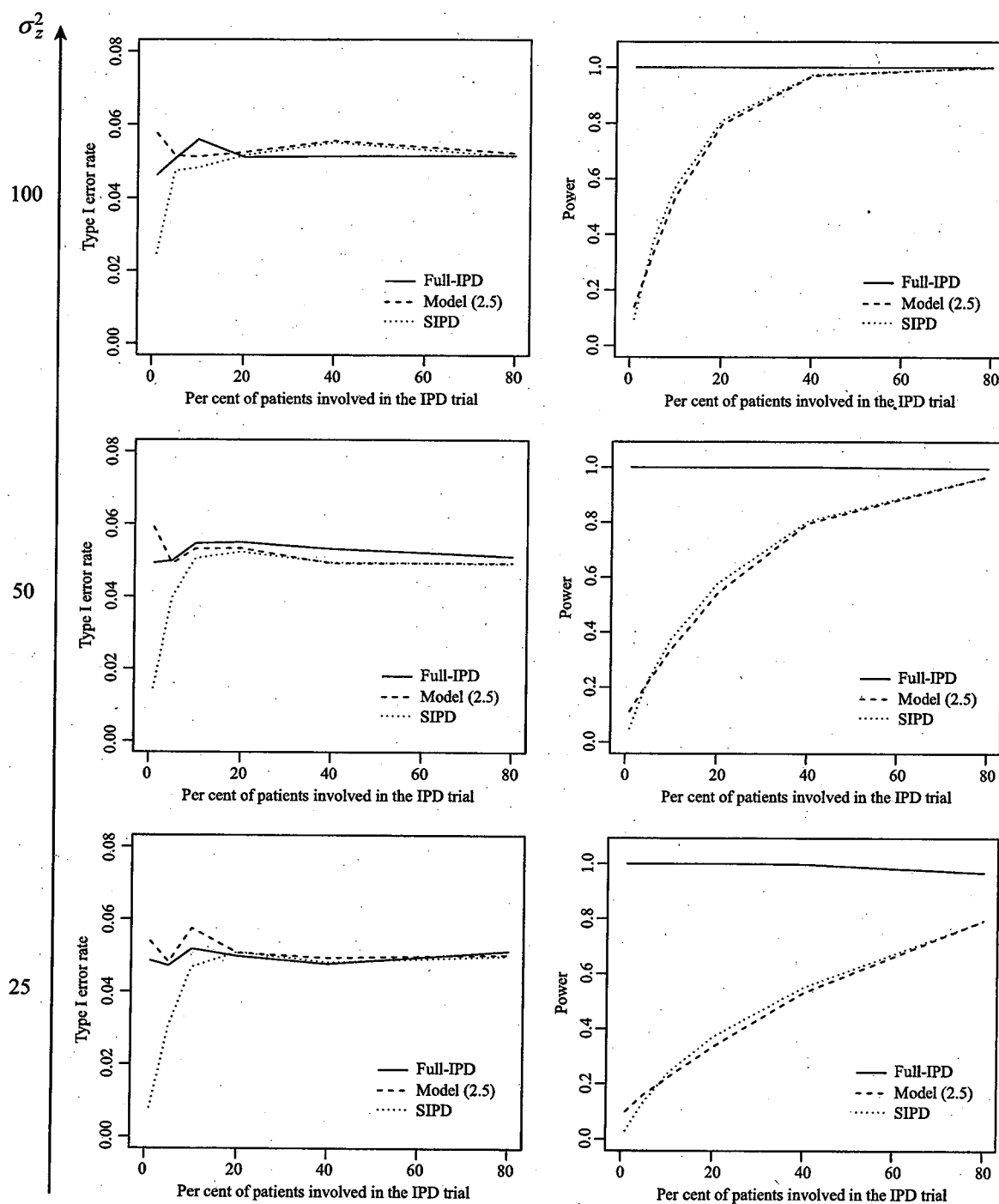


Figure 13. Type I error rates (three panels on the left side) and powers (three panels on the right side) for within-trial treatment-covariate interaction effect.

4.6 Simulation 4: Performance of the proposed method based on simulated IPD in the situation of binary outcome and covariate

Through an application to the hypertension data and Simulation 3, in the situation of continuous outcome and covariate, the SIPD method via Bayesian procedure improved the existing method by Riley *et al.* (2008) in the sense of matching the full IPD analysis for inference of the within-trial interaction. We here supposed a situation where a single binary outcome and covariate are observed from each patient, and focused on the across-trial and the within-trial interaction effect estimated by fitting the existing model (2.9) and the SIPD method via Bayesian procedure described in Chapter 3.2. The methods were applied for the case that some trials provide IPD and the others provide only AD. We were again interested in how estimates of the within-trial interaction from the proposed method became close to those from the full IPD analysis according to the proportion of trials providing IPD.

4.6.1 Design of Simulation 4

We considered that the true models for generating individual-specific outcomes and covariates from each study were written as follows:

$$\begin{aligned}
 x_{ij} &= \begin{cases} 0, & j \in C. \\ 1, & j \in T \end{cases}, \\
 z_{ij} &\sim \text{Bernoulli}(m_{zi}), \\
 y_{ij}|x_{ij}, z_{ij}, \bar{z}_i &\sim \text{Bernoulli}(q_{ij}), \\
 \log \frac{q_{ij}}{1 - q_{ij}} &= -2 - 0.5x_{ij} + 0.5z_{ij} + 2x_{ij}\bar{z}_i + x_{ij}(z_{ij} - \bar{z}_i), \\
 j &= 1, \dots, 400, \quad i = 1, \dots, 10
 \end{aligned} \tag{4.4}$$

where the numbers of patients for the treatment group in each trial were assumed to be equivalent to those for the control group; i.e. $n_{iT} = n_{iC} = 200$ for $i = 1, \dots, 10$. We gave the total number of trials by 10, and controlled the number of trials providing IPD by 4 scenarios of 1, 2, 4, 8 trials (the numbers of AD trials were given by $N \in \{9, 8, 6, 2\}$). We also controlled the true proportion of patients with $Z = 1$ in each trial, m_{zi} for $i = 1, \dots, 10$, by 3 scenarios of low-, moderate- and high-heterogeneity across trials. The true proportions of patients with $Z = 1$ for all 10 trials were given by $\{0.40, 0.40, 0.45, 0.45, 0.50, 0.50, 0.55,$

