STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:1870–1893
Published online 11 December 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3165

Meta-analysis of continuous outcomes combining individual patient data and aggregate data

Richard D. Riley[1,*,†], Paul C. Lambert[2], Jan A. Staessen[3], Jiguang Wang[4], Francois Gueyffier[5], Lutgarde Thijs[3] and Florent Boutitie[6]

[1]*Centre for Medical Statistics and Health Evaluation, Faculty of Medicine, University of Liverpool, Shelley's Cottage, Brownlow Street, Liverpool L69 3GS, U.K.*
[2]*Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, 2nd Floor, Adrian Building, University Road, Leicester LE1 7RH, U.K.*
[3]*Studies Coordinating Centre, Division of Hypertension and Cardiovascular Rehabilitation, Department of Cardiovascular Diseases, University of Leuven, Campus Gasthuisberg, Herestraat 49/702, B-3000 Leuven, Belgium*
[4]*Centre for Epidemiological Studies and Clinical Trials, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Ruijin 2nd Road 197, Shanghai 200025, China*
[5]*INSERM, CIC201, Lyon F-69000, France*
[6]*UMR 5558, Laboratoire Biostatistique Santé, Pierre-Bénite F-69495, France*

SUMMARY

Meta-analysis of individual patient data (IPD) is the gold-standard for synthesizing evidence across clinical studies. However, for some studies IPD may not be available and only aggregate data (AD), such as a treatment effect estimate and its standard error, may be obtained. In this situation, methods for combining IPD and AD are important to utilize all the available evidence. In this paper, we develop and assess a range of statistical methods for combining IPD and AD in meta-analysis of continuous outcomes from randomized controlled trials.

The methods take either a one-step or a two-step approach. The latter is simple, with IPD reduced to AD so that standard AD meta-analysis techniques can be employed. The one-step approach is more complex but offers a flexible framework to include both patient-level and trial-level parameters. It uses a dummy variable to distinguish IPD trials from AD trials and to constrain which parameters the AD trials estimate. We show that this is important when assessing how patient-level covariates modify treatment effect, as aggregate-level relationships across trials are subject to ecological bias and confounding. We thus develop models to separate within-trial and across-trials treatment–covariate interactions; this ensures that only IPD trials estimate the former, whilst both IPD and AD trials estimate the latter in addition to the pooled treatment effect and any between-study heterogeneity. Extension to multiple correlated outcomes is also considered. Ten IPD trials in hypertension, with blood pressure the continuous outcome of interest,

*Correspondence to: Richard D. Riley, Centre for Medical Statistics and Health Evaluation, Faculty of Medicine, University of Liverpool, Shelley's Cottage, Brownlow Street, Liverpool L69 3GS, U.K.
†E-mail: richard.riley@liv.ac.uk

Contract/grant sponsor: Department of Health's National Coordinating Centre for Research Capacity Development

Received 17 May 2007
Accepted 31 October 2007

Copyright © 2007 John Wiley & Sons, Ltd.

are used to assess the models and identify the benefits of utilizing AD alongside IPD. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: meta-analysis; individual patient data; aggregate data; treatment–covariate interaction; ecological bias

1. INTRODUCTION

In evidence-based medical research, meta-analysis methods combine the quantitative evidence across studies to produce results based on a whole body of research [1]. A traditional meta-analysis involves synthesis of aggregate data (AD) obtained from study publications or study authors; for continuous outcomes, such AD often include the mean treatment difference and its variance. Meta-analysis produces a weighted average of the AD across studies to give an overall measure of treatment effect [1]. A random-effects meta-analysis accounts for between-study heterogeneity in treatment effect [2], and meta-regression models [3] assess how study-level covariates explain this heterogeneity [4].

An alternative to the AD approach is a meta-analysis of individual patient data (IPD), where the raw data from each study are obtained and synthesized directly [5]. IPD meta-analysis is the 'gold-standard' as it has many advantages over AD [6]. For example, it allows one to obtain information unavailable from publications and apply sophisticated modelling techniques [7–10]. IPD also allows patient-level covariates (e.g. age) to be modelled, which enables one to assess how such covariates modify treatment effect, that is, to estimate the interaction between covariate and treatment [11]. This informs clinical decisions about how to tailor treatment strategies for the individual patient. It is preferable to estimate treatment–covariate interactions in an IPD meta-analysis rather than a meta-regression. Meta-regression often has low power to detect treatment–covariate interactions [12], as it assesses across-trials relationships between study-level summaries (e.g. mean age) and treatment effect, rather than within-trial relationships between patient-level values and treatment effect. Across-trials relationships are also prone to confounding as they are based on observational associations, unlike the experimental associations within randomized controlled trials (RCTs) [13]. Across-trials relationships may also truly differ from within-trial relationships [14], a phenomenon known as ecological bias [15].

IPD meta-analyses are increasing [16], but practitioners may face the practical problem of obtaining IPD from only a proportion of studies [17]. This may be resource related, for example, due to lack of time, but is often related to the study authors themselves, who may not be contactable, may have lost their IPD, or may be unwilling to collaborate. A review of 175 applied IPD meta-analysis articles encouragingly found that 58 per cent obtained IPD from 90 per cent or more of the total number of studies [17]. Yet, also 29 per cent of the articles only obtained IPD for less than 80 per cent of the studies. In such situations, results from an IPD-only meta-analysis may be biased if unavailability of IPD is related to the study results [18]. It may thus help to supplement the available IPD with AD for those studies where IPD are not available. For example, a review in multiple myeloma involved 20 studies (4930 patients) with IPD and seven other studies (1703 patients) with AD [19]. A key statistical question, therefore, is how does one suitably combine IPD and AD in meta-analysis, especially when treatment–covariate interactions are of interest? Surprisingly few papers discuss this issue. A review [17] found three unpublished (Richardson *et al.*, Best *et al.*, Collete *et al.*, *personal communications*) and five published [5, 20–23] articles,

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

and we are aware of two others published subsequently [24, 25]. These 10 articles discuss four methods for combining IPD and AD. The simplest and most common is the two-step method, where the IPD are first reduced to AD and then a standard meta-analysis of AD is employed [5, 16]. More sophisticated methods use one-step models that distinguish between IPD and AD trials [22, 23]. For example, Goldstein *et al.* [22] use a multi-level model that uses a dummy variable to distinguish between IPD and AD trials. However, there has been little statistical assessment of this approach, with only one clinical application [24].

In this paper, we develop and empirically assess statistical models that use the Goldstein framework to combine IPD and AD in meta-analysis of continuous outcomes from RCTs. We also introduce related two-step models and show how the approaches can be used to estimate treatment–covariate interactions in relation to a pooled treatment effect and between-study heterogeneity, using a combination of IPD and AD. In Section 2, we introduce a data set in hypertension where IPD are available from 10 trials, with blood pressure the continuous outcome of interest. This data set is used to generate scenarios involving IPD and AD, so that subsequent models developed can be empirically assessed. In Section 3, we then introduce one-step and two-step meta-analysis models for estimating a pooled treatment effect, and we outline the basic framework for combining IPD and AD. In Section 4 the models are extended to include treatment–covariate interactions, and we show the importance of separating within-trial and across-trials relationships. Extensions to multiple covariates and multiple outcomes are considered in Section 5, and Section 6 provides a discussion of our work, with suggestions for further research.

2. THE HYPERTENSION DATA

Wang *et al.* [26] performed a quantitative overview of trials in hypertension to investigate to what extent lowering of systolic blood pressure (SBP) and diastolic blood pressure (DBP) contributed to cardiovascular prevention. They selected RCTs 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 [27] or at the Studies Coordinating Centre in Leuven (Belgium) [28–30]. Ten trials were ultimately included, and these provided IPD for a total of 28 581 patients. The mean age, proportion male, and mean baseline SBP and DBP values are shown in Table I, and the groups appear to be well balanced in each trial. Interestingly there is large variation in the mean age of patients across trials, ranging from about 42 to 75 years. Similarly, the proportion male varies considerably. Table I also shows the mean treatment effect on change in blood pressure (baseline minus follow-up) in each trial, with negative estimates indicating a beneficial treatment effect. Both unadjusted and adjusted treatment effects are shown, with the latter adjusted for baseline values as is recommended (see Section 3.1.1) [31]. In every trial the active treatment reduces both SBP and DBP more than placebo on average, and it is clearly clinically important to assess how patient covariates modify this treatment effect.

The hypertension data will be used in this paper to demonstrate and critically assess the models developed; those interested in more clinical conclusions are referred elsewhere [26]. We will analyse the effect of treatment on blood pressure and consider how age and sex modify treatment effectiveness. To imitate situations involving IPD for some trials and only AD for others, we generated scenarios where we assumed that only a proportion of the trials (ranging from 0 to 90 per cent) gave their IPD and that other trials just gave AD as presented in Table I, which is typical of the AD available to meta-analysts in practice. In each scenario we will fit: (i) models that use

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Table I. Summary of the 10 trials included in the meta-analysis of Wang *et al.* [26].

ID	Trial name*	Number of patients		Age (years)		Sex (0 = female, 1 = male)		SBP baseline (mmHg)		DBP baseline (mmHg)		Unadjusted treatment effect (baseline − follow-up)			Adjusted† treatment effect (baseline − follow-up)		
		Control	Treatment	Control mean (sd)	Treatment mean (sd)	Control per cent male	Treatment per cent male	Control mean (sd)	Treatment mean (sd)	Control mean (sd)	Treatment mean (sd)	SBP mean (var)	DBP mean (var)	Correlation (mean SBP, mean DBP)	SBP mean (var)	DBP mean (var)	Correlation (mean SBP, mean DBP)
1	ATMH	750	780	42.36 (5.34)	42.17 (5.39)	70.00	69.36	153.05 (15.73)	152.28 (15.25)	100.22 (8.65)	100.08 (8.75)	−6.13 (1.02)	−2.88 (0.40)	0.75	−6.66 (0.72)	−2.99 (0.27)	0.78
2	HEP	199	150	69.57 (5.39)	69.71 (5.18)	37.19	32.67	191.55 (17.64)	189.94 (16.15)	85.08 (7.06)	85.80 (8.78)	−13.23 (5.86)	−8.33 (1.71)	0.37	−14.17 (4.73)	−7.87 (1.44)	0.45
3	EWPHE	82	90	74.11 (8.69)	72.64 (7.99)	20.73	25.56	178.23 (15.06)	177.33 (15.85)	91.70 (1.45)	91.61 (1.56)	−12.68 (10.53)	−6.01 (1.76)	0.58	−12.88 (10.31)	−6.01 (1.77)	0.59
4	HDFP	2371	2427	41.54 (5.48)	41.58 (5.53)	53.73	54.64	151.00 (19.53)	151.68 (19.83)	101.15 (9.65)	101.18 (9.54)	−9.13 (0.42)	−5.13 (0.14)	0.76	−8.71 (0.30)	−5.11 (0.10)	0.77
5	MRC-1	3445	3546	45.17 (5.86)	45.38 (6.00)	59.01	58.83	156.65 (15.96)	156.60 (16.09)	97.86 (5.75)	97.92 (5.74)	−8.62 (0.19)	−4.68 (0.06)	0.61	−8.70 (0.14)	−4.64 (0.05)	0.66
6	MRC-2	1337	1314	70.43 (2.72)	70.41 (2.74)	41.81	41.25	182.13 (12.73)	182.19 (12.63)	83.39 (8.96)	83.44 (8.90)	−10.65 (0.72)	−5.59 (0.24)	0.46	−10.60 (0.58)	−5.56 (0.18)	0.49
7	SHEP	2371	2365	71.54 (6.68)	71.64 (6.72)	42.68	43.72	170.12 (9.24)	170.49 (9.50)	76.48 (9.06)	76.81 (9.17)	−11.51 (0.31)	−4.14 (0.09)	0.48	−11.36 (0.30)	−3.98 (0.27)	0.50
8	STOP	131	137	75.90 (3.95)	76.00 (3.75)	24.43	27.01	194.15 (11.16)	194.68 (12.21)	91.37 (2.40)	91.25 (2.63)	−18.18 (6.28)	−6.45 (1.35)	0.60	−17.93 (5.82)	−6.54 (1.31)	0.61
9	Sy-Chi	1139	1252	66.77 (5.67)	66.42 (5.34)	63.65	65.02	170.25 (11.41)	170.73 (10.90)	85.89 (6.85)	86.27 (6.65)	−6.84 (0.49)	−2.29 (0.13)	0.41	−6.55 (0.41)	−2.08 (0.11)	0.45
10	Sy-Eur	2297	2398	70.21 (6.67)	70.26 (6.73)	33.83	32.53	173.94 (10.09)	173.75 (9.86)	85.48 (5.91)	85.47 (5.83)	−10.18 (0.21)	−3.49 (0.05)	0.52	−10.26 (0.20)	−3.49 (0.04)	0.51

SBP, systolic blood pressure; DBP, diastolic blood pressure; sd, standard deviation; var, variance.

*Trial names are consistent with Wang *et al.* [26], where further details and trial publications can be found.

†Adjusted for baseline blood pressure values.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

only the IPD trials available and also (ii) models that use both IPD and AD trials. In both parts (i) and (ii), the analyses will be run for each possible combination of IPD and AD trials, with the results then averaged across analyses. For example, in the scenario involving 50 per cent of trials providing IPD (i.e. five IPD trials and five AD trials) in part (ii) we will perform 252 analyses, one for each combination, of which five trials provide IPD and five provide AD; we will then average across results from all 252 analyses. In each scenario, the results obtained will be compared with those from a meta-analysis of IPD from all 10 trials (i.e. a 100 per cent IPD analysis), allowing us to empirically assess the performance of the models and identify the value of combining IPD and AD in practice.

3. METHODS FOR ESTIMATING A POOLED TREATMENT EFFECT

In this section we introduce meta-analysis models for estimating a pooled treatment effect across studies, using notation similar to other related IPD articles [9, 11]. We describe both a two-step and a one-step approach and consider when all trials provide IPD, and also when a proportion of trials provide only AD in the form of their treatment effect estimate and its variance. The models are then assessed using the hypertension data.

3.1. Two-step approach

3.1.1. All trials provide IPD. Consider that there is one continuous outcome response of interest, say SBP after treatment, and assume that there is a treatment group (T) and a control group (C) in each trial ($i=1$–N), with n_{iT} and n_{iC} patients in each group. Thus, the jth patient provides their SBP after treatment, which we denote by y_{ij}. Given IPD for each trial, the common approach for meta-analysis is a two-step method [5, 16], where the IPD are first analysed separately in each trial and then the trial parameter estimates are combined in a standard meta-analysis. For example, in the first step, a common IPD model to fit to each of the i trials separately is

$$\begin{aligned} y_{ij} &= \phi_i + \theta_i x_{ij} + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathrm{N}(0, \sigma_i^2) \end{aligned} \tag{1}$$

where ϕ_i is the fixed trial effect, x_{ij} is coded 0/1 to denote control/treatment group, θ_i is the underlying treatment effect in study i, and σ_i^2 is the residual variance of the responses in trial i after accounting for the treatment effect. Model (1) and all subsequent models in the paper can be estimated using restricted maximum likelihood (REML) within suitable software for mixed models, for example, SAS Proc Mixed [32]. From this, the treatment effect estimate $\hat{\theta}_i$ and its variance $V(\hat{\theta}_i)$ can be obtained for each trial. If the continuous outcome of interest is a change from baseline (e.g. SBP at baseline minus SBP after treatment) then it is appropriate to extend model (1) to adjust for baseline SBP responses, y_{0ij}, in each trial [31]. This is done by changing ϕ_i to $\phi_i + \beta_i y_{0ij}$, and it generally leads to more efficient treatment effect estimates [31], as seen in Table I.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

In the second part of the two-step approach, the $\hat{\theta}_i$ values are then suitably combined across trials, for example, in a standard random-effects meta-analysis model, as follows:

$$\begin{aligned} \hat{\theta}_i &= \theta_i + \varepsilon_i \\ \theta_i &= \theta + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon_i &\sim \mathrm{N}(0, V(\hat{\theta}_i)) \end{aligned} \tag{2}$$

In model (2), the $V(\hat{\theta}_i)$ estimates are assumed to be known, which is a common assumption in the meta-analysis field [1], and u_i denotes a random effect indicating that the treatment effect in the ith trial, θ_i, is normally distributed about a pooled treatment effect, θ, with between-study variance, τ^2. The pooled treatment effect estimate ($\hat{\theta}$) will be a weighted average of the $\hat{\theta}_i$ s, with trial weights equal to the inverse of $V(\hat{\theta}_i) + \hat{\tau}^2$ [2]. Setting τ^2 to zero reduces model (2) to a fixed-effects meta-analysis, although we take a random-effects approach in this paper as heterogeneity in treatment effect often exists in practice.

3.1.2. Some trials provide IPD; some trials provide only AD. Consider now that N_{IPD} trials provide IPD and N_{AD} trials provide only AD in the form of their treatment effect estimate ($\hat{\theta}_i$) and its variance ($V(\hat{\theta}_i)$), where the total number of trials (N) in the meta-analysis is given by $N = N_{\mathrm{IPD}} + N_{\mathrm{AD}}$. The two-stage approach can easily accommodate such AD trials. Essentially, one would fit model (1) separately to each IPD trial and obtain their $\hat{\theta}_i$ and $V(\hat{\theta}_i)$ estimates. Then model (2) can be applied to all the available $\hat{\theta}_i$ and $V(\hat{\theta}_i)$, from both IPD and AD trials, to estimate the pooled treatment effect, θ, and the between-study variance, τ^2.

3.2. One-step approach

3.2.1. All trials provide IPD. As an alternative to the two-stage approach, one can meta-analyse IPD trials in a one-step approach using a mixed model, such as follows:

$$\begin{aligned} y_{ij} &= \phi_i + \theta_i x_{ij} + \varepsilon_{ij} \\ \theta_i &= \theta + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon_{ij} &\sim \mathrm{N}(0, \sigma_i^2) \end{aligned} \tag{3}$$

As before, ϕ_i is a fixed trial effect; u_i is a random effect indicating that the treatment effect in the ith trial, θ_i, is normally distributed about a pooled treatment effect, θ, with between-study variance, τ^2; σ_i^2 is the residual variance of the responses in trial i after accounting for the treatment effect, and τ^2 represents the unexplained heterogeneity in treatment effect across trials. Model (3) thus makes the same assumptions as models (1) and (2) combined, and indeed it should give very similar results for $\hat{\theta}$ and $\hat{\tau}^2$ [33, 34]. Slight differences may arise as the one-step method estimates the σ_i^2's simultaneously alongside $\hat{\theta}$ and $\hat{\tau}^2$ in model (3), whilst the two-step method estimates each σ_i^2 separately in model (1) and independent of $\hat{\theta}$ and $\hat{\tau}^2$ in model (2). As discussed for model (1),

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

if change from baseline is the continuous outcome of interest, then model (3) should be extended to adjust for baseline values by replacing ϕ_i with $\phi_i + \beta_i y_{0ij}$.

Model (3) forms the basis for further one-step models in this paper. We note, though, that it could be simplified, for example, by assuming a common residual variance in each trial (i.e. $\sigma_i^2 = \sigma^2$) or by assuming a fixed, rather than a random, treatment effect across trials (i.e. $\theta_i = \theta$ for all i, so that $\tau^2 = 0$ and $u_i = 0$). It could also be made more complex by assuming that the ϕ_i are randomly distributed about an overall trial effect [4]; however, this is controversial in the meta-analysis field [9], as the available trials are then assumed to be a random sample from a distribution of trials; hence, it is not considered further here.

3.2.2. Some trials provide IPD; some trials provide only AD. Model (3) can be extended to allow IPD trials to be combined with AD trials that provide their treatment effect estimate ($\hat{\theta}_i$) and its variance ($V(\hat{\theta}_i)$), using a few modelling tricks based on Goldstein *et al.* [22]. Firstly, for the IPD trials we continue to set the patient y_{ij} values as the response variable; however, for each AD trial we assume that there is only one patient and set the response variable to be $\hat{\theta}_i$. Similarly, for each IPD trial we continue to estimate their residual variance, σ_i^2, but for each AD trial we set the residual variance to be known as $V(\hat{\theta}_i)$. Finally, we include in the model a dummy variable D_i to distinguish between responses from IPD trials ($D_i = 1$) and responses from AD trials ($D_i = 0$). This ensures that only the IPD trials estimate their trial effect, ϕ_i, and their residual variance, σ_i^2, but allows both the IPD trials and the AD trials to estimate the pooled treatment effect, θ, and the between-study variance, τ^2. The model can be expressed as follows:

$$\begin{aligned} y_{ij}^* &= D_i \phi_i + \theta_i x_{ij} + \varepsilon_{ij}^* \\ \theta_i &= \theta + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon_{ij}^* &\sim \mathrm{N}(0, V_i^*) \end{aligned} \tag{4}$$

For each IPD trial, $y_{ij}^* = y_{ij}$ and $V_i^* = \sigma_i^2$ as in model (3). For each AD trial, there is only one response ($j = 1$) and we set $x_{i1} = 1$, $y_{i1}^* = \hat{\theta}_i$, and $V_i^* = V(\hat{\theta}_i)$, with the latter assumed to be known. Model (4) is essentially an amalgam of IPD model (3) and AD model (2), which allows them to be fitted simultaneously. The necessary SAS code is available on request.

3.3. Application to the hypertension data

Consider now application to the hypertension data in each of the generated scenarios described in Section 2, with change in SBP (before minus after treatment) the outcome of interest. In each scenario, for the AD trials we assumed that the treatment effect for change in SBP, $\hat{\theta}_i$, and its variance, $V(\hat{\theta}_i)$, were available for each trial, with $\hat{\theta}_i$ adjusted for baseline SBP. There is a small difference between unadjusted and adjusted $\hat{\theta}_i$ values (Table I). The adjusted estimates are derived from the correct method for analysing change from baseline [31], which one would hope trial authors had implemented. The impact of having only unadjusted $\hat{\theta}_i$ values for AD trials will be discussed in Section 4.3.3. For the analysis of IPD trials in each scenario, SBP after treatment was taken as the response variable and the analyses also correctly adjusted for baseline SBP, as described in Sections 3.1.1 and 3.2.1; their standardized residuals appeared close to normally

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

distributed in each trial. In each scenario, meta-analysis was performed using the one-step model (4) to each combination of trials providing IPD and trials providing AD. The results for each scenario, averaged across all possible combinations, are given in Table II.

In the scenario where IPD are available from all 10 trials, the pooled treatment effect estimate was −10.16 (SE 0.93), indicating that the treatment is significantly effective in reducing SBP by, on average, 10.16 mmHg more than placebo. However, there is also a large between-study variance of 7.13, indicating that the treatment effect varies much across the trials. This is perhaps expected given the different patient characteristics across trials, and this will be investigated in Section 4. In the scenario involving only AD from all trials, the results obtained are almost identical to the 100 per cent IPD analysis, highlighting why AD in the form of $\hat{\theta}_i$ and $V(\hat{\theta}_i)$ are considered sufficient for meta-analysis when only the pooled treatment effect and between-study variance are of interest [33, 34].

In the scenarios involving IPD and AD trials, we are interested in how estimates differ from those of the 100 per cent IPD analysis. Model (4) combines IPD and AD to produce average estimates that are almost identical to those from the 100 per cent IPD analysis (Table II). The benefit of combining IPD with AD also increases as the proportion of IPD trials decreases. The IPD-only analyses (i.e. those using only available IPD trials) overestimate the treatment effect and between-study variance on average; they also increase the standard error of treatment effect, with the exception of the 10 per cent IPD scenario where the standard error is artificially small in the IPD-only analyses as the between-study heterogeneity is ignored. These findings show that model (4) performs well and that combining AD and IPD is potentially important in practice. Rather than model (4), we also applied the two-step approach of models (1) and (2) to combine IPD and AD; the results obtained were practically identical to those presented in Table II.

4. METHODS FOR ESTIMATING TREATMENT–COVARIATE INTERACTIONS

Clinicians wish to know how treatment effect is modified by patient characteristics. Furthermore, if there is between-study heterogeneity it is important to explain its cause [35] and thus reduce $\hat{\tau}^2$ in the above models. We thus now extend the models in Section 3 to include a patient-level covariate and show how to estimate a treatment–covariate interaction given IPD from all trials or a mixture of IPD and AD trials. The models again take a two-step or a one-step approach, and AD trials are assumed to provide their treatment effect estimate, its variance, and covariate means. This is typical AD available in practice. In each trial we also assume that treatment groups are balanced in the covariate under investigation, as is usually true for RCTs. However, the impact of unbalanced groups and having only unadjusted AD is considered further in Section 4.3.3.

4.1. Two-step approach

4.1.1. All trials provide IPD. Using only within-trial information: Let z_{ij} be a patient-level covariate (e.g. the age of patient j in trial i) observed for all patients in each trial, and let m_i be the mean of this covariate in trial i (e.g. the mean age in trial i). Simmonds and Higgins [11] suggest a two-step method for estimating treatment–covariate interactions, and we extend their

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Table II. Model (4) parameter estimates when assessing change in systolic blood pressure (before–after treatment) from the hypertension data in a range of different scenarios.

	Per cent of trials providing IPD Trials used in the analysis											
	100 per cent	90 per cent		70 per cent		50 per cent		30 per cent		10 per cent		0 per cent
Parameter estimate	IPD trials only	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	AD trials only
Pooled treatment effect, $\hat{\theta}$	−10.16	−10.18	−10.16	−10.23	−10.17	−10.28	−10.17	−10.38	−10.17	−10.78	−10.17	−10.17
Standard error of $\hat{\theta}$	0.93	0.99	0.93	1.16	0.93	1.42	0.93	1.93	0.93	1.55	0.93	0.93
Between-study variance, $\hat{\tau}^2$	7.13	7.31	7.16	7.80	7.17	8.47	7.17	9.51	7.18	NA	7.18	7.18

Values shown are the average estimates across all possible combinations of IPD trials and AD trials within each scenario.

The average standard error of $\hat{\theta}$ was calculated by taking the square root of the average variance of $\hat{\theta}$.

IPD, individual patient data; AD, aggregate data; NA, not applicable, as $\hat{\tau}^2$ is not estimatable given only one trial.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

approach here. As in Section 3.1, the first stage involves fitting a model to the IPD from each trial separately as follows:

$$\begin{aligned} y_{ij} &= \phi_i + \alpha_i x_{ij} + \mu_i z_{ij} + \gamma_i x_{ij} z_{ij} + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathrm{N}(0, \sigma_i^2) \end{aligned} \tag{5}$$

The parameters here are as specified for model (1), but with the addition of a treatment–covariate interaction term, γ_i, which indicates the mean change in treatment effect for a one-unit increase in z_{ij} for trial i; also introduced are μ_i, the mean change in control group response for a one-unit increase in z_{ij}, and also α_i, the mean treatment effect when $z_{ij}=0$. Incorporation of μ_i takes into account any imbalance in the covariate distribution between treatment groups, and inclusion of further confounding factors can be made, which may be more important for observational studies than RCTs. Note that z_{ij} could be a binary or a continuous covariate, and model (5) assumes a linear relationship between a continuous z_{ij} and the patient response in each treatment group. Estimation of model (5) produces a treatment–covariate interaction estimate, $\hat{\gamma}_i$, and its variance, $V(\hat{\gamma}_i)$, for each trial. The second stage involves synthesizing the $\hat{\gamma}_i$ values across trials in a standard meta-analysis model, for example, as follows:

$$\begin{aligned} \hat{\gamma}_i &= \gamma_{\mathrm{W}} + \varepsilon_i \\ \varepsilon_i &\sim \mathrm{N}(0, V(\hat{\gamma}_i)) \end{aligned} \tag{6}$$

Here, $V(\hat{\gamma}_i)$ is assumed to be known and the model estimates the pooled treatment–covariate interaction, γ_{W}, where 'W' denotes that it is based on only within-trial information (see below). Thus, $\hat{\gamma}_{\mathrm{W}}$ denotes how the treatment effect changes for a one-unit increase in z_{ij}. Model (6) can be extended to allow γ_{W} to vary across trials if necessary, although this specific issue requires further research in the meta-analysis field.

Using within-trial and across-trials information: A limitation of model (5) is that γ_i cannot be estimated in trials where z_{ij} is identical for all patients; for example, when all patients in the trial are male there is no *within-trial information* regarding how sex modifies treatment effect. Such trials would thus be excluded from model (6); however, they could be combined with the other trials in a meta-regression that assesses the across-trials relationship between treatment effect and mean covariate value m_i (e.g. proportion male). Essentially, one would fit model (1) for each trial separately and then regress $\hat{\theta}_i$ against m_i, weighting by the inverse of $V(\hat{\theta}_i)+\hat{\tau}^2$ as follows:

$$\begin{aligned} \hat{\theta}_i &= \alpha_i + \gamma_{\mathrm{A}} m_i + \varepsilon_i \\ \alpha_i &= \alpha + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon_i &\sim \mathrm{N}(0, V(\hat{\theta}_i)) \end{aligned} \tag{7}$$

Model (7) is a meta-regression model and the parameters are as defined previously, with the addition of α, which denotes the pooled treatment effect for trials with $m_i=0$, and γ_{A}, which denotes how treatment effect changes across trials for a one-unit increase in m_i, with $\hat{\gamma}_{\mathrm{A}}$ thus based solely on across-trials information ('A'). Caution is required when interpreting $\hat{\gamma}_{\mathrm{A}}$, as it estimates the across-trials interaction and not the within-trial interaction, γ_{W}. Across-trials relationships are prone to confounding and ecological bias and thus may not reflect within-trial relationships. For this reason,

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

meta-regression has been criticised [15], but the approach does have increased power when m_i varies considerably across trials (e.g. some trials contain only males, others only females), so that the effect of a covariate on treatment shows up in differences in treatment effect across trials [11]. In this situation, to improve efficiency, one option is to mix within-trial and across-trial estimates, e.g. $\hat{\gamma}_A$ from model (7) could be pooled with $\hat{\gamma}_W$ from model (6) in a further meta-analysis, suitably accounting for $V(\hat{\gamma}_A)$ and $V(\hat{\gamma}_W)$. This would produce an amalgam of the within-trial and across-trial interactions ($\hat{\gamma}_{WA}$, say), improving efficiency compared with the separate estimates alone but at the risk of introducing bias by incorporating $\hat{\gamma}_A$. Unless there is good justification to assume $\gamma_W = \gamma_A$, we generally recommend making inferences about treatment–covariate interactions using only the within-trials estimate, $\hat{\gamma}_W$, to avoid the threat of bias from $\hat{\gamma}_A$.

4.1.2. Some trials provide IPD; some trials provide only AD. AD trials providing $\hat{\theta}_i, V(\hat{\theta}_i)$, and m_i can be combined with IPD trials in model (7) so as to help estimate the across-trials interaction, γ_A. They cannot help estimate the within-trial interaction, γ_W, in model (6), unless one makes the strong assumption that $\gamma_W = \gamma_A$ as discussed; if one does, then a further meta-analysis can be used to combine $\hat{\gamma}_W$ (from the meta-analysis of model (6) using just IPD trials) with $\hat{\gamma}_A$ (from the meta-regression of model (7) using both IPD and AD trials). This will improve efficiency, especially when the proportion of AD trials is large; however, generally it seems most sensible to use only $\hat{\gamma}_W$ to make inferences about the true treatment–covariate interaction within trials, even if there are only a few IPD trials. Section 4.3 explores this issue further.

4.2. One-step approach

4.2.1. All trials provide IPD. Separating within-trial and across-trials information: Treatment–covariate interactions can also be estimated using a one-step meta-analysis approach that extends model (3). Care is needed, though, to separate within-trial and across-trial relationships, an issue in meta-analysis [4, 5] and also related contexts [36–39]. The literature suggests centring z_{ij} by the mean covariate value, m_i, in each trial, whilst including an interaction term between m_i and treatment effect, as follows:

$$\begin{aligned} y_{ij} &= \phi_i + \alpha_i x_{ij} + \mu_i z_{ij} + \gamma_W x_{ij}(z_{ij} - m_i) + \gamma_A x_{ij} m_i + \varepsilon_{ij} \\ \alpha_i &= \alpha + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon_{ij} &\sim \mathrm{N}(0, \sigma_i^2) \end{aligned} \tag{8}$$

The parameters in model (8) are as defined previously, and it separates the pooled within-trial treatment–covariate interaction, γ_W, from the across-trials interaction, γ_A. Applying model (8) to the hypertension data gives independent $\hat{\gamma}_W$ and $\hat{\gamma}_A$ (see Section 4.3), and other applications also indicate this [38], even when the number of subjects per cluster is small. Note that, consistent with previous models, model (8) specifies a linear relationship between a continuous z_{ij} and treatment effect, and γ_W is assumed to be common across trials. Such assumptions can be changed; but here we take model (8) as our basic one-step framework for separating within-trial and across-trial interactions. It can be extended to include baseline values and confounding factors if necessary, as discussed for model (5). An appealing property of model (8) is that it simultaneously estimates γ_W and γ_A alongside the remaining between-study variance ($\hat{\tau}^2$) and the average treatment effect

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

in a trial with covariate mean m_i ($\hat{\alpha}+\hat{\gamma}_A m_i$). The latter is maybe most informative when m_i is set to m, the mean z_{ij} across all patients in all trials, to give an average treatment effect similar in interpretation to $\hat{\theta}$ from a standard random-effects meta-analysis (model (2)). These estimates are also obtainable from the two-stage approach, but more tediously a couple of two-stage models is needed; one to obtain $\hat{\gamma}_W$ (model (5) followed by model (6)) and another to obtain $\hat{\tau}^2$, $\hat{\alpha}$, and $\hat{\gamma}_A$ (model (1) followed by model (7)).

Mixing within-trial and across-trial information: Estimation of model (8) allows a test against the null hypothesis that $H_0: \gamma_W = \gamma_A$, i.e. the fact that the within-trial and across-trials interactions are equal. However, the test will have low power as the standard error of $\hat{\gamma}_A$ will often be large, due to $\hat{\gamma}_A$ being dependent on the number of trials [12], which is usually small; hence, the test result should be treated with caution. If one can justify setting $\gamma_W = \gamma_A = \gamma_{WA}$, then model (8) simplifies to

$$\begin{aligned} y_{ij} &= \phi_i + \alpha_i x_{ij} + \mu_i z_{ij} + \gamma_{WA} x_{ij} z_{ij} + \varepsilon_{ij} \\ \alpha_i &= \alpha + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon_{ij} &\sim \mathrm{N}(0, \sigma_i^2) \end{aligned} \tag{9}$$

Estimation of model (9) gives $\hat{\gamma}_{WA}$, a weighted average of $\hat{\gamma}_W$ and $\hat{\gamma}_A$ from model (8). We recommend that whenever $\hat{\gamma}_{WA}$ is reported, $\hat{\gamma}_W$ and $\hat{\gamma}_A$ should also be given so as to make any differences between the within-trial and across-trial interactions transparent.

4.2.2. Some trials provide IPD; some trials provide only AD. Models (8) and (9) can be extended to combine IPD trials with AD trials that give their treatment effect estimate ($\hat{\theta}_i$), its variance ($V(\hat{\theta}_i)$), and covariate means m_i. For example, using the framework introduced in model (4), model (8) can be extended as follows:

$$\begin{aligned} y^*_{ij} &= D_i \phi_i + \alpha_i x_{ij} + D_i \mu_i z^*_{ij} + D_i \gamma_W x_{ij} (z^*_{ij} - m_i) + \gamma_A x_{ij} m_i + \varepsilon^*_{ij} \\ \alpha_i &= \alpha + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon^*_{ij} &\sim \mathrm{N}(0, V_i^*) \end{aligned} \tag{10}$$

For each IPD trial, $D_i = 1$, $y^*_{ij} = y_{ij}$, $V_i^* = \sigma_i^2$, and $z^*_{ij} = z_{ij}$. For each AD trial, there is only one response ($j = 1$) and we set $D_i = 0$, $x_{i1} = 1$, $y^*_{i1} = \hat{\theta}_i$, $V_i^* = V(\hat{\theta}_i)$, and $z^*_{ij} = m_i$ with $V(\hat{\theta}_i)$ assumed to be known. The model framework ensures that the AD trials help estimate only the across-trials parameters (α, γ_A, and τ^2), whereas the IPD trials help estimate all the parameters. Both IPD and AD trials contribute to a test of $H_0: \gamma_W = \gamma_A$, and the model assumptions can be modified as discussed for other models. Model (10) is essentially an amalgamation of IPD model (8) and meta-regression model (7) that allows them to be fitted simultaneously with suitable partition of the within-trial and across-trials interactions. The SAS Proc Mixed code to fit model (10) is provided in the Appendix.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

4.3. Application to the hypertension data

Consider again the hypertension data and, in each of the scenarios described in Section 2, we now assess how age and sex modify the treatment effect on SBP. For simplicity, we assess each covariate independently; Section 5.1 makes extension to multiple covariates. Model (10) was fitted for each covariate, and the within-trial (γ_W) and across-trials (γ_A) interactions were estimated. The results are shown in Table III, and we note that the equivalent two-stage models gave almost identical results to the one-step analyses.

4.3.1. Within-trial versus across-trials interaction. Consider first just the scenario involving IPD from all 10 trials; model (10) is equivalent to model (8) in this situation. The correlation between $\hat{\gamma}_W$ and $\hat{\gamma}_A$ was -0.0003 in the sex analysis and -0.00005 in the age analysis, highlighting the independence of γ_W and γ_A in models (8) and (10). For sex, $\hat{\gamma}_W = 0.89$ (SE 0.41) and $\hat{\gamma}_A = 15.10$ (SE 3.18), both indicating a worse treatment effect in males than females, but with $\hat{\gamma}_A$ suggesting a much larger sex effect than $\hat{\gamma}_W$. Figure 1 also shows this difference between $\hat{\gamma}_A$ and $\hat{\gamma}_W$; the within-trial relationships (dashed lines) have flat gradients, especially in the larger trials, but the across-trials relationship (solid line) has a steep gradient, which highlights the danger of using $\hat{\gamma}_A$ to make inferences about γ_W, with ecological bias and confounding a likely cause. $\hat{\gamma}_A$ also has a large standard error and low power, as the variability of sex within trials is large relative to the variability of the proportion male, m_i, across trials (Table I). This is confirmed by the $\widetilde{Q}_e$ statistic [11],

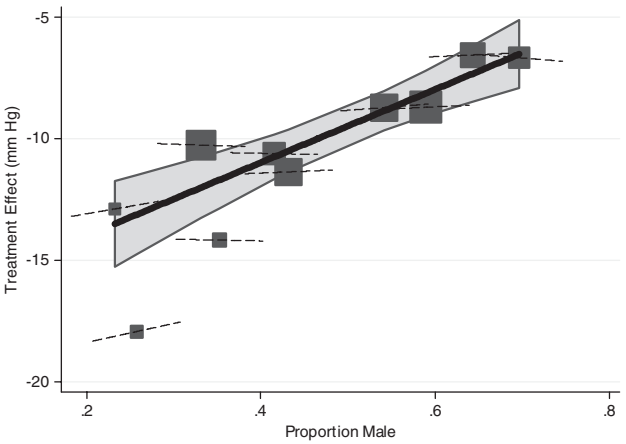

Figure 1. Plot showing: (i) the across-trials relationship ($\hat{\gamma}_A$, solid line) between proportion male (m_i) and treatment effect ($\hat{\theta}_i$) as estimated by meta-regression model (7), with a 95 per cent confidence interval around it; and (ii) the within-trial relationships ($\hat{\gamma}_i$, dashed lines) between sex and treatment effect as estimated separately within each trial using IPD and model (5). The gradient of each dashed line indicates the change in treatment effect from females to males within each trial ($\hat{\gamma}_i$); the length of the dashed lines is unimportant and is kept the same for each trial simply to aid clarity. Each block represents a trial and is centred at m_i in each trial; the block size is proportional to the inverse of the standard error of $\hat{\theta}_i$ in each trial.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Table III. Model (10) parameter estimates when analysing change in systolic blood pressure (before–after treatment) from the hypertension data in a range of different scenarios for each of sex and age.

	Per cent of trials providing IPD Trials used in the analysis											
	100 per cent	90 per cent		70 per cent		50 per cent		30 per cent		10 per cent		0 per cent
Parameter estimate	IPD trials only	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	IPD trials only	IPD & AD trials	AD trials only
Analysis of sex												
Meta-regression intercept, $\hat{\alpha}$	−17.04	−17.15	−17.02	−17.50	−16.99	−17.96	−16.99	−18.85	−17.01	—	−17.06	−17.06
(s.e)	(1.59)	(1.68)	(1.58)	(1.97)	(1.58)	(2.60)	(1.58)	(5.41)	(1.59)		(1.60)	(1.60)
Within-trial interaction: sex∗treatment, $\hat{\gamma}_W$	0.89	0.89	0.89	0.87	0.87	0.85	0.85	0.87	0.87	1.61	1.61	—
(s.e.)	(0.41)	(0.44)	(0.44)	(0.50)	(0.50)	(0.052)	(0.052)	(0.081)	(0.081)	(3.50)	(3.50)	
Across-trials interaction: sex∗treatment, $\hat{\gamma}_A$	15.10	15.28	15.07	15.89	15.02	16.70	15.03	18.43	15.05	—	15.16	15.16
(s.e.)	(3.18)	(3.26)	(3.07)	(3.93)	(3.06)	(5.55)	(3.06)	(14.04)	(3.08)		(3.10)	(3.11)
Between-study variance, $\hat{\tau}^2$	0.85	0.85	0.84	0.95	0.84	1.13	0.84	2.07	0.85	—	0.88	0.88
Treatment effect for a trial with proportion male m, $\hat{\alpha}+\hat{\gamma}_A m$	−9.52	−9.54	−9.52	−9.59	−9.51	−9.64	−9.51	−9.67	−9.51	—	−9.51	−9.51
(s.e.)	(0.40)	(0.42)	(0.40)	(0.52)	(0.40)	(0.78)	(0.40)	(2.37)	(0.40)		(0.40)	(0.40)
Analysis of age												
Meta-regression intercept, $\hat{\alpha}$	−1.84	−1.86	−1.82	−1.33	−1.99	3.34	−1.94	9.64	−1.93	—	−1.96	−1.97
(s.e)	(3.71)	(4.02)	(3.72)	(4.92)	(3.68)	(7.65)	(3.67)	(34.86)	(3.67)		(3.63)	(3.61)
Within-trial interaction: Age∗treatment, $\hat{\gamma}_W$	−0.050	−0.050	−0.050	−0.050	−0.050	−0.051	−0.051	−0.046	−0.046	0.068	0.068	—
(s.e.)	(0.034)	(0.036)	(0.036)	(0.041)	(0.041)	(0.052)	(0.052)	(0.082)	(0.082)	(0.29)	(0.29)	
Across-trials interaction: Age∗treatment, $\hat{\gamma}_A$	−0.14	−0.13	−0.14	−0.14	−0.13	−0.21	−0.13	−0.30	−0.13	—	−0.13	−0.13
(s.e.)	(0.060)	(0.064)	(0.060)	(0.077)	(0.059)	(0.12)	(0.059)	(0.50)	(0.059)		(0.058)	(0.058)
Between-study variance, $\hat{\tau}^2$	4.97	5.02	5.01	5.04	4.85	5.12	4.83	5.02	4.82	—	4.70	4.67
Treatment effect for a trial with mean age m, $\hat{\alpha}+\hat{\gamma}_A m$	−9.68	−9.66	−9.68	−9.50	−9.64	−8.68	−9.63	−7.70	−9.63	—	−9.63	−9.63
(s.e.)	(0.82)	(0.88)	(0.82)	(1.05)	(0.81)	(1.54)	(0.81)	(6.35)	(0.81)		(0.80)	(0.79)

Values shown are the average estimates across all possible combinations of IPD trials and AD trials within each scenario.

The average standard errors were calculated by taking the square root of the average variances.

IPD, individual patient data; AD, aggregate data; s.e., standard error; Sex = 0 for females and 1 for males; m, the mean covariate value across all trials; $m = 57.82$ for age, 0.50 for sex.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

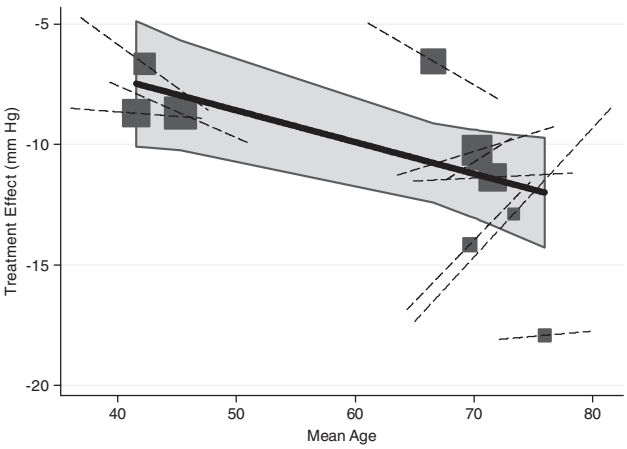

Figure 2. Plot showing: (i) the across-trials relationship ($\hat{\gamma}_A$, solid line) between mean age (m_i) and treatment effect ($\hat{\theta}_i$) as estimated by meta-regression model (7), with a 95 per cent confidence interval around it; and (ii) the within-trial relationships ($\hat{\gamma}_i$, dashed lines) between age and treatment effect as estimated separately within each trial using IPD and model (5). The gradient of each dashed line indicates the change in treatment effect for a one year increase in age within each trial ($\hat{\gamma}_i$). The width of the dashed line about the centre of each block is defined by 1 times the standard deviation of age in each trial. Each block represents a trial and is centred at m_i in each trial; the block size is proportional to the inverse of the standard error of $\hat{\theta}_i$ in each trial.

as $\widetilde{Q}_e/10=0.08$, far less than 1, which indicates that $\hat{\gamma}_A$ has low power. It is thus clear that separating within-trial and across-trials interactions is a sensible strategy for sex. Indeed the test against $H_0: \gamma_W = \gamma_A$ is highly significant ($p<0.0001$). Note, though, that modelling $\hat{\gamma}_A$ is itself important for sex, as it significantly reduces $\hat{\tau}^2$ from 7.13, when sex is excluded (Table II), to 0.85; this in turn reduces the standard error of the average treatment effect estimate across patients from 0.93, when sex is excluded (Table II), to 0.40, a reduction of 57 per cent. Of course, due to the potential for confounding, it may be that an unobserved confounding factor is causing the 'proportion male' covariate to explain the heterogeneity here.

For age, model (10) gives $\hat{\gamma}_W = -0.050$ (SE 0.034) and $\hat{\gamma}_A = -0.14$ (SE 0.060), both of which indicate that the treatment effect increases for a year increase in age. $\hat{\gamma}_W$ and $\hat{\gamma}_A$ are much closer than for sex, and $\widetilde{Q}_e/10$ equals 6.54 for age, indicating that $\hat{\gamma}_W$ may be powerful enough to detect the true treatment–covariate interaction within trials [11]. This is due to the variability of the m_i values across trials being relatively large for age (Table I). There was also no significant evidence against $H_0: \gamma_W = \gamma_A$ ($p=0.21$), and Figure 2 shows that the across-trials relationship is quite similar to the within-trial relationship in many of the larger trials. For these reasons, one might consider applying model (9) to combine $\hat{\gamma}_W$ and $\hat{\gamma}_A$ and thus allow the across-trials information help estimate the treatment–age interaction. On applying model (9), the treatment–covariate interaction, γ_{WA}, is estimated as -0.072, a weighted combination of $\hat{\gamma}_W = -0.050$ and $\hat{\gamma}_A = -0.14$ from model (8). Also the standard error of $\hat{\gamma}_{WA}$ is 0.029, smaller than the standard error of 0.034 for $\hat{\gamma}_W$ in model (8) due to the extra across-trials information being utilized. This gain in precision clearly arises from the simplifying assumption of $\gamma_W = \gamma_A$, which may not be true and is heavily dependent on the assumption of linearity both within trials and across trials (see Section 4.3.4). Further, $\hat{\gamma}_A$ may still

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

be subject to confounding, and hence $\hat{\gamma}_W$ and $\hat{\gamma}_A$ are perhaps still better modelled separately for age.

4.3.2. Results when combining IPD and AD. Consider now the scenarios involving a mixture of IPD and AD trials, where for each AD trial we assume that $\hat{\theta}_i$ (adjusted for baseline SBP), $V(\hat{\theta}_i)$, and m_i are available. Model (10) was fitted to each combination of IPD and AD trials in each scenario, for each of sex and age. We also fitted model (8) to just the IPD trials in each combination to assess the benefits of combining IPD and AD over just using IPD. The results, averaged across all possible combinations, are summarized in Table III and we are again interested in how estimates differ from those of the 100 per cent IPD analysis. Firstly, it is evident from Table III that model (10) correctly allows only the IPD trials to estimate the within-trial interaction, γ_W. For example, where five trials provide IPD and five only AD, $\hat{\gamma}_W$ is 0.85 (SE 0.052) for sex regardless of whether AD trials are included or not. This explains why the standard error of $\hat{\gamma}_W$ increases as the proportion of IPD trials decreases and emphasizes why it is better to obtain IPD from all trials. Yet, although its standard error is increased, $\hat{\gamma}_W$ itself remains close to the true answer (from the 100 per cent IPD analysis) even when just three trials provide IPD, in both age and sex analyses. The benefit of including AD trials is evident in the results for the across-trials parameters. For example, in the sex analysis, where three trials provide IPD and seven only AD, $\hat{\tau}^2$ is 2.07 in the IPD-only analysis and 0.85 in the combined analysis, the latter of which agrees with $\hat{\tau}^2$ from the 100 per cent IPD analysis. Similarly, the inclusion of AD trials produces estimates of the mean treatment effect when $m_i = 0\,(\hat{\alpha})$, the across-trials interaction ($\hat{\gamma}_A$), and the mean treatment effect in a trial with covariate mean $m\,(\hat{\alpha}+\hat{\gamma}_A m)$ that all closely agree with those from the 100 per cent IPD analysis. The standard error of these estimates is also much smaller in the analyses combining IPD and AD compared with the IPD-only analyses.

4.3.3. The impact of different adjustment factors in IPD and AD trials. For the age analysis, the benefits of combining IPD and AD over an IPD-only analysis are as discussed for the sex analysis, with the exception of $\hat{\tau}^2$. In some scenarios, such as where 30 per cent of trials provide IPD, the average value of $\hat{\tau}^2$ is actually closest to the true value in the IPD-only analysis. This is due to a subtle imbalance in age between treatment groups. In each AD trial the treatment effect estimate, $\hat{\theta}_i$, is adjusted for baseline SBP but not for age. However, on application of model (10) the IPD trials are adjusted for baseline SBP and also age. Given the treatment groups appeared balanced for age (Table I), this was not foreseen as a problem. Yet, including age as a confounding factor does subtly alter the treatment effect estimate in each trial, which leads to $\hat{\tau}^2$ being slightly different in the 100 per cent IPD analysis ($\hat{\tau}^2 = 4.97$) compared with the 100 per cent AD-only analysis ($\hat{\tau}^2 = 4.67$). Further differences may have arisen if the AD had not been adjusted for baseline SBP. In an ideal world, the $\hat{\theta}_i$ available from AD trials would be adjusted for all possible confounding factors; however, this is unlikely in practice. An advantage of IPD trials is that one can adjust for confounding using the IPD itself. For such reasons some may argue that including AD trials alongside IPD trials may lead to bias. However, such bias should only be small when the $\hat{\theta}_i$ are from RCTs, as any imbalance in treatment groups is itself usually only small, and importantly IPD-only meta-analyses may themselves introduce bias by ignoring AD trials. For example, in the scenario where 30 per cent of trials provide IPD, combining IPD and AD gives a bias of -0.15 in $\hat{\tau}^2$, slightly worse than the bias of 0.05 in the IPD-only analysis. Yet, the bias in $\hat{\alpha}$, $\hat{\gamma}_A$, and $\hat{\alpha}+\hat{\gamma}_A m$

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

is much worse in the IPD-only analyses, where it is 11.48, −0.16, and 1.98, respectively, compared with the analyses combining IPD and AD where it is −0.09, 0.01, and 0.05, respectively. We note, though, that unadjusted $\hat{\theta}_i$ from observational studies may be more biased than those from RCTs, as any imbalance between groups is likely to be more serious.

4.3.4. Non-linear interaction between treatment and age. Three trials have a mean age around 43, whereas the other seven trials have a mean age around 70 (Table I). Figure 2 suggests that the relationship between treatment effect and age may be non-linear, with a flatter within-trial interaction observed in the trials of older patients. We thus considered it a sensible, although admittedly data driven, exercise to extend the age analysis and fit model (10) allowing a separate within-trial treatment–age interaction for 'young' patients and 'old' patients. In each trial we defined patients aged 55 years or less to be 'young', as few patients in the trials with younger people were aged >55 and few patients in other trials were aged <55. In the 100 per cent IPD scenario, the extended model gives $\hat{\gamma}_W = -0.17$ (SE 0.056) in young patients and $\hat{\gamma}_W = 0.0030$ (SE 0.039) in old patients, a significant difference ($p = 0.007$) confirming the difference in the treatment–age interaction for young and old patients. Note that we did not fit a separate across-trials interaction for each subgroup of trials, as there are only three trials in the younger subgroup; thus the across-trial parameter results are as in Table III. Interestingly $\hat{\gamma}_A = -0.14$ is very similar to $\hat{\gamma}_W = -0.17$ for the younger patients; yet it is nonsensical to consider combining $\hat{\gamma}_A$ with $\hat{\gamma}_W$ here, as $\hat{\gamma}_A$ relates to a linear trend across all 10 trials, which contain both young and old patients, not just young patients.

5. EXTENSIONS TO THE MODELS

5.1. A general framework for the one-step approach

The one-step models in Section 4 can be generalized to allow multiple patient-level covariates. For example, model (10) can include each of $p = 1$ to P patient-level covariates, z_{pij}, which have means m_{pi} in each trial, as follows:

$$\begin{aligned} y^*_{ij} &= D_i\phi_i + \alpha_i x_{ij} + D_i \sum_{p=1}^{P} \mu_{pi} z^*_{pij} + D_i x_{ij} \sum_{p=1}^{P} \gamma_{Wp}(z^*_{pij} - m_{pi}) + x_{ij} \sum_{p=1}^{P} \gamma_{Ap} m_{pi} + \varepsilon^*_{ij} \\ \alpha_i &= \alpha + u_i \\ u_i &\sim \mathrm{N}(0, \tau^2) \\ \varepsilon^*_{ij} &\sim \mathrm{N}(0, V^*_i) \end{aligned} \tag{11}$$

When assessing multiple patient-level covariates, this one-step model is more convenient than a two-stage approach, due to the correlation between multiple interactions within trials. One-step models simultaneously estimate the within-trial interactions and thus inherently account for their correlation. A two-stage approach must more tediously obtain the correlation between interaction estimates from each trial separately and then, in the second stage, jointly synthesize the multiple interaction estimates in a multivariate model that accounts for this correlation [40]. Note that some

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

of the z_{pij} in model (11) could also be study-level covariates (e.g. trial location), but for these the associated $\gamma_{\mathrm{W}p}$ term would disappear as such z_{pij} are constant across patients within a trial. Model (11) could also include interaction terms between the multiple covariates if needed, both at the within-trial level and at the across-trials level. One has to be careful, though, when including multiple across-trials parameters as the power to assess them depends on the number of trials.

5.2. Multiple outcomes

Model (11) can be extended to assess multiple correlated continuous outcomes ($k=1$ to K). Consider two correlated outcomes, such as SBP and DBP. There is a correlation at the within-trial level, as responses for outcome 1 (y_{ij1}) are correlated with responses for outcome 2 (y_{ij2}), leading to a within-study correlation, $\hat{\rho}_{\mathrm{W}i}$, between $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$ in each trial. Further, there is a between-study correlation, ρ_{B}, as the underlying treatment effects, θ_{i1} and θ_{i2}, may be related across trials. For k outcomes, model (11) can be extended to

$$\begin{aligned}
y^*_{ijk} &= D_i\phi_{ik}+\alpha_{ik}x_{ij}+D_i\sum_{k=1}^{K}\sum_{p=1}^{P}\mu_{pik}z^*_{pij}\\
&\quad +D_i x_{ij}\sum_{k=1}^{K}\sum_{p=1}^{P}\gamma_{\mathrm{W}pk}(z^*_{pij}-m_{pi})+x_{ij}\sum_{k=1}^{K}\sum_{p=1}^{P}\gamma_{\mathrm{A}pk}m_{pi}+\varepsilon^*_{ijk} \qquad (12)\\
\alpha_{ik} &= \alpha_k+u_{ik}\\
u_{ik} &\sim \mathrm{N}(0,\tau_k^2)\\
\varepsilon^*_{ijk} &\sim \mathrm{N}(0,V^*_{ik})\\
\mathrm{cov}(u_{i1},u_{i2}) &= \rho_{\mathrm{B}}\tau_1\tau_2\\
\mathrm{cov}(\varepsilon^*_{ijk},\varepsilon^*_{ijk'}) &= \rho_{ikk'}\sqrt{V^*_{ik}V^*_{ik'}}
\end{aligned}$$

This model assumes a different patient-level correlation ($\rho_{ikk'}$) between each pair of different outcomes (k and k') in each study, but simpler correlation structures can be specified if desired. As an example of model (12), consider again just two outcomes ($k=1$ or 2). For each IPD trial, $D_i=1, y^*_{ijk}=y_{ijk}, V^*_{ik}=\sigma^2_{ik}$, and $\mathrm{corr}(\varepsilon^*_{ij1},\varepsilon^*_{ij2})=\rho_{i12}$. For AD trials, we assume that $\hat{\theta}_{i1}, \hat{\theta}_{i2}, V(\hat{\theta}_{i1}), V(\hat{\theta}_{i2})$, and $\hat{\rho}_{\mathrm{W}i}$ are available. We thus assume that there is one patient ($j=1$) with two responses and we set $D_i=0, x_{i1}=1, y^*_{i11}=\hat{\theta}_{i1}, y^*_{i12}=\hat{\theta}_{i2}, V^*_{i1}=V(\hat{\theta}_{i1}), V^*_{i2}=V(\hat{\theta}_{i2})$, and $\mathrm{corr}(\varepsilon^*_{ij1},\varepsilon^*_{ij2})=\hat{\rho}_{\mathrm{W}i}$ with each of these covariance terms assumed to be known. The model can accommodate trials providing only one outcome [41], under a missing at random assumption. The AD trials again help estimate only the across-trial parameters, whereas the IPD trials help estimate all parameters.

Model (12) was applied to the hypertension data in each of the scenarios generated in Section 2, with SBP and DBP the two outcomes of interest. For each AD trial, we assumed that the adjusted

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

$\hat{\theta}_{i1}, \hat{\theta}_{i2}, V(\hat{\theta}_{i1}), V(\hat{\theta}_{i2})$, and $\hat{\rho}_{Wi}$ were available (Table I). For brevity, we discuss here only the pooled SBP and DBP estimates. The results obtained were consistent regardless of the number of trials providing IPD. Across scenarios, the average pooled treatment effect estimate was -10.15 (SE 0.91) for SBP and -4.61 (SE 0.51) for DBP, indicating a beneficial treatment effect for both outcomes. The results are almost identical to those from analysing the outcomes independently (see Table II for SBP), suggesting that the correlation has little impact. This is generally true for complete data scenarios but the benefit of utilizing correlation becomes pronounced when there are missing outcomes [41]. One problem for model (12) in practice is that $\hat{\rho}_{Wi}$ may be unavailable for AD trials. Proposals exist to limit this (for example, see [42]), but here one could obtain the $\hat{\rho}_{Wi}$ in IPD trials and use them to inform the missing $\hat{\rho}_{Wi}$ in AD trials.

5.3. AD trials also provide treatment–covariate interactions

So far we have assumed that AD trials provide only $\hat{\theta}_i$, $V(\hat{\theta}_i)$, and m_i; however, occasionally they may also provide their within-trial treatment–covariate interaction, $\hat{\gamma}_i$, with $V(\hat{\gamma}_i)$. Given just a single treatment–covariate of interest, such AD trials could easily be utilized alongside the $\hat{\gamma}_i$ and $V(\hat{\gamma}_i)$ from IPD trials in model (6), the second part of the two-stage approach. However, for multiple treatment–covariate interactions it is more complex, as AD trials also need to provide the correlations between these interactions, as discussed in Section 5.1. Such correlations are unlikely to be available; hence, as for the multiple outcomes scenario, it may be necessary to assume that such correlations are similar to those observed in IPD trials. Further research of this issue is required, alongside if and how $\hat{\gamma}_i$ estimates from AD trials can be incorporated in the framework of one-step model (12).

6. DISCUSSION

In the context of continuous outcomes, this paper has developed and assessed a range of methods for combining IPD and AD. These methods were built on previous work [4, 5, 11, 16, 22, 38] and use either a two-step or a one-step approach. The two-step approach is more traditional, with IPD reduced to AD so that standard AD meta-analysis models can be used. The one-step approach is more complex but it provides a flexible framework for including both patient-level and trial-level parameters, with a dummy variable in dictating which terms the AD trials estimate. The one-step approach is also more convenient when non-linear effects or multiple covariates are assessed, as reasoned in Section 5.1.

The separation of within-trial and across-trials relationships is an important component of our work. Clinicians prefer treatment–covariate interaction estimates based on within-trial information, as these relate patients' clinical characteristics to treatment response [13]. Meta-regression results, however, assess across-trials relationships between group-level summaries (e.g. mean age) and treatment effect, which may not reflect the within-trial relationship for a variety of reasons and are difficult to interpret [14, 15]. This is shown in Figure 1 for sex, with the steep across-trials gradient dramatically different to the mostly flat gradients observed within trials. A simulation study [12] found that the across-trials interaction was an unbiased estimate of the true treatment–covariate interaction within trials; crucially, though, it did not generate data assuming any ecological bias or confounding across studies. Thus, generally assuming that the across-trials interaction is an unbiased estimate of the within-trials interaction, as in model (9), is not recommended. To ever

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

take this approach, one must assume that the within-trial and across-trial relationships are identical (e.g. both linear) and that no confounding or ecological bias affects the across-trials interaction. This situation is unlikely, and even if it did occur the across-trials interaction will usually have low power anyway [12]. For example, in Section 4.3.4 it was sensible only to fit linear effects across trials due to their small number; however, the IPD enabled us to fit different within-trial relationships for two patient subgroups. If some trials had included both young and old patients, we may also have fitted non-linear continuous functions within trials, such as polynomials and splines [43]. Note that in models (8) and (10)–(12) the across-trials interaction can be removed entirely if desired, leaving just the within-trial interaction and overall pooled effect [4]. This is at the expense of not explaining the between-study heterogeneity, which is not ideal [35]; yet it avoids patient-level responses being regressed against population-level factors (e.g. mean age in the trial), which is perhaps non-intuitive in clinical studies [4], unlike in other contexts where population factors truly affect individuals (e.g. mean ability in a school class [22]).

Thompson and Higgins [13] concur that 'the estimated relations between the extent of treatment benefit and patients' characteristics are derived only from within-trial information, so that confounding because of differences across trials is avoided'. We have shown how to do this in IPD trials, with the within-trials interaction separated from the across-trials interaction. For AD trials to contribute towards within-trial interactions, they need to provide their treatment–covariate interaction estimate and its variance (Section 5.3). Yet, usually AD trials provide only their treatment effect estimate and its variance, alongside group-level summaries. We focused on this situation in Sections 4 and 5 and showed (e.g. in model (10)) how to combine IPD and AD trials so that only IPD trials estimate the within-trial relationships, but both IPD and AD trials estimate the pooled treatment effect, the between-study variance, and the across-trials relationships. AD trials thus still contribute towards important meta-analysis results, but appropriately only IPD trials estimate the treatment–covariate interaction within trials. The argument to include AD trials becomes stronger as the amount of missing IPD only increases, as the IPD-only analyses generally move further from the truth with increasing uncertainty (Tables II and III). However, some may argue that AD is far less reliable than IPD and that including AD alongside IPD may lead to bias. For example, in Section 4.3.2 we highlighted that whilst IPD trials can adjust for confounding, AD trials may provide only unadjusted results. With this in mind, practitioners may wish to assess the sensitivity of meta-analysis results to the incorporation of AD and also explore any differences between IPD and AD trials. For example, IPD and AD hernia trials were found to be of comparable quality [44], justifying the inclusion of AD alongside IPD within meta-analyses of such trials [45].

The models in this paper can be extended to three or more treatment groups and adapted to binary or survival outcomes. For binary outcomes, one must specify two models linked by common parameters: (i) an IPD-trial model that assesses patients' binary responses in relation to within-trial and across-trial covariates and (ii) an AD-trial model that assesses the log-odds ratio from each trial in relation to trial-level covariates. Similarly, for survival outcomes patients' survival times can be modelled in the IPD-trial model and log-hazard ratios modelled in the AD-trial model. For further research, an investigation is needed regarding if and when random effects should be placed on the treatment–covariate interactions, $\hat{\gamma}_i$, and the underlying trial effects, ϕ_i [4]; in such situations one must also account for the correlation between the now multiple random effects. An investigation of how including AD trials affects publication bias assessments would also be interesting. Bayesian hierarchical-related regression [23] is an alternative, yet similar framework to our frequentist one-step models. This uses Markov chain Monte Carlo methods to simultaneously estimate IPD-trial and AD-trial models linked by common parameters. The approach has been

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

developed in the context of ecological studies, where individual-level data supplement the aggregate information across different groups [21], such as geographical areas. Sutton *et al.* [25] have also applied the method to clinical studies varying in design and whether they provided IPD or AD. Their model assumes that the across-trials interaction is an unbiased estimate of the within-trials treatment–covariate interaction; however, it could be extended to mirror model (10) if preferred. The Bayesian approach [4] also accounts for all parameter uncertainty (e.g. in $\hat{\sigma}_i^2$ and $\hat{\tau}^2$) and can incorporate pertinent prior information where available.

APPENDIX

SAS Proc Mixed code to fit model (10) when the first seven trials in Table I provide aggregate data and the last three trials provide IPD, with age as a covariate.

```
/* initiate the MIXED procedure, state that restricted maximum likelihood (REML) is to
be used to estimate the parameters, and specify the dataset containing AD for the first seven
trials and IPD for the remaining three trials */
proc mixed cl method=reml data=ad7ipd3;

/* denote the class variables, where 'idnr' is an identification number for each patient in
the dataset */
class idnr trial;

/* specify model (11) as in Section 4, with 'ipd' the dummy variable, 'agemean' denoting
the mean age in the trial, and 'agecent' denoting the patient age minus the mean age in
the trial /*
model diff=trial*ipd trial*age*ipd treat treat*agecent*ipd treat*agemean/noint s
cl covb;

/* specify that the treatment effect is randomly distributed across trials */
random treat/type=un subject=trial;

/* specify a residual variance for each trial */
repeated/type=un subject=trial(idnr)group=trial;

/* enter the (starting) values for the covariance parameters */
parms
/* firstly the between-study variance starting value*/
(1)
/* now the variance of the treatment effect in the 7 AD trials; these are fixed */
(0.72)
(4.73)
(10.31)
(0.30)
(0.14)
(0.58)
(0.30)
```

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

```
/* now the starting value for the residual variance in the 3 IPD trials */
(1)
(1)
(1)
/* specify that the variances given for the seven AD trials are to be kept fixed */
/eqcons=2 to 8;

/* test whether the within-trial interaction equals the across-trials interaction */
estimate 'diff' treat*agecent*ipd -1 treat*agemean 1 / cl;
run;
```

ACKNOWLEDGEMENTS

Richard Riley is funded as a Research Scientist in Evidence Synthesis by the Department of Health's National Coordinating Centre for Research Capacity Development. We would like to thank Samantha Rowland, from SAS technical Support, who helped us to generate the full set of trial combinations in the hypertension analyses. We also thank the Associate Editor and two reviewers whose comments have greatly improved the paper.

REFERENCES

1. Whitehead A. *Meta-analysis of Controlled Clinical Trials*. Wiley: West Sussex, 2002.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
3. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**:395–411.
4. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *Journal of Clinical Epidemiology* 2004; **57**: 683–697.
5. Simmonds MC. Statistical methodology of individual patient data. *Ph.D. Thesis*, University of Cambridge, 2005.
6. Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *The Lancet* 1993; **341**:418–422.
7. Tudur-Smith C WP, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Statistics in Medicine* 2005; **24**:1307–1319.
8. Trikalinos TA, Ioannidis JP. Predictive modeling and heterogeneity of baseline risk in meta-analysis of individual patient data. *Journal of Clinical Epidemiology* 2001; **54**:245–252.
9. Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* 2001; **20**:2219–2241.
10. Whitehead A, Omar RZ, Higgins JP, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* 2001; **20**:2243–2260.
11. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine* 2007; **26**:2982–2999.
12. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 2002; **55**:86–94.
13. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? *The Lancet* 2005; **365**:341–346.
14. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *The Lancet* 1998; **351**:123–127.
15. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine* 2002; **21**:371–387.
16. Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2005; **2**:209–217.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

17. 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.
18. 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.
19. Group MTC. Combination chemotherapy versus melphalan plus prednisone as treatment for multiple myeloma: an overview of 6,633 patients from 27 randomized trials. *Journal of Clinical Oncology* 1998; **16**:3832–3842.
20. Tudur C, Williamson PR, Kahan S, Best LY. The value of the aggregated data approach to meta-analysis with time-to-event outcomes. *Journal of the Royal Statistical Society*, *Series A* 2001; **164**:357–370.
21. Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data. *Statistics in Medicine* 2006; **25**:2136–2159.
22. Goldstein H, Yang M, Omar RZ, Turner RM, Thompson SG. Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society*, *Series C—Applied Statistics* 2000; **49**:399–412.
23. 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* 2007; DOI: 10.1111/j.1467-985X.2007.00500.x.
24. Zhu CR, Liu ZH, Sun KH, Ni ZZ. The multi-level meta analysis model combining individual level data with aggregative level data. *Sichuan Da Xue Xue Bao Yi Xue Ban* 2005; **36**:888–891.
25. Sutton AJ, Kendrick D, Coupland C. Meta-analysis of individual and aggregate level data. *Statistics in Medicine* 2007; DOI: 10.1002/sim.2916.
26. 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.
27. Gueyffier F, Boutitie F, Boissel JP, Coope J, Cutler J, Ekbom T, Fagard R, Friedman L, Perry HM, Pocock S *et al*. INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results. *Therapie* 1995; **50**:353–362.
28. 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.
29. 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.
30. 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.
31. Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. *British Medical Journal* 2001; **323**:1123–1124.
32. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*. SAS Institute Inc.: Cary, NC, 1996.
33. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* 1998; **54**:317–322.
34. Mathew T, Nordstrom K. On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics* 1999; **55**:1221–1223.
35. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal* 1994; **309**:1351–1355.
36. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; **54**:638–645.
37. Mancl LA, Leroux BG, DeRouen TA. Between-subject and within-subject statistical information in dental research. *Journal of Dental Research* 2000; **79**:1778–1781.
38. 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.
39. Dwyer T, Blizzard L. A discussion of some statistical methods for separating within-pair associations from associations among all twins in research on fetal origins of disease. *Paediatric and Perinatal Epidemiology* 2005; **19**:48–53.

Copyright © 2007 John Wiley & Sons, Ltd.

40. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624.
41. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 2007; **7**:3.
42. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2007; DOI: 10.1093/biostatistics/kxm023.
43. Lambert PC, Abrams KR, Jones DR, Halligan AW, Shennan A. Analysis of ambulatory blood pressure monitor data using a hierarchical model incorporating restricted cubic splines and heterogeneous within-subject variances. *Statistics in Medicine* 2001; **20**:3789–3805.
44. McCormack K, Scott N, Grant A. Are trials with individual patient data available different from trials without individual patient data available?. *Ninth Annual Cochrane Colloquium Abstracts*, Lyon, 2001.
45. Scott NW, McCormack K, Graham P, Go PM, Ross SJ, Grant AM. Open mesh versus non-mesh for repair of femoral and inguinal hernia. *Cochrane Database of Systematic Reviews* 2002; CD002197.

Copyright © 2007 John Wiley & Sons, Ltd.