Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Meta-analysis of continuous outcomes: using pseudo IPD created from aggregate data to adjust for baseline imbalance and assess treatment-by-baseline modification

Katerina Papadimitropoulou*[1,2] | Theo Stijnen[3] | Richard D. Riley[4] | Olaf M. Dekkers[1] | Saskia le Cessie[1,3]

[1]Clinical Epidemiology, Leiden University Medical Center, The Netherlands
[2]Data Science and Biometrics, Danone Nutricia Research, The Netherlands
[3]Biomedical Data Sciences, Leiden University Medical Center, The Netherlands
[4]Centre for Prognosis Research, Research Institute for Primary Care & Health Sciences, Keele University, UK

Correspondence
*Katerina Papadimitropoulou, Clinical Epidemiology, Leiden University Medical Center. Email: a.papadimitropoulou@lumc.nl

Summary

Meta-analysis of individual participant data (IPD) is considered the "gold-standard" for synthesizing clinical study evidence. However, gaining access to IPD can be a laborious task (if possible at all) and in practice only summary (aggregate) data are commonly available. In this work we focus on meta-analytic approaches of comparative studies where aggregate data are available for continuous outcomes measured at baseline (pre-treatment) and follow-up (post-treatment). We propose a method for constructing pseudo individual baselines and outcomes based on the aggregate data. These pseudo IPD can be subsequently analysed using standard analysis of covariance (ANCOVA) methods.

Pseudo IPD for continuous outcomes reported at two timepoints can be generated using the sufficient statistics of an ANCOVA model i.e., the mean and standard deviation at baseline and follow-up per group, together with the correlation of the baseline and follow-up measurements. Applying the ANCOVA approach, which crucially adjusts for baseline imbalances and accounts for the correlation between baseline and change scores, to the pseudo IPD results in identical estimates to the ones obtained by an ANCOVA on the true IPD. In addition, an interaction term between baseline and treatment effect can be added. There are several modelling options available under this approach, which makes it very flexible.

Methods are exemplified using reported data of a previously published IPD meta-analysis of 10 trials investigating the effect of antihypertensive treatments on systolic blood pressure, leading to identical results compared with the true IPD analysis and of a meta-analysis of fewer trials, where baseline imbalance occurred.

KEYWORDS:
meta-analysis, pseudo individual participant data, ANCOVA, sufficient statistics

1 | INTRODUCTION

Meta-analysis methods of individual participant data or individual patient data (IPD) are considered the "gold-standard" for clinical studies' evidence synthesis[1,2,3,4]. IPD meta-analysis has several advantages over the traditional aggregate data (AD) meta-analysis approach, which synthesizes summary statistics per study, often retrieved from published sources. For example when continuous outcomes are available at baseline and follow-up, IPD meta-analysis enables the meta-analyst to perform adjustments for baseline imbalances and detailed explorations of treatment-covariate interactions[5,6,7]. In addition, it comes with a large toolbox of methods and greater flexibility to analyse the data in an one-stage or two-stage approach[8,9,10,11].

There are, however, challenges as access to IPD can be problematic because of time and cost constraints and privacy issues, and often it is not feasible to retrieve the IPD of all studies to be synthesized. It is possible to generate/back-calculate IPD for different types of AD, such as for binary, ordinal and time to event outcomes[12,13,14,15]. For aggregate data of continuous outcomes reconstructing the original outcome values is not possible. However, we recently proposed an algorithm to construct pseudo IPD for an one-stage meta-analysis with one continuous outcome, using the sufficient statistics for linear mixed models i.e., group means, standard deviations and sample sizes[16]. In this way the analysis using the pseudo IPD yields exactly the same results as the analysis of the original IPD. The pseudo IPD approach allowed more flexible modeling, using standard linear mixed model software, for example enabling common or different residual variances for treatment and control groups in each study.

In this paper we extend the original method of creating pseudo IPD from reported AD to the situation where continuous outcomes are reported both at baseline and follow-up. We discuss how pseudo IPD can be derived, taking the correlation between baseline and follow-up/final measurements into account, using the summary observed group means, standard deviations at baseline and post-treatment, and the group correlation of the baseline and post-treatment values (or equivalently the standard deviations of the difference between baseline and post-baseline values in both groups). These summary measures are the *sufficient statistics* for an analysis of covariance (ANCOVA) approach under the linear mixed model (LMM) framework. The generated pseudo IPD can be analysed using standard software for linear mixed models, and a linear mixed model analysis of the pseudo IPD will yield identical results to the ones obtained when it is applied on the original IPD.

We describe the advantages of this approach, compared with the standard methods to synthesize aggregate baseline and follow-up data: using mean *follow-up* (post-treatment/final) scores, ignoring the baseline values and mean *change scores*, subtracting the follow-up value from the baseline[17,18].

It is possible to perform a meta-analysis in an one-stage or a two-stage approach using the pseudo IPD, using the toolbox of available IPD methods[8,9,10,11]. A plethora of modelling options is available and we discuss several options, assuming stratified and random study intercepts and random treatment effect models.

The flexibility of the linear mixed modelling framework makes it possible to correct for potential baseline imbalances. Although imbalance at baseline is not expected in a randomised trial, it can occur by chance, particularly in small trials[19] or due to flaws in the randomisation process[20].

Treatment effects may also differ between patients, depending on their baseline values. For example, in a trial for hypertension, patients with low systolic blood pressure at baseline are expected to experience less improvement after administration of treatment, compared with patients having high baseline pressure values. Similarly, severely depressed patients with high values on a depression score may profit more from treatment than patients with mild depression. When generating and analysing pseudo IPD using an ANCOVA approach we can cope with the correlation between the baseline value and the change score by introducing an interaction term between the baseline measurement and the treatment effect. In this way treatment heterogeneity depending on the baseline values can be further explored.

The paper is organized as follows. In Section 2 we introduce two illustrating meta-analysis datasets: one in hypertension where group-level AD of systolic blood pressure (SBP) at baseline and at follow-up for anti-hypertensive treatments versus placebo/no treatment are available from a previous IPD meta-analysis publication[21] and a second example where active versus sham treatments in obstructive sleep apnea are compared and baseline imbalance occured between the treatment groups[22]. In Sections 3 and 4, we describe some of the existing modelling options for one-stage and two-stage IPD meta-analyses, respectively, including models for treatment-by-baseline interaction. In Section 5, we explain how pseudo IPD baselines and outcomes can be generated from the aggregate continuous data in the case of correlated baseline and final measurements. In Section 6, we apply our proposed method to the hypertension dataset in/excluding an investigation of the interaction between

[0]**Abbreviations:** AD, aggregate data; CI: confidence interval, IPD, individual patient (participant) data; RCT, randomised controlled trial

baseline and treatment and compare the results with those obtained when using the original IPD as previously reported in the work of Riley *et al.*[21,23] and with standard two-stage methods on the AD. In addition, we apply the pseudo IPD approach on the sleep apnea dataset and compare the results of the pseudo IPD ANCOVA models, while varying group-correlations coefficients (as sensitivity analysis), with change scores AD meta-analysis. Brief final comments are provided in Section 7.