0.55, 0.60, 0.60}, {0.30, 0.30, 0.40, 0.40, 0.50, 0.50, 0.60, 0.60, 0.70, 0.70} and {0.10, 0.10, 0.30, 0.30, 0.50, 0.50, 0.70, 0.70, 0.90, 0.90} respectively; one for each trial. The standard errors of $\hat{\gamma}_A$ are expected to become smaller in the scenario of high-heterogeneity on the mean covariate (Lambert *et al.*, 2002). Each patient in each trial was allocated to a patient subgroup with $Z = 1$ or $Z = 0$, using predefined proportions for the trial, which varied from trial to trial.

The implementing procedure was as follows. Firstly, we set the number of IPD trials and m_{zi} for $i = 1, \dots, 10$ among 12 scenarios, and then according to the true model (4.4), we generated 5,000 sets of meta-analysis data. Secondly, for each scenario, we summarised the IPD from trials $i = 1, \dots, N$ to the AD. Finally, we meta-analysed a mixture of IPD and AD by: (i) fitting model (2.9) to the mixture of IPD and AD, (ii) applying the SIPD method via Bayesian procedure to the mixture of IPD and AD. In each analysis, we computed RMSE, mean bias and mean standard error for $\hat{\gamma}_A$ and $\hat{\gamma}_W$. We also computed sample mean of absolute differences between estimates of γ_A (and γ_W) obtained by fitting model (2.9) or the SIPD method and those obtained from the full IPD analysis. Moreover, we estimated the type I error rate and the statistical power for γ_W with two-sided hypothesis test at 5 per cent level of significance for $H_0 : \gamma_W = 0$ and $H_1 : \gamma_W \neq 0$.

4.6.2 Results of Simulation 4

The results of MSE, mean bias and mean standard error for the across-trial interaction are shown in Table XII. In each scenario, the SIPD method provided similar results to the full IPD analysis. By contrast, the estimates from model (2.9) had a negative bias, especially in the scenario of low-heterogeneity on the mean covariate. The results of MSE, mean bias and mean standard error for the within-trial interaction are shown in Table XIII. In each scenario, the SIPD method provided substantially smaller MSEs and mean standard errors in comparison with model (2.9), especially when the number of trials providing IPD was small (e.g. 1 or 2 IPD trials) and the heterogeneity on the mean covariate was high. 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 (2.9). The difference between the results from model (2.9) and the SIPD method was the largest for the scenario of 1 IPD trial and high-heterogeneity on the mean covariate, and became smaller as the number of trials providing IPD was larger and heterogeneity on the mean covariate was lower. Figure 14 also shows the estimates of the within-trial interaction

for the scenario of high-heterogeneity on the mean covariate, which are arranged in ascending order of estimates from model (2.9) and then suitably smoothed by taking an average of each 100 estimates. From Figure 14, the difference between results obtained from model (2.9) and the SIPD method were seen dynamically. In particular, for 1 IPD trial and 9 AD trials in panel (a), it was confirmed that the estimates from the SIPD method were much closer to those from the full IPD analysis in comparison with model (2.9). Figure 15 shows the type I error rates and the power for $\hat{\gamma}_W$ estimated by the 3 methods for each scenario. The powers of the SIPD method to detect the true positive interaction were higher than those from model (2.9), especially when the number of trials providing IPD was small (e.g. 1 or 2 IPD trials) and the heterogeneity on the mean covariate was high.

Table XII. Mean-square errors, mean biases and mean standard errors for estimator of across-trial treatment-covariate interaction effect, and sample means of absolute differences between estimates from model (2.9) or the SIPD method and those from the full IPD analysis (for IPD from all 10 trials).

Number of trials providing IPD	Covariate heterogeneity	Mean-square error of $\hat{\gamma}_A$			Mean bias of $\hat{\gamma}_A$			Mean standard error of $\hat{\gamma}_A$			Mean of absolute difference	
		Model (2.9)	SIPD	(Full-IPD)*	Model (2.9)	SIPD	(Full-IPD)*	Model (2.9)	SIPD	(Full-IPD)*	Model (2.9)	SIPD
1 / 10	Low	1.119	1.275	1.170	-0.150	-0.002	0.011	1.064	1.099	1.088	0.221	0.189
	Moderate	0.304	0.351	0.315	-0.156	-0.018	-0.004	0.550	0.568	0.561	0.159	0.097
	High	0.082	0.090	0.083	-0.072	0.005	0.012	0.285	0.291	0.288	0.086	0.041
2 / 10	Low	1.123	1.243	1.156	-0.182	-0.044	-0.035	1.060	1.087	1.081	0.202	0.167
	Moderate	0.299	0.331	0.312	-0.118	0.010	0.018	0.549	0.563	0.559	0.146	0.086
	High	0.079	0.086	0.081	-0.079	-0.009	-0.003	0.285	0.290	0.288	0.077	0.036
4 / 10	Low	1.226	1.310	1.262	-0.089	0.017	0.019	1.069	1.088	1.085	0.169	0.144
	Moderate	0.308	0.331	0.316	-0.096	0.002	0.003	0.552	0.561	0.560	0.111	0.073
	High	0.085	0.089	0.087	-0.037	0.018	0.019	0.286	0.289	0.288	0.058	0.029
8 / 10	Low	1.177	1.200	1.178	-0.019	0.020	0.020	1.080	1.086	1.085	0.085	0.079
	Moderate	0.316	0.322	0.317	-0.015	0.019	0.019	0.557	0.560	0.560	0.049	0.040
	High	0.083	0.084	0.083	-0.013	0.006	0.006	0.287	0.288	0.288	0.023	0.017

Model (2.9): Fit model (2.9) to the mixture of IPD and AD, SIPD: Apply the SIPD method to the mixture of IPD and AD.

*Results by fitting model (2.6) to the full IPD from all trials.

Table XIII. 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 (2.9) or the SIPD method and those from the full IPD analysis (for IPD from all 10 trials).

Number of trials providing IPD	Covariate heterogeneity	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 (2.9)	SIPD	(Full-IPD)*	Model (2.9)	SIPD	(Full-IPD)*	Model (2.9)	SIPD	(Full-IPD)*	Model (2.9)	SIPD
1 / 10	Low	0.259	0.254	0.026	0.023	0.030	-0.003	0.513	0.490	0.163	0.388	0.388
	Moderate	0.281	0.208	0.027	0.022	0.032	0.008	0.531	0.454	0.167	0.398	0.348
	High	0.827	0.231	0.037	0.007	0.040	0.006	0.714	0.473	0.194	0.579	0.380
2 / 10	Low	0.128	0.130	0.026	0.015	0.018	0.010	0.361	0.338	0.162	0.266	0.267
	Moderate	0.132	0.115	0.027	-0.001	0.007	0.006	0.373	0.331	0.167	0.257	0.245
	High	0.192	0.115	0.038	-0.003	0.006	-0.001	0.445	0.344	0.194	0.311	0.259
4 / 10	Low	0.065	0.065	0.028	0.001	0.004	0.001	0.256	0.231	0.162	0.156	0.156
	Moderate	0.073	0.067	0.027	0.005	0.009	0.002	0.264	0.234	0.167	0.170	0.164
	High	0.092	0.066	0.037	0.001	0.009	0.003	0.309	0.257	0.194	0.186	0.165
8 / 10	Low	0.032	0.032	0.026	0.004	0.005	0.002	0.181	0.168	0.162	0.068	0.067
	Moderate	0.036	0.036	0.030	0.010	0.011	0.010	0.187	0.173	0.167	0.066	0.065
	High	0.049	0.045	0.038	0.021	0.020	0.020	0.216	0.200	0.194	0.078	0.075

Model (2.9): Fit model (2.9) to the mixture of IPD and AD, SIPD: Apply the SIPD method to the mixture of IPD and AD.

*Results by fitting model (2.6) to the full IPD from all trials.

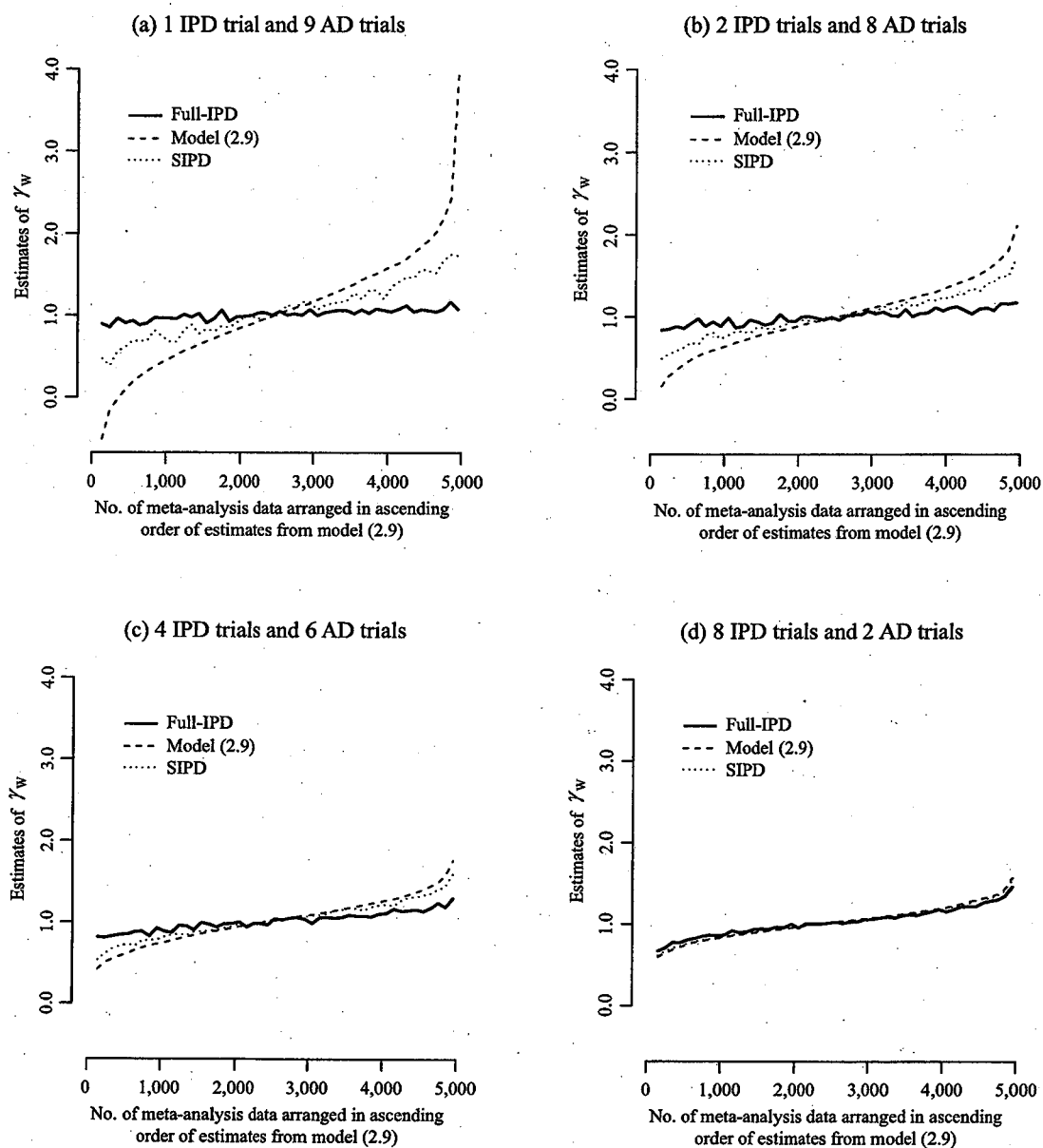


Figure 14. Estimates of within-trial treatment-covariate interaction effect for the scenario of high-heterogeneity on mean covariate.

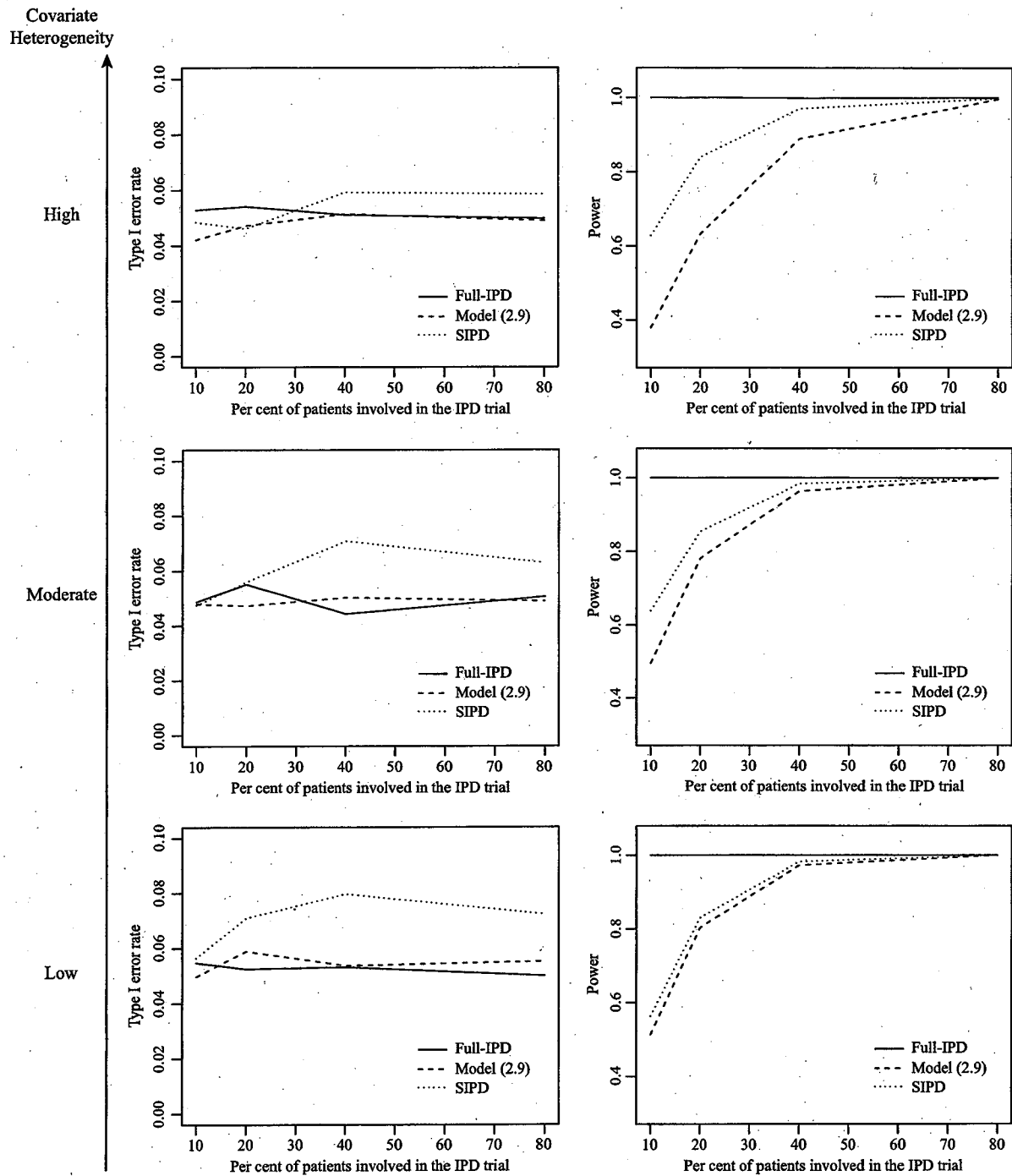


Figure 15. Type I error rates (three panels on the left side) and powers (three panels on the right side) for within-trial treatment-covariate interaction effect.

5 Discussion and further developments

Meta-analysis with marginalising the missing IPD

We proposed a new meta-analysis method for estimating the treatment-covariate interaction. As pointed out by Riley *et al.* (2008), separation of the treatment-covariate interaction into the across-trial and the within-trial effect is a clinically meaningful operation because these effects might be different of each other due to ecological bias and/or trial-level confounding. Indeed, for the hypertension data, we cannot conclude a negative interaction between the treatment effect on change in SBP and age, with the across-trial effect estimated by using the mean age in each trial. This is because the within-trial interaction, which was estimated by using age of individual patients, acted in the opposite direction of the across-trial interaction. Similarly for the home safety education data, the across-trial interaction between the intervention effect on provision of the stair gate and the proportion of male participants was very different from the within-trial interaction between the treatment effect and gender of individual patients. We here advocate that the MR model (2.2) or model (2.7) does not give sufficient evidence for the patient characteristics.

In the proposed method, we assume the IPD meta-analysis model for the missing IPD and then marginalise its density with respect to the missing IPD. These processes produce the log-likelihood (3.8) for AD available, and the use of this log-likelihood is useful to estimate the within-trial interaction even when all trials provide only AD. The simulation studies suggested that the proposed method provided the within-trial interaction effect estimates with moderately small RMSEs, and worked better when the variance in within-trial covariate distribution was large. The proposed method assumes that all trials are similar (exchangeable) to each other apart from having a separate baseline (intercept). This strong exchangeability assumption means that, conditional on the AD available (means and standard deviations of each group), the missing information (such as the within-trial interaction, γ_w , and the covariate effect, μ) can be informed approximately by meta-analysing the AD.

The proposed method is also simply extended to the case for mixture of IPD and AD. 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, Simmonds and Look, 2007; Ahmed, Sutton and Riley, 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. Through simulation studies, the proposed method provided smaller biases and smaller MSEs for estimator of the within-trial interaction, $\hat{\gamma}_w$, in comparison with the existing method by Riley *et al.* (2008), especially when the number of trials providing IPD was small. And also, simulation studies suggested how the biases and the MSEs for $\hat{\gamma}_w$ from the proposed method changed according to the number of trials providing IPD. These results could offer a useful guidance if one considers how many IPD trials should be collected to preserve a desired level of statistical power.

However, we recognise that the proposed 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 and within-trial interaction are fixed across trials. It would be useful to extend the method to random effects models to allow for heterogeneity if possible (Higgins, Thompson and Spiegelhalter, 2009), and also allow a trial-specific covariate effect (μ_i) and a trial-specific error variance (σ_{yi}^2). 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). Riley *et al.* (2012) 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 to this situation. Moreover, in the proposed method, we assume that the covariate is normally distributed. It would be necessary to discuss how sensible the results from the proposed method are with respect to this assumption. Finally, we only consider models for estimating one interaction, but of course in practice multiple interactions might be of interest. Nonetheless, where the assumed criteria are considered plausible or worth consideration in a sensitivity analysis, the proposed method is a promising method for meta-analysts faced with combining IPD and AD.

Meta-analysis based on simulated IPD

The SIPD method proposed offers a novel framework for meta-analysis, and is also flexible enough to estimate the treatment-covariate interaction whilst separating across-trial and

within-trial effects because it involves the scheme of marginalising the missing IPD. Through the application to the hypertension data, we demonstrated that the SIPD method provided results for the within-trial interaction closer to those from the full IPD analysis than the existing method 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 (2.5) 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 to model (2.5), can provide estimates and standard errors closer to those from a full IPD analysis. This is particularly true when given over 10 per cent and under 40 per cent 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.

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 in the cases when the proportion of patients with available IPD was under 10 per cent. 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 per cent of patients in the IPD trials, the SIPD method performed well.

Using the SIPD enables one to apply any IPD meta-analysis approaches, and could have a huge possibility to produce novel findings which is never provided by the existing meta-analysis methods. Through the application to the hypertension data, we used the SIPD method to meta-analyse patient subgroups within each trial identified age of patients. The estimates of the within-trial interaction and its standard errors for each patient subgroup could be utilised for a flexible trial design. To seek and find the further potential benefits of using the SIPD will be discussed as a future problem.

Appendices

Appendix A

The difficulty of using (3.16) is to integrate the density over a restricted sample space. Tsiatis (2006) gave a general calculation to solve such problems. We here brief the Tsiatis's (2006) approach, and describe how to integrate the normal density over the sample space that a sample mean and variance are fixed. This calculation follows a discussion by Pullin (1979) which proposed a method for generating random samples from a normal distribution with known sample mean and variance.

Let $X = (X_1, \dots, X_K)$ be K random variables, and assume there exists a dimensional-reduction transformation $h(X)$, that is a K' -dimensional variable ($K' < K$). Also, assume there exists a $(K - K')$ -dimensional variable $g(X)$ that

$$X \leftrightarrow \{h(X), g(X)\}$$

is one-to-one for all $h(X)$. Let $f(X)$ and $f(h(X), g(X))$ be the density of X and $(h(X), g(X))$ respectively. Consider random samples of X , $x = (x_1, \dots, x_K)$, and suppose that K' -dimensional summary statistics, $h(x)$, are only available. Then, an integration of $f(X)$ over a sample space with fixed $h(x)$ is equivalent to an integration of $f(h(x), g(X))$ with respect to $g(X)$; i.e.

$$\int_{h(x)} f(X) dX = \int f(h(x), g(X)) dg(X). \quad (\text{A.1})$$

If we consider X as normal random variables, $h(x)$ as a sample mean and variance of x , the desired likelihood (3.16) can be derived by using the relationship (A.1).

Let $x_1, \dots, x_K$ be independent random samples from a normal distribution with mean μ and variance σ^2 . The joint distribution of $(x_1, \dots, x_K)$ is given by

$$dF(x_1, \dots, x_K) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{k=1}^K (x_k - \mu)^2 \right] dx_1, \dots, dx_K.$$

We first consider the following Helmert's transformation:

$$y_1 = \frac{1}{K^{1/2}} \sum_{k=1}^K (x_k - K\mu), \quad (\text{A.2})$$

$$y_l = \frac{1}{[(K+1-l)(K+2-l)]^{1/2}} \sum_{k=1}^K x_k - (K+1-l)x_{l-1}, \quad l = 2, \dots, K.$$

The inverse of (A.2) is readily obtained as

$$x_1 = \frac{1}{K} \left[K\bar{x} - \sqrt{K(K-1)} y_2 \right], \quad (\text{A.3})$$

$$x_l = x_{l-1} + \frac{1}{(K+1-l)^{1/2}} \left[(K+2-l)^{1/2} y_l - (K-l)^{1/2} y_{l+1} \right], \quad l = 2, \dots, K.$$

Using (A.3), a known result is led as

$$dF(y_1, \dots, y_K) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp \left[-\frac{1}{2\sigma^2} \left(y_1^2 + \sum_{k=2}^K y_k^2 \right) \right] dy_1, \dots, dy_K \quad (\text{A.4})$$

and

$$\bar{x} = \mu + \frac{y_1}{\sqrt{K}}, \quad s^2 = \frac{1}{K} \sum_{k=2}^K y_k^2.$$

Then, $y_1, \dots, y_K$ independently follow a normal distribution with mean zero and variance σ^2 . We now consider a transformation of $y_1, \dots, y_K$ which follow the normal distribution as follows:

$$dF(y_2, \dots, y_K) \propto \frac{1}{(\sigma^2)^{(K-1)/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{k=2}^K y_k^2 \right] dy_2, \dots, dy_K. \quad (\text{A.5})$$

The solution requires slightly different treatment by whether K is odd or even. Now let

$$\eta = (K-1)/2 + \xi$$

where $\xi = 0$ if K is odd and $\xi = 1/2$ if K is even, and introduce the set of transformations

$$y_{2m} = \sqrt{e_m} \cos \nu_m, \quad (\text{A.6})$$

$$y_{2m+1} = \sqrt{e_m} \sin \nu_m, \quad m = 1, \dots, \eta - 2\xi$$

where $0 \leq \nu_m \leq 2\pi$ for all m . If K is odd, we use (A.6) directly. If not, we must add

$$y_N = \pm \sqrt{e_m} \quad (\text{A.7})$$

to (A.6) with equal probability $1/2$ for the $+$ and $-$. Substituting (A.6) and (A.7) into (A.5), we have

$$dF(e_1, \dots, e_\eta, \nu_1, \dots, \nu_{\eta-2\xi}) \propto \exp \left[-\frac{1}{2\sigma^2} \sum_{m=1}^{\eta} e_m \right] \prod_{i=1}^{\eta-2\xi} de_i d\nu_i [e_\eta^{-1/2} de_\eta]^{2\xi} \quad (\text{A.8})$$

while the sample variance is given by

$$s^2 = \frac{1}{N} \sum_{m=1}^{\eta} e_m. \quad (\text{A.9})$$

Finally, we introduce a transformation of $e_1, \dots, e_\eta$ into $(s^2, z_1, \dots, z_{\eta-1})$ as

$$e_m = K s^2 z_m \prod_{j=1}^{m-1} (1 - z_j), \quad m = 1, \dots, \eta - 1, \quad (\text{A.10})$$

$$e_\eta = K s^2 \prod_{j=1}^{\eta-1} (1 - z_j).$$

where each of the z_m , $m = 1, \dots, \eta - 1$, is located in an interval $(0,1)$. The Jacobian of this transformation is given by

$$\left| \frac{\partial(e_1, \dots, e_\eta)}{\partial(s^2, z_1, \dots, z_{\eta-1})} \right| = K^\eta (s^2)^{\eta-1} \prod_{j=1}^{\eta-2} (1 - z_j)^{\eta-j-1}. \quad (\text{A.11})$$

Substituting (A.10) and (A.11) into (A.8) and (A.9), we have

$$dF(s^2, z_1, \dots, z_{\eta-1}, \nu_1, \dots, \nu_{\eta-2\xi}) \propto$$

$$(s^2)^{(K-3)/2} \exp \left[-\frac{K s^2}{2\sigma^2} \right] d(s^2) \prod_{m=1}^{\eta-2\xi} d\nu_m \prod_{m=1}^{\eta-1} (1 - z_m)^{\eta-(m+\eta+1)} dz_m. \quad (\text{A.12})$$

Therefore, from (A.4) and (A.12), the density of (μ, σ^2) is given as follows except for terms unrelated to (μ, σ^2) :

$$\frac{1}{(\sigma^2)^{K/2}} \exp \left[-\frac{K}{2\sigma^2} \{(\bar{x} - \mu)^2 + s^2\} \right]. \quad (\text{A.13})$$

By using (A.13) in the context of the continuous meta-analysis data, we obtain (3.16):

Appendix B

In Step (2) of the SIPD method, the uncollected IPD must be drawn from the conditional distribution given available AD; 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), Lindqvist and Taraldsen (2005). They gave general formula to calculate the conditional expectation based on the conditional distribution given the sufficient statistics. Especially, an issue of sampling from the conditional distributions were considered. We here brief this approach, proposed by Lindqvist and Taraldsen (2005), for a simple case of univariate normal distribution.

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

$$\begin{aligned}\chi(U, \theta) &\equiv (\mu + \sigma U_1, \dots, \mu + \sigma U_K), \\ \tau(U, \theta) &\equiv (\mu + \sigma \bar{U}, \sigma S_U),\end{aligned}\tag{B.1}$$

where $\bar{U}$ and S_U^2 stand for the mean and the variance similarly defined to $\bar{X}$ 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 $t = (\bar{x}, s_x)$ and U , $\hat{\theta} \equiv \hat{\theta}(U, t)$ in which $\tau(U, \hat{\theta})$ is held is uniquely determined as follows.

$$\hat{\theta}(U, t) \equiv (\hat{\mu}(U, t), \hat{\sigma}(U, t)) = \left(\bar{x} - \frac{\bar{U}}{S_U} s_x, \frac{1}{S_U} s_x \right).\tag{B.2}$$

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

$$X_t = \chi(U, \hat{\theta}(U, t)) = \left(\bar{x} + \frac{U_1 - \bar{U}}{S_U} s_x, \dots, \bar{x} + \frac{U_K - \bar{U}}{S_U} s_x \right).\tag{B.3}$$

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 give the mean and the variance is as follows: (i) generate random numbers $u = (u_1, \dots, u_K)$ of U , (ii) substituting u and t to equation (A.3), we get

$$x_t = \left(\bar{x} + \frac{u_1 - \bar{u}}{s_u} s_x, \dots, \bar{x} + \frac{u_K - \bar{u}}{s_u} s_x \right).\tag{B.4}$$

References

- Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analysis using individual participant data: a database survey. *The British Medical Journal* 2012; **344**, DOI:10.1136/bmj.d7762.
- Amery A, Birkenhager W, Brixko P, Bulpitt C, Clement D, Deruyttere M, De Schaepdryver A, Dollery C, Fagard R, Forette F *et al.* Mortality and morbidity results from the European Working Party on high blood pressure in the elderly trial. *The Lancet* 1985; **1**:1349–1354.
- Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Statistics in Medicine* 2003; **22**:2591–2602.
- Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**:395–411.
- Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient versus group-level data meta-regression for the investigation of treatment effect modifiers: Ecological bias rears its ugly head. *Statistics in Medicine* 2002; **21**:371–387.
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Wiley, 2009.
- Cheng RCH. Generation of inverse Gaussian variates with given sample mean and dispersion. *Applied Statistics* 1984; **33**:309–316.
- Clamp M, Kendrick D. A randomised controlled trial of general practitioner safety advice for families with children under 5 years. *The British Medical Journal* 1998; **316**:1576–1579.
- Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; **10**:101–129.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B* 1977; **39**:1–37.
- Engen S, Lillegard M. Stochastic simulations conditioned on sufficient statistics. *Biometrika* 1997; **84**:235–240.

- Fog A. Sampling methods for Wallenius' and Fisher's noncentral hypergeometric distributions. *Communications in Statistics: Simulation and Computation* 2008; **37**:241–257.
- Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman and Hall, 1995.
- Gielen AC, McDonald EM, Wilson MEH, Hawang WT, Serwint JR, Andrews JS *et al*. Effects of improved access to safety counseling products and home visits on parents' safety practices: results of a randomized trial. *Archives Pediatric and Adolescent Medicine* 2002; **156**:33–40.
- Goldstein H, Yang M, Omar RZ, Thompson SG. Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistics* 2000; **49**:399–412.
- Gueyffier F, Bouitrie F, Boissel JP, Coope J, Cutler J, Ekblom T, Fagard R, Friedman L, Perry HM, *et al*. INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results. *Therapie* 1995; **50**:353–362.
- Haneuse S, Wakefield J. Hierarchical models for combining ecological and case-control data. *Biometrics* 2007; **63**:128–136.
- Haneuse S, Wakefield J. The Combination of Ecological and Case-Control Data. *Journal of the Royal Statistical Society: Series B* 2008; **70**:73–93.
- Heitjan DF, Rubin DB. Ignorability and coarse data. *The Annals of Statistics* 1991; **19**:2244–2253.
- Higgins JPT. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology* 2008; **37**:1158–1160.
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **21**:1539–1558.
- Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A* 2009; **172**:137–159.
- Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* 2001; **20**:2219–2241.
- Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data. *Statistics in Medicine* 2006; **25**:2136–2159.

- Jackson C, Best N, Richardson S. Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *Journal of the Royal Statistical Society: Series A* 2008; **171**:159–178.
- Jones D. Of medicine and meta-analysis. *Nature Reviews Drug Discovery* 2008; **7**:376–377.
- Kendrick D, Illingworth R, Woods A, Watts K, Collier J, Dewey M *et al.* Promoting child safety in primary care: a cluster randomised controlled trial to reduce baby walker use. *British Journal of General Practice* 2005; **55**:582–588.
- Kendrick D, Marsh P, Fielding K, Miller P. Preventing injuries in children: cluster randomised controlled trial in primary care. *The British Medical Journal* 1999; **318**:980–983.
- Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariate in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 2002; **55**:86–94.
- Lindqvist BH, Taraldsen G. Monte Carlo conditioning on a sufficient statistic. *Biometrika* 2005; **92**:451–464.
- Little RJA, Rubin DB. *Statistical Analysis with Missing Data, 2nd edition*. New York: John Wiley and Sons, 2002.
- 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.
- McDonald EM, Solomon B, Shields W, Serwint JR, Jacobsen H, Weaver NL *et al.* Evaluation of a Kiosk-based tailoring to promote household safety behaviours in an urban pediatric primary care practice. *Patient Education and Counseling* 2005; **58**:161–181.
- Morgenstern H. Uses of ecological analysis in epidemiologic research. *American Journal of Public Health* 1982; **72**:1336–1334.
- Nansel TR, Weaver N, Donlin M, Jacobsen H, Kreuter MW, Simons-Morton BG. Baby, be safe: the effect of tailored communications for pediatric injury prevention provided in a primary care setting. *Patient Education and Counseling* 2002; **46**:175–190.

- Neuhaus LA, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; **54**:638–645.
- Pullin DI. Generation of normal variates with given sample mean and variance. *Journal of Statistical Computation and Simulation* 1979; **9**:303–309.
- Robins JM, Wang N. Inference for imputation estimators. *Biometrika* 2000; **87**:113–124.
- Riley RD. Commentary: Like it and lump it? Meta-analysis using individual participant data. *International Journal of Epidemiology* 2010; **39**:1359–1361.
- Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods* 2010; **1**:2–19.
- Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *British Medical Journal* 2010; **340**:c221.
- Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Bouitrie F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine* 2008; **27**:1870–1893.
- Riley RD, Simmonds MC, Look MP. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology* 2007; **60**:431–439.
- Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Statistics in Medicine* 2008; **27**:6111–6136.
- Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *British Medical Journal* 2011; **342**:d549.
- Riley RD, Kauser I, Bland M, Thijs L, Staessen JA, Wang J, Gueyffier F, Deeks JJ. Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. (submitted).
- Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons, 1987.
- Rubin DB. A new perspective. in Wachter KW, Straf ML. (eds.), *The future of Meta-Analysis*, Russel Sage Foundation, New York, 1990; Chapter 14:155–165.

- Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ and Thompson S.G. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2005; **2**:209–217.
- Simmonds MC, Higgins JPT. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine* 2007; **26**:2982–2999.
- Sutton AJ, Higgins JPT. Recent development in meta-analysis. *Statistics in Medicine* 2008; **27**:625–650.
- Sutton AJ, Kendrick D, Coupland CAC. Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine* 2008; **27**:651–669.
- Staessen JA, Fagard R, Thijs L, Celis H, Arabidze GG, Birkenhager WH, Bulpitt CJ, de Leeuw PW, Dollery CT, Fletcher AE *et al.* 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* 1997; **350**:757–764.
- Steele RJ, Wang N, Raftery AE. Inference from multiple imputation for missing data using mixtures of normals. *Statistical Methodology* 2010; **7**:351–365.
- Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and the Health Professions* 2002; **25**:76–97.
- Sznajder M, Leduc S, Janvrin MP, Bonnin MH, Aegerter P, Baudier F *et al.* Home delivery of an injury prevention kit for children in four French cities: a controlled randomized trial. *Injury Prevention* 2003; **9**:261–265.
- Tanner MA, Wong WH. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* 1987; **82**:528–540.
- Tsiatis AA. *Semiparametric Theory and Missing Data*. Springer, 2006.
- Thompson SG, Higgins JPT. Can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* 2005; **365**:341–346.

- Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research* 1993; **2**:147–160.
- Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 2002; **21**:1559–1573.
- Thompson SG, Turner RM, Warn DE. Multilevel models for meta-analysis, and their application to absolute risk difference. *Statistical Methods in Medical Research* 2001; **10**:375–392.
- Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2000; **19**:3417–3432.
- Wakefield J, Salway R. A statistical framework for ecological and aggregate studies. *Journal of the Royal Statistical Society: Series A* 2001; **164**:119–137.
- Wakefield J, Haneuse A, Dobra A, Teeple E. Bayes computation for ecological inference. *Statistics in Medicine* 2011; **30**:1381–1396.
- Wakefield C. Ecological inference for 2×2 tables (with discussion). *Journal of the Royal Statistical Society: Series A* 2004; **167**:385–445.
- Wang JG, Staessen JA, Franklin SS, Fagard R, Gueyffier F. Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome. *Hypertension* 2005; **45**:907–913.
- Wang N, Robins JM. Large-sample theory for parametric multiple imputation procedure. *Biometrika* 1998; **85**:935–948.
- Watson M, Kendrick D, Coupland C, Woods A, Futers D, Robinson JL. Providing child safety equipment to prevent injuries: randomised controlled trial. *The British Medical Journal* 2005; **330**:178–181.
- Wei GCG, Tanner MA. A Monte Carlo implementation of the EM algorithm and the Poor Man's data augmentation. *Journal of the American Statistical Association* 1990; **85**:699–704.

Whitehead A, Omar RZ, Higgins JPT, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* 2001; **20**:2243–2260.

List of publications

Papers for journals (peer-reviewed)

- Yamaguchi Y, Sakamoto W, Goto M, Staessen JA, Wang J, Gueyffier F, Riley RD. Meta-analysis of a continuous outcome combining individual patient data and aggregate data: a method based on simulated individual patient data (submitted).
- Yamaguchi Y, Sakamoto W, Shirahata S, Goto M. An evaluation of treatment-covariate interaction in meta-analysis with marginalising the missing individual patient data. *Journal of the Japanese Society of Computational Statistics* (accepted).
- Yamaguchi Y, Okamura H, Dohi T. Approximate computation for posterior distribution with parameter correlation using variational Bayes. *Transactions of the Japan Society for Industrial and Applied Mathematics* 2011; 21:73–88 (in Japanese).
- Yamaguchi Y, Okamura H, Dohi T. Variational Bayesian approach for estimating parameters of a mixture of Erlang distributions. *Communications in Statistics: Theory and Methods* 2010; 39:2333–2350.

Papers for international conferences

- Yamaguchi Y, Sakamoto W, Shirahata S, Goto M. Reconstruction of individual patient data for meta-analysis via Bayesian approach. *Joint Meeting of the 2011 Taipei International Statistical Symposium and 7th Conference of the Asian Regional Section of the IASC: Abstract* 2011; 73–74. (2011.12, Taipei, Taiwan).
- Yamaguchi Y, Sakamoto W, Shirahata S, Goto M. A meta-analysis method based on simulated individual patient data. *Proceedings of the 58th Conference of the International Statistical Institute* 2011; CD-ROM. (2011.8, Dublin, Ireland).

Papers for domestic conferences

- Yamaguchi Y, Sakamoto W, Goto M. Meta-analysis of a binary outcome based on simulated individual patient data. *Proceedings of the 26th Annual Meeting of the Japanese Society of Computational Statistics* 2012; 91–94. (2012.5, Kagawa, Japan) (in Japanese).

Yamaguchi Y, Sakamoto W, Shirahata S, Goto M. Meta-analysis combining with individual patient data: a method based on simulated individual patient data. *Proceedings of the 2011 Annual Meeting of the Japanese Federation of Statistical Science Associations* 2011; 303. (2011.9, Fukuoka, Japan) (in Japanese).

Yamaguchi Y, Sakamoto W, Shirahata S, Goto M. Evaluation of the interaction effect in meta-analysis: simulation-based method. *Proceedings of the 25th Annual Meeting of the Japanese Society of Computational Statistics* 2011; 43–46. (2011.5, Hokkaido, Japan) (in Japanese).

Yamaguchi Y, Sakamoto W, Shirahata S, Goto M. Recovery of the data in clinical evaluation and its applications. *Proceedings of the 24th Symposium of the Japanese Society of Computational Statistics* 2010; 93–96. (2010.11, Osaka, Japan) (in Japanese).

Yamaguchi Y, Sakamoto W, Shirahata S, Goto M. Method of synthesizing statistics and meta-analysis based on simulation. *Proceedings of the 25th Annual Meeting of the Japanese Society of Computational Statistics* 2009; 163–164. (2009.5, Fukuoka, Japan) (in Japanese).