2 | ILLUSTRATING EXAMPLES

2.1 | Aggregate data from 10 trial in hypertension with baseline imbalance and artificial baseline imbalance

We use the reported aggregate data for studies originally contained in an IPD meta-analysis of Wang *et al.*[24], and subsequently analysed by Riley *et al.*[21] investigating the effect of hypertension treatments on systolic blood pressure (SBP). The authors included IPD of trials comparing antihypertensive treatments against placebo/no treatment[25,26,27,28]. A total of 28 851 patients from 10 trials were included. Each trial measured blood pressure at baseline and after treatment. The aggregate data for each trial, including the mean, standard deviation and correlation of the baseline and the final SBP values (in mmHg) are shown in Table 1. Riley *et al.*[21] compared several meta-analytic approaches to estimate the summary treatment effect of antihypertension treatments in reducing SBP using the original IPD and compared them to standard AD methods. In this article, we re-analyse these data using only the aggregate group means, standard deviations and correlations of the baseline and the final values and apply our algorithm to generate pseudo IPD. We also perform standard AD meta-analysis using change scores and provide a comparison of the different methods. Riley *et al.*[21] explored the effect of large baseline imbalance by modifying the original hypertension dataset. This was achieved by subtracting 5 mmHg from the baseline and final SBP values of patients in the treatment group of trials 1 and 2; 20 mmHg of patients in the treatment group of trials 4 and 5 and 10 mmHg of the baseline and final values of patients in the treatment group of trial 6 accordingly, such that five studies have lower baseline values in the treatment group compared with the control group. We also demonstrate our method on the aggregate version of this modified dataset.

TABLE 1 Aggregate data of the 10 hypertension trials included in the meta-analysis of Wang *et al.*[24] as reported by Riley *et al.*[21]

		Number of subjects		SBP baseline (mmHg)		SBP final (mmHg)		Correlation (SBP baseline, SPB final)	
				Treatment	Control	Treatment	Control		
ID	Trial name	Treatment	Control	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Treatment	Control
1	ATMH	780	750	152.28 (15.25)	153.05 (15.73)	132.85 (16.72)	139.75 (17.85)	0.265	0.284
2	HEP	150	199	189.94 (16.15)	191.55 (17.64)	165.06 (20.03)	179.89 (22.15)	0.335	0.331
3	EWPHE	90	82	177.33 (15.85)	178.23 (15.06)	156.88 (21.26)	170.45 (26.91)	0.462	0.534
4	HDFP	2427	2370	151.68 (19.83)	151.00 (19.53)	130.09 (19.25)	138.54 (21.26)	0.337	0.408
5	MRC-1	3546	3445	156.60 (16.09)	156.65 (15.96)	135.49 (16.32)	144.25 (17.58)	0.346	0.416
6	MRC-2	1314	1337	182.19 (12.63)	182.13 (12.73)	153.99 (20.13)	164.58 (19.71)	0.178	0.137
7	SHEP	2365	2371	170.49 (9.5)	170.12 (9.24)	145.10 (19.05)	156.24 (20.12)	0.315	0.253
8	STOP	137	131	194.68 (12.21)	194.15 (11.16)	171.46 (19.29)	189.11 (21.9)	0.177	0.414
9	Sy-Chi	1252	1139	170.73 (10.9)	170.25 (11.41)	150.2 (15.84)	156.55 (16.86)	0.199	0.347
10	Sy-Eur	2398	2297	173.75 (9.86)	173.94 (10.07)	154.87 (16.31)	165.24 (16.33)	0.319	0.431

ATMH: Australian Trial in Mild Hypertension, HDFP: Hypertension Detection and Follow-up Programme, EWPHE: European Working Party on High Blood Pressure in the Elderly, MRC: Medical Research Council, SBP: systolic blood pressure, SD: standard deviation, SHEP: Systolic Hypertension in the Elderly Programme, STOP: Swedish Trial in Old Patients with Hypertension, Sy-Chi: Systolic Hypertension in China, Sy-Eur: Systolic Hypertension in Europe

2.2 | Aggregate data from 8 trials in obstructive sleep apnea with baseline imbalance

Aggregate data from a review of treatments for obstructive sleep apnea in adults[22] were used. We focus on a meta-analysis summarising the treatment effect of an active continuous positive airway pressure (CPAP) device versus a sham CPAP. Eight studies, of in total 311 patients, recorded the apnea-hypopnea index (AHI), which is defined as the number of apnea and hypoapnea events divided by the total hours of sleep, at baseline and follow-up. The authors[22] estimated a statistically significant mean difference in change scores of AHI between active CPAP and sham, favoring CPAP (difference -46 events/hour 95% CI: [-57, -36]; blue/triangle, Figure 1). We re-analysed these data, taking into account the considerable baseline imbalance which occured between the treatment groups (difference of 5 events/hour, 95% CI [0, 11]– the subjects randomised in the active CPAP arm suffered more severely from sleep apnea; red/circle, Figure 1), and explored whether patients with higher AHI at baseline benefitted more from treatment. For comparison purposes, we have additionally included the summary estimates of the final values analysis, which is not preferred due to baseline imbalance (green/square, Figure 1).

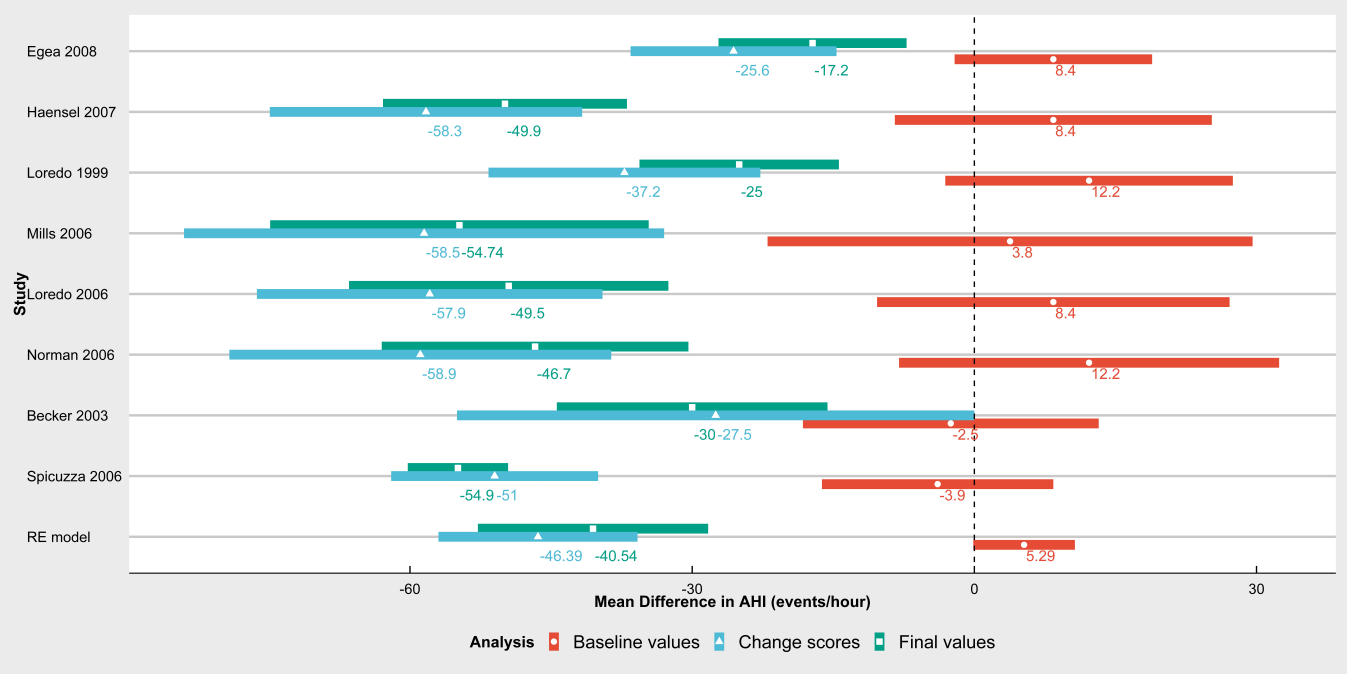

FIGURE 1 Obstructive sleep apnea meta-analysis example: forest plot of three different summary measures : a) difference in final values between mean AHI in the active CPAP group and mean AHI in the sham CPAP group (green/square); b) between groups difference in mean change from baseline (blue/triangle); c) between groups difference in mean AHI score at baseline (red/circle). The lowest line gives the results from a standard random effects meta-analysis.

3 | ONE-STAGE IPD META-ANALYSIS USING LINEAR MIXED MODELS (LMM)

In this section, we introduce notation and modelling options, for an one-stage meta-analysis of IPD of studies measuring continuous outcomes ast baseline and follow-up. The data we consider have the following format: let Y_{Bij} denote the *continuous outcome of interest* (i.e., SBP) at baseline/pre-treatment of patient j in study $i(1, ..., N)$ and Y_{Fij} the outcome, of each patient post-treatment (at follow-up). Also, let X_{ij} be a dummy variable to indicate the treatment group; X_{ij}=1 for patients in the treatment group and 0 for patients in the control group, respectively. There are many IPD meta-analysis ANCOVA type model options. A number of them are presented in this section; a similar description of the ANCOVA model can be found in Burke *et al.*[8]

3.1 | Analysis of covariance (ANCOVA)

3.1.1 | Stratified study model

An analysis of covariance (ANCOVA) model, with study-specific stratified intercepts and stratified adjustment terms for baseline measurements may be written as follows:

$$Y_{Fij} = \beta_{0i} + (\beta_1 + b_{1i})X_{ij} + \beta_{2i}(Y_{Bij} - \bar{Y}_{Bi}) + \epsilon_{ij}, \tag{1}$$

where β_{0i} is the mean outcome in the control group in study i for individuals with the mean baseline value, β_1 the summary (average) treatment effect and β_{2i} is the study-specific adjustment term for baseline values. A random effect b_{1i} is added to the overall treatment effect, which is assumed to be normally distributed with mean 0 and between-study variance equal to τ_1^2. Although a random treatment effect is preferred, one can assume a common (fixed) treatment effect by constraining τ_1^2=0. There are several modelling options for the variance of the within-study residuals, ϵ_{ij}, on which we elaborate later on.

3.1.2 | Random study model

An alternative approach to using stratified study intercepts and slopes is to assume a random intercept and a random baseline adjustment effect, resulting in the following ANCOVA model:

$$Y_{Fij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{ij} + (\beta_2 + b_{2i})(Y_{Bij} - \bar{Y}_{Bi}) + \epsilon_{ij}, \tag{2}$$
$$where \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} & \tau_{02} \\ \tau_{01} & \tau_1^2 & \tau_{12} \\ \tau_{02} & \tau_{12} & \tau_2^2 \end{bmatrix}\right)$$

Parameters are as in Equation (1), except for a random intercept and baseline adjustment coefficient; with τ_1^2 denoting the between-study variance of the treatment effect. In the literature, is it often assumed that the random effects are independent (i.e, $\tau_{ij} = 0$ for $i \neq j$), although under the LMM it is possible to estimate their covariances.

3.2 | Analysis of covariance (ANCOVA) including treatment-by-baseline interaction

To investigate potential treatment effect modification by the baseline value, the equations (1) and (2) can be extended by including the interaction term between baseline and treatment effect. The stratified study model (1) incorporating the "treatment-covariate interaction" is as follows:

$$Y_{Fij} = \beta_{0i} + (\beta_1 + b_{1i})X_{ij} + \beta_{2i}(Y_{Bij} - \bar{Y}_{Bi}) + (\beta_3 + b_{3i})[(Y_{Bij} - \bar{Y}_{Bi})X_{ij}] + \beta_{4i}(\bar{Y}_{Bi}X_{ij}) + \epsilon_{ij} \tag{3}$$

While the other parameters are as in Equation (1), β_3 denotes the mean increase in treatment effect for a one-unit increase in the baseline values and the random effect b_{3i} allows for between studies heterogeneity in the treatment-covariate interaction. This estimate reflects the within-trial interaction effect and β_{4i} estimates the increase in the treatment effect associated with a one-unit increase between the mean baseline of two studies, which reflects the across-trial interaction. Centering the baseline values and appropriately separating within- and across trial-associations avoids ecological bias, a phenomenon where the associations are erroneously equated[29]. Note that if the $\beta_{4i}(\bar{Y}_{Bi}X_{ij})$ is omitted from model (4), then the interaction term will reflect a weighed average of β_3 and the magnitude of the ecological bias[30].

Similarly, equation (2) can be extended yielding a random study ANCOVA model allowing for the interaction between baseline and treatment, which is formulated as follows:

$$Y_{Fij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{ij} + (\beta_2 + b_{2i})(Y_{Bij} - \bar{Y}_{Bi}) + (\beta_3 + b_{3i})[(Y_{Bij} - \bar{Y}_{Bi})X_{ij}] + \beta_{4i}(\bar{Y}_{Bi}X_{ij}) + \epsilon_{ij} \tag{4}$$

This model has four random effects ($b_{0i}, b_{1i}, b_{2i}, b_{3i}$), the covariance matrix of which may either be completely unspecified or may be modelled, for example by assuming independence of the different random effects.

Although, many other modelling specifications are possible, in this work we consider models (1) to (4).

3.3 | Within-study residual variances

The within-study residuals ϵ_{ij} are assumed to follow a normal distribution with mean 0. The within-study residual variance σ^2_{ik} may depend on the study i and group k. We explore *four structures* for modelling σ^2_{ik}: all variances assumed different (arm- and study-specific): $\epsilon_{ik} \sim N(0, \sigma^2_{ik})$, study-specific variances: $\sigma^2_{ik} = \sigma^2_{i.}$, one variance for control and one variance for treated group $\sigma^2_{ik} = \sigma^2_{.k}$, which are the same for all studies and one overall variance: $\sigma^2_{ik} = \sigma^2$.

4 | TWO-STAGE IPD META-ANALYSIS APPROACH

Instead of modelling all IPD in one model, in practice it may be more convenient to use a two-step approach. In the first stage, a separate ANCOVA is fitted in each of the studies $i = 1$ to N.

$$Y_{Fij} = \beta_{0i} + \beta_{1i} X_{ij} + \beta_{2i} Y_{Bij} + \epsilon_{ij} \tag{5}$$

This yields N treatment effects $\hat{\beta}_{1i}$ with standard errors se_i.
At the second stage a common (fixed)-effect or random-effects meta-analysis is run on the estimated study-specific β_{1is}.

In principle, the one-stage and two-stage approaches produce very similar results yet minor differences may arise as the former estimates the within-study residual variances simultaneously with β_{1i} and τ^2_1 while under the two-stage approach the within-study residual variances are estimated separately as seen in Eq (5) and independently of β_{1i} and τ^2_1 in the second stage. In particular, the stratified study one-stage model (1) and two-stage IPD meta-analysis approaches will yield very similar results, under the same underlying (modelling) assumptions, for example equal variance for treatment and control within studies[8]. For small sample sizes the results may deviate slightly. Equation (5) can also be extended to estimate the interaction between baseline values and treatement effect by introducing the interaction term similar to term β_3 from Eq (3).

5 | CONSTRUCTION OF PSEUDO IPD FROM AGGREGATE DATA

In our previous work we developed a method to generate pseudo IPD for a single continuous outcome per subject without baseline values[16]. The method generates data with the same observed means, standard deviations and sample sizes, the so-called pseudo IPD. Because the means and standard deviations are the sufficient statistics, the likelihood function for the IPD, using the linear mixed model is identical to the likelihood of the unknown true IPD. This means that analysing the pseudo IPD with LMM will yield identical results to the analysis of the true IPD.

In this article we extend our method to creating pseudo IPD from available aggregate data for a continuous outcome, reported at two timepoints, at baseline and follow-up. Appropriate sufficient statistics for an analysis of covariance (ANCOVA) approach are, for each study separately, the means and standard deviations of the continuous outcome at baseline and follow-up in each group, together with the group correlation of the baseline and follow-up values. Our premise is to create pseudo IPD that have exactly these sample means, standard deviations, and correlations, so that the subsequent pseudo IPD meta-analysis will produce the same results as if the original IPD were available.

The algorithm to construct for each of the studies and groups pseudo data with exactly the same mean, standard deviation and group correlation between baseline and follow-up measurement is as follows: let in a certain study arm, $\bar{Y}_B$, sd_B and $\bar{Y}_F$, sd_F be the observed means and SDs at baseline and follow-up, respectively and let r be the correlation between baseline and follow-up measurement, and let n be the sample size. Then for *each group in each study* separately, execute the following steps:

1. Simulate two samples $Y^*_{i1}(i = 1, ..., n)$ and $Y^*_{i2}(i = 1, ..., n)$, from a certain distribution, for example a standard normal distribution.
2. Standardise both samples to obtain $\bar{Y}^*_1 = 0$ and $\bar{Y}^*_2 = 0$, and $sd^*_1 = sd^*_2 = 1$ and calculate the correlation r^* between Y^*_{i1} and Y^*_{i2}.
3. Regress Y^*_{i2} on Y^*_{i1} and keep the regression coefficients $\hat{\beta}$ and the residuals $\hat{\epsilon}_i$. Note that since $sd^*_1 = sd^*_2 = 1$, it follows that $\hat{\beta}_i = r^*$ and $\hat{\epsilon}_i = Y^*_{i2} - r^* Y^*_{i1}$. Also note that the residuals are uncorrelated to Y^*_{i1} and have variance $1 - r^{*2}$.
4. Generate $Y^*_{i3} = Y^*_{i1} r + \hat{\epsilon}_i \sqrt{1 - r^2} [\sqrt{1 - r^{*2}}]^{-1}$. Note that $var(Y^*_{i3}) = 1$ and its correlation with Y^*_{i1} is r.

5. Generate the pseudo baseline as follows: $Y_{Bi} = Y^*_{i1} sd_B + \bar{Y}_B$.

   One can immediately verify that the pseudo baseline measurements have mean $\bar{Y}_B$ and standard deviation sd_B.

6. Generate the pseudo follow-up outcome as follows: $Y_{Fi} = Y^*_{i3} sd_F + \bar{Y}_F$.

   Similarly, the pseudo follow-up outcomes have mean $\bar{Y}_F$ and standard deviation sd_F and $cor(Y_{Bi}, Y_{Fi}) = r$.

This algorithm can be easily carried out in standard statistical software. In the Supplementary material we show how this algorithm can be carried out in R[31] and SAS[32]. The pseudo IPD can now be analysed using the LMM methods for IPD of Sections 3 and 4.

In practice, the group correlations are rarely reported. However, the mean change from baseline, with the standard deviation or standard error are more often provided. When the standard deviation at baseline, at follow-up and the change from baseline sd_{Change} are reported, the group correlation can be directly calculated as follows:

$$r = \frac{sd^2_B + sd^2_F - sd^2_{Changescores}}{2sd_B sd_F} \tag{6}$$

For more details see the Cochrane Handbook[33], Chapter 16. Alternatively, if the standard error of the difference between groups in mean change scores is provided and the pre/post correlations are assumed to be equal between the two groups; the correlation can be calculated as:

$$r = \frac{sd^2_{BT}/n_T + sd^2_{FT}/n_T + sd^2_{BC}/n_C + sd^2_{FC}/n_C - se^2_{difChangescores}}{2sd_{BT}sd_{FT}/n_T + 2sd_{BC}sd_{FC}/n_C} \tag{7}$$

where T and C are the indexes for treatment and control group, respectively[21]. When the group correlation cannot be derived from the available data, one could resort to imputation methods[34,35,36].

6 | APPLICATION OF THE METHODS TO THE DATA

We generated pseudo IPD baselines and outcomes for the aggregate hypertension data of Table 1, the aggregate hypertension dataset with artificial baseline imbalance and the AD of the obstructive sleep apnea example (given in the Supplementary material). Using these pseudo IPD we subsequently fitted the LMM models (1) to (4) discussed in Section 3; stratified study models and random study models, both with and without the interaction between treatment and baseline measurements. For the stratified models including the interaction term of baseline with the treatment effect, we assumed an unstructured variance-covariance matrix for the two random effects. For the random study models, we centered the groups when specifying the random effects, and assumed independent random effects due to memory issues. The parameters in the models were estimated using restricted maximum likelihood (REML[37]).

We fitted all models using the LMM program of SAS, PROC MIXED because SAS has explicit options for modelling the within-study residual variances and allows for additional flexibility using different methods to calculate the degrees of freedom and hence confidence intervals of the treatment effect. We used two different approaches, the default method where the degrees of freedom are calculated using the “between within" method in SAS, as it was the method also used in our previous work and also the Satterthwaite approximation method[38], following the recommendations of Legha *et al.*[11], who performed an extensive simulation study comparing the models in Section 3 under different CI derivations options.

In the Supplementary material we provide details on the SAS code and on how to fit the same models in R using nlme[39]. For comparison purposes with the results of Riley *et al.*[21], we only show the CIs derived using the between-within method.

6.1 | Results of the hypertension example with baseline balance

Results of the analyses using the pseudo IPD generated from the aggregate data on hypertension were compared with the two-stage IPD meta-analysis results of Riley *et al.*[21], who (unlike us) had access to the original IPD. As mentioned a two-stage IPD meta-analysis is very similar to the stratified study model of Equation (1) assuming equal residual variances between the treatment and the control group per study, i.e study-specific variances: $\sigma^2_{ik} = \sigma^2_{i.}$. We also performed a two-stage ANCOVA using the pseudo IPD. For completeness we also present the results of an AD meta-analysis using the change scores.

The results for the baseline balanced example are shown in the top two rows of Table 2. Across all competing models, the treatment effect estimates were negative indicating that the hypertension treatment reduced systolic blood pressure values.

The estimated treatment effect and corresponding standard error of the one-stage pseudo IPD ANCOVA analysis assuming study-specific residual variances, were identical to the results based on the analysis of the true IPD by Riley *et al.*[21]; -10.17 (SE=0.93) vs -10.17 (SE=0.93). There are slight differences in the 95% CIs as they were derived by different methods; under the Satterthwaite correction method were slightly wider. In addiiton, a two-stage analysis on the pseudo IPD assuming study-specific residual variances yielded identical results to model (1) and the analysis of the true IPD[21]: a summary treatment effect of -10.17, SE= 0.93.

We compared the AIC values[40] of different within-study residual variance structures for the stratified study models and for the random study models. In both model blocks the lowest value was found for the assuming all within-study residual variances to be free (arm-specific and study specific; 243387.2), although AIC values were found to be very similar across the different within-study variance options, suggesting that one could potentially adopt a simpler model when opting for a more parsimonious model. The study stratified model assuming within-study variances to be study-specific had the second lowest AIC value (243411.9) in that model block and was adopted as the final model. This model showed a summary treatment effect of -10.17 [95% CI: (-12.27, -8.06)], indicating that on average antihypertension treatments have a positive effect on SBP levels, reducing them by 10.17 mmHg more compared with control/no treatment.

The last column of Table 2 shows the results of the standard AD analysis following a change scores approach; a summary treatment effect -10.10 [95% CI: (-12.33, -7.87)], slightly lower than the ANCOVA estimate using one-stage or two-stage pseudo IPD.

6.2 | Results of the hypertension example with baseline imbalance

For the aggregate data with baseline imbalance, the effect of the active hypertension treatments compared with control is more pronounced (bottom rows of Table 2). We adopt the stratified study model as the final model which produces a summary treatment effect of -14.55 [95% CI: (-18.31, -10.80)], identical to the ANCOVA result of the true IPD presented in Riley *et al.*[21] Using a two-stage analysis of the pseudo IPD assuming study-specific residual variances resulted also in a summary treatment effect of -14.55 [95% CI: (-18.30, -10.80)].

The results of the pseudo IPD analysis were substantially different from the standard AD meta-analysis of change scores, because of the induced baseline imbalance.

TABLE 2 Meta-analysis results of *summary treatment effect* using the pseudo IPD approach compared with the true IPD and standard AD modelling approaches of Riley *et al.* [21]

			pseudo IPD meta-analysis				**true IPD meta-analysis**	**AD meta-analysis**
			one-stage ANCOVA				ANCOVA model results as described in Riley *et al.* [21]	Change scores
Dataset	Model	Results	σ^2_{ik}	$\sigma^2_{i.}$	$\sigma^2_{.k}$	σ^2	$\sigma^2_{i.}$	
Hypertension (balanced)	Stratified study (Eq. 1)	$\hat{\beta}_1$	-10.17	-10.17	-10.34	-10.34	-10.17	-10.10
		SE	0.93	0.93	0.98	0.98	0.93	0.99
		95% CI	(-11.99, -8.34)	(-12.27, -8.06)	(-12.26, -8.43)	(-12.26, -8.43)	(-12.28, -8.05)	(-12.33, -7.87)
		τ^2_1	7.12	7.11	8.17	8.17	7.15	6.56
	Random study (Eq. 2)	$\hat{\beta}_1$	-10.45	-10.46	10.56	-10.56		
		SE	0.99	0.99	1.01	1.01		
		95% CI	(-12.39, -8.52)	(-12.39, -8.53)	(-12.53, -8.58)	(-12.54, -8.59)		
		τ^2_1	8.61	8.62	9.13	9.16		
Hypertension (imbalanced)	Stratified study (Eq. 1)	$\hat{\beta}_1$	-14.57	-14.55	-14.58	-14.57	-14.55	-10.10
		SE	1.65	1.66	1.64	1.65	1.66	0.99
		95 % CI	(-17.81, -11.33)	(-18.31, -10.80)	(-17.80, -11.36)	(-17.78, -11.36)	(-18.30, -10.80)	(-12.33, -7.87)
		τ^2_1	25.28	25.47	25.30	25.19	25.43	6.56
	Random study (Eq. 2)	$\hat{\beta}_1$	-14.45	-14.46	-14.49	-14.48		
		SE	1.65	1.65	1.63	1.63		
		95% CI	(-17.69, -11.20)	(-17.67, -11.20)	(-17.69, -11.29)	(-17.68, -11.29)		
		τ^2_1	25.34	25.23	25.10	24.99		

CI: Confidence Interval, SE: standard error, σ^2_{ik}: study- and arm-specific variances, $\sigma^2_{i.}$: study-specific variances, $\sigma^2_{.k}$: two variance parameters; one for control and one for treatment, σ^2: one overall variance

6.3 | Including the interaction between baseline and treatment effect

To investigate potential treatment-by-baseline modification, we included the interaction term β_3 between baseline and treatment effect in the pseudo IPD LMM models. We compared the pseudo IPD models (3) and (4) with the two-stage IPD meta-analysis of Riley *et al.*[21] with interaction, and with a random-effects meta-regression of the final values on the mean baseline of the treatment group. The estimate obtained from the AD meta-regression is actually comparable to the β_4 term, which quantifies tha across-trial interaction. In the results we focus on the within-trial interaction estimate β_3 which reflects the treatment-by-baseline interaction.

In the balanced example case, the derived pseudo IPD ANCOVA interaction term under the stratified study model assuming all within-study residual variances to be free was equal to -0.09 [95% CI: (-0.17, -0.01)], providing some evidence that the treatment effect is slightly higher for the more severe hypertensive patients at baseline with higher SBP baseline values (top row of Table 3). In addition, the result from model (3) assuming study-specific residual variances was found to be identical to the two-stage model fitted in Riley *et al.*[21], -0.09 (SE:0.038). Using a two-stage analysis of the pseudo IPD assuming study-specific residual variances in SAS yielded a summary treatment-by-baseline interaction effect of -0.09 [95% CI: (-0.18, -0.00)]. We also replicated the two-stage analysis in STATA using the DerSimonian-Laird method[41] to combine the effects, where the results were found identical to the analysis in Riley *et al*[21].

The meta-regression results using the mean baseline value of the treatment group were higher compared with the pseudo IPD ANCOVA model (-0.16 vs -0.09).

The estimates of the interaction effect in the imbalanced baseline dataset using the pseudo IPD were found to be very similar to the ones in the balanced case. However, the meta-regression estimate was in the opposite direction of the effect compared with the ANCOVA pseudo IPD results. The across-trial interaction as estimated from a standard AD meta-analysis can differ from the within-trial interaction, i.e. the difference in treatment effect of two patients in the same study differing one unit at baseline, as estimated from a true IPD or pseudo IPD meta-analysis. The assumption that they are the same is often not plausible due to the fact that across-trial interaction can suffer from confounding[5]. This phenomenon is called ecological or aggregation bias. Therefore the across-trials interaction should be carefully interpreted. Also note that the statistical power for the estimation of the within-trial interaction is usually much larger than for the across-trials interaction, as reflected by the standard errors (Table 3).

TABLE 3 Meta-analysis results of *interaction* of baseline with treament using the pseudo IPD approach compared with the true IPD and standard AD modelling approaches of Riley *et al.* [21]

			pseudo IPD meta-analysis				**true IPD meta-analysis**	**AD meta-analysis**
			one-stage ANCOVA: including the interaction between baseline and treatment				ANCOVA model results as described in Riley *et al.* [21]	Meta-regression
Dataset	Model	Results	σ^2_{ik}	$\sigma^2_{i.}$	$\sigma^2_{.k}$	σ^2	$\sigma^2_{i.}$	Using $\bar{Y}_{BTi}$
Hypertension (balanced)	Stratified study (Eq. 3)	$\hat{\beta}_3$	-0.09	-0.09	-0.10	-0.10	-0.09	-0.16
		SE	0.04	0.04	0.04	0.04	0.03	0.05
		95% CI	(-0.17, -0.01)	(-0.17, -0.01)	(-0.18, -0.01)	(-0.18, -0.01)	(-0.16, -0.03)	(-0.28, -0.04)
		τ^2_3	0.01	0.01	0.01	0.01	0.01	3.16
	Random study (Eq. 4)	$\hat{\beta}_3$	-0.09	-0.09	-0.10	-0.10		
		SE	0.04	0.04	0.04	0.04		
		95% CI	(-0.17, -0.01)	(-0.17, -0.01)	(-0.18, -0.01)	(-0.17, -0.02)		
		τ^2_3	0.01	0.01	0.01	0.01		
Hypertension (imbalanced)	Stratified study (Eq. 3)	$\hat{\beta}_3$	-0.09	-0.09	-0.10	-0.10	-0.09	0.20
		SE	0.04	0.04	0.04	0.04	0.03	0.11
		95% CI	(-0.17, -0.01)	(-0.17, -0.01)	(-0.18, -0.01)	(-0.18, -0.01)	(-0.16, -0.03)	(-0.76, 0.50)
		τ^2_3	0.01	0.01	0.01	0.01	0.01	47.85
	Random study (Eq. 4)	$\hat{\beta}_3$	-0.09	-0.09	-0.10	-0.11		
		SE	0.04	0.04	0.04	0.0		
		95% CI	(-0.17, -0.01)	(-0.17, -0.01)	(-0.18, -0.01)	(-0.19, -0.02)		
		τ^2_3	0.010	0.011	0.012	0.012		

CI: Confidence Interval, SE: standard error, $\bar{Y}_{BTi}$: mean baseline SBP value of the treated group per trial, σ^2_{ik}: study- and arm-specific variances, $\sigma^2_{i.}$: study-specific variances, $\sigma^2_{.k}$: two variance parameters; one for control and one for treatment, σ^2: one overall variance

6.4 | Results of the obstructive sleep apnea example

In this second example, it was possible to calculate the group correlations (assumed to be equal between active and sham) using Equation (7); the derived correlations values varied slightly across studies [median: 0.498, IQR: 0.496-0.503]. We additionally performed sensitivity analyses by imputing three values of r (0.5, 0.6 and 0.7), to simulate cases where deriving the correlations from available data would not be possible. The R package ggplot2[42] was used to visualise the results of the competing models.

Figure 2 shows the results of the one-stage stratified study model assuming different options for the within-study residual variances. Results consistently showed that CPAP statistically significantly reduces AHI compared with the sham device (~ 41 events/hour). When r was calculated from the summary data (blue line/circle estimate), the point estimates across competing models vary slightly between 41 and 42 less events per hour in favor of active CPAP. The lowest AIC value was found for the most flexible model assuming arm and study residual variances to be free (AIC = 2273). Overall, AIC values did not differ greatly across the models hence simpler structures can also be adopted, e.g., study-specific within-study residual variances model.

The point estimates and 95% CIs were found to vary little across the imputed values of r, and the differences were not deemed to be clinically significant. The differences within the blocks of the more flexible modelling options (study- and arm- specific, and study-specific within-study residual variables) were more pronounced compared with the results of the more restricted models (group specific and one overall variance). Overall, the results based on the different imputed values within the same model block and across models did not seem to materially differ.

For this example, no direct comparison is feasible with the true IPD, thus we present the results of the one- and two-stage pseudo IPD analysis (using the calculated r value) and the original meta-analysis[22], and compare them with each other (Table 4). The one-stage stratified study model and the two-stage ANCOVA model, which form a natural comparison with one anoother, produced identical results when rounded in two decimal places (rows 3-4, Table 4). The point estimate of the standard AD change score analysis was larger compared with the ANCOVA results of the pseudo IPD, which may be explained by the negative correlation of the change scores with the baseline scores and the worse baseline of the subjects randomised in the active group. Generating the pseudo IPD enabled us to explore the interaction of baseline values with the treatment effect which in this example was found to be statistically significant (last two rows of Table 4), suggesting that the treatment effect is higher for the patients randomised in the active CPAP arm who were found to suffer more at baseline compared to the control patients.

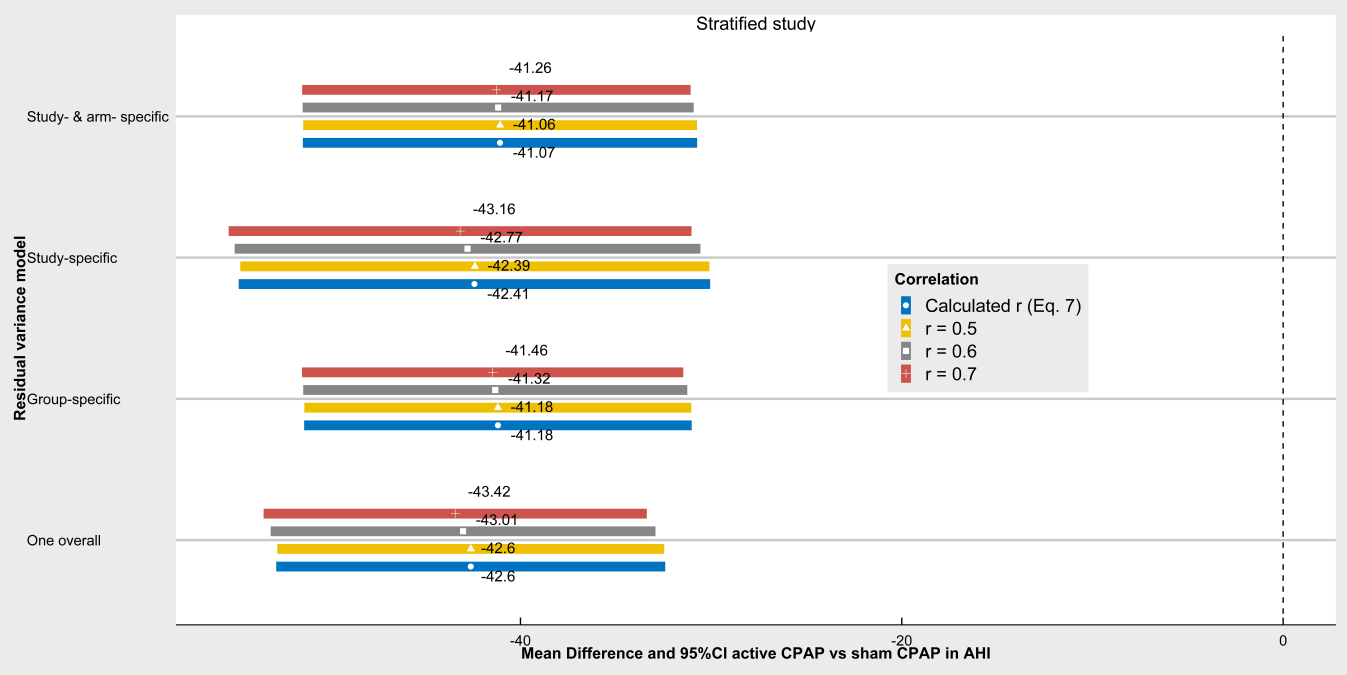

FIGURE 2 Obstructive sleep apnea meta-analysis results: estimates of overall mean difference of active CPAP vs sham and 95% CI in AHI across different residual variance models and varying group correlation coefficients between baseline and follow-up values.

TABLE 4 Meta-analysis results of *summary treatment effect and interaction effect* using the pseudo IPD approach compared with standard change score AD methods.

Approach	Method	Estimate	Standard Error	95% CI
standard AD	Difference in Change scores as in Balk *et al.* [22]	-46.39	5.39	(-56.97, -35.81)
pseudo IPD	One-stage ANCOVA*† Eq. (1)	-42.41	5.23	(-54.77, -30.05)
	Two-stage ANCOVA Eq. (5)	-42.41	5.23	(-54.77, -30.04)
	One-stage ANCOVA interaction effect Eq. (3)	-0.40	0.07	(-0.54, -0.25)
	Two-stage ANCOVA interaction effect	-0.40	0.07	(-0.54, -0.25)

CI: Confidence Interval, †Assumed $r_{CPAP} = r_{sham} = 0.5$, *A study-stratified model with study specific variances was used, $\bar{Y}_{BTi}$: mean baseline AHI values of the treated group per trial

7 | DISCUSSION

We have shown how aggregate data from comparative studies of continuous outcomes measured at baseline and follow-up can be analysed by generating pseudo IPD. These pseudo IPD enable us to use the complete palette of techniques available for IPD meta analyses. In particular, we are able to (1) perform an ANCOVA, where we can adjust for baseline imbalances between treatment and control groups and to (2) explore interactions between baseline values and treatment effects. Different modelling approaches of increasing complexity can be applied by using the linear mixed model (LMM) framework. Since the LMM analyses are likelihood-based, one-stage and two-stage results derived using the pseudo IPD baseline and follow-up outcomes are identical to the ones of the original IPD. The proposed methods can be applied in any standard statistical software therefore eliminating the need for training on a special purpose meta-analytic software.

In this article we have described modelling situations of comparing two treatment groups using the follow-up and baselines values. However, the LMM is a broad framework which offers rather staightfoward extensions of this work; the algorithm is directly generalisable to repeated measures meta-analysis and to multiple-treatments meta-analysis. Extension of the method for meta-analysis of cross-over trials is also applicable with some modifications albeit beyond the scope of this work. In addition, incorporation of non-linear covariates or non-linear interactions of treatment with continuous covariates could be a topic of future research as in this work we included the baseline (our covariate of interest) as a linear term in the ANCOVA model. Our algorithm could be extended to incorporate other covariates than only the baseline if the required summary statistics are available, in this case the variance-covariance matrix per group. These summaries are practically never reported however it is much easier to request them from the authors compared to the true IPD, as no privacy issues are involved. Bonofiglio and authors recently proposed a similar approach under distributed computing setting framework using only IPD summaries to recreate the marginal distributions of the original IPD considering eight baseline predictors in a multivariable logistic regression model[43].

The proposed approach successfully addresses the problem of IPD disclosure which is seldom possible due to various reasons with respect to data privacy and data security. In the case of continuous outcomes measured at baseline and follow-up often the sufficient aggregate data may be only partially available; for example often only means and standard deviations at baseline and mean change from baseline scores with the respective standard deviation or standard error are reported. Less frequently the mean and the standard deviation values at follow-up are provided. In that case, we could resort to algebraic calculations or imputation methods[36,34]. In principle, the minimally required set of aggregate data is the means and standard deviations at baseline and follow-up and also the standard deviation of the change from baseline. If these three standard deviations are provided, the correlation coefficient of baseline and follow-up can be calculated[33]. If one of these standard deviations are missing, they can potentially be algebraically extracted by other commonly reported summary statistics, e.g., confidence interval of mean difference, standard error of mean difference, paired t-test or a p-value from a paired t-test[44,45,46]. In cases where the post-baseline standard deviation is missing, it is common practice to assume it equal to the standard deviation at baseline and thus enable the calculation of the within-group correlation. Another commonly used approach is to impute the missing

SDs at post-baseline from other similar studies, with respect to study and patient characteristics, included in the meta-analysis. Recently, Weir and colleagues[36] proposed fifteen methods for addressing missing standard deviations (and by extension group correlations) in continuous data meta-analysis, building on the empirical review of Wiebe and colleagues in 2006[34]. Interested readers are referred to these reviews as a lengthy description of available methods for calculating or imputing the missing summary data is beyond the scope of this work. We also encourage contacting the authors of the original studies to provide the aggregate data also at follow-up, when confidentiality issues prohibit the direct provision of IPD.

We compared our pseudo IPD approach to standard meta-analytic approaches for aggregate data: random effects meta-analysis using change scores and meta-regression of the final scores on the baseline values of the treatment group to compare their performance with the pseudo IPD models. In case of imbalanced baseline values, the AD methods based on change scores tend to provide biased treatment and interaction effect estimates compared with the pseudo IPD ANCOVA methods.

Another advantage of the pseudo IPD approach is that it allows us to make more realistic and flexible assumptions regarding the within-study residual variances. In the absence of computational or estimations issues, we propose to use a realistic structure of the within-study residual variance. This flexibility is not possible in the standard AD analysis. Moreover, the standard AD assumes the standard errors of the treatment effects to be fixed and known, while using pseudo IPD ANCOVA methods may account for the fact that these are estimated.

When the appropriate AD are available (i.e., two means, standard deviations and correlation per group), we strongly recommend our proposed methodology to construct the pseudo IPD and perform an ANCOVA, if needed including the treatment-by-baseline interaction term. The advantage of our method is highlighted particularly in the case of baseline imbalance and in the case of treatment-baseline interaction, as the standard AD methods for interaction are known to suffer from low power and the potential of ecological-bias.

HIGHLIGHTS

What is already known?

The meta-analysis of IPD has been advocated as the "gold-standard" of evidence synthesis for many years. The generally preferred method to analyse IPD with continuous measurements at baseline and follow-up is linear mixed effects ANCOVA model. However access to IPD is often impossible. Researchers thus resort in an AD meta-analysis where in case of baseline imbalances, the treatment effects, derived by other methods than ANCOVA, may be biased.

What is new?

We provide an algorithm which makes use of summary reported AD of continuous measurements at baseline and follow up for to construct pseudo IPD. These pseudo IPD can be analysed in the same way as the original IPD using ANCOVA, producing identical results. Therefore we can adjust for baseline imbalances between treatment and control groups and explore interactions between baseline values and treatment effects. In the example dataset where the true IPD have been synthesized, the results of our analysis were identical to the true original IPD results.

What is the potential impact for RSM readers outside the author's field?

To enable reproducibility and dissemination of the method, we have provided implementation code of the algorithm both in R and SAS. Meta-analysis is a statistical technique undertaken by researchers from various fields and thus being able to use the provided code in easily accessible free and commercial software can only improve the quality of their work.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

References

1. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the health professions* 2002; 25(1): 76–97.

2. Stewart L, Tierney J, Burdett S. Do systematic reviews based on individual patient data offer a means of circumventing biases associated with trial publications?. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* 2005: 261–286.

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

4. Riley RD. Commentary: like it and lump it? Meta-analysis using individual participant data. *International journal of epidemiology* 2010; 39(5): 1359–1361.

5. 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(3): 371–387.

6. 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(1): 86–94.

7. Riley R, Debray T, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: statistical recommendations for conduct and planning. *Statistics in Medicine* 2020. doi: 10.1002/sim.8516

8. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in Medicine* 2017; 36(5): 855–875.

9. Simmonds MC, Higginsa JP, Stewartb LA, Tierneyb 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(3): 209–217.

10. Debray T, Moons KGM, Valkenhoef vG, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Research Synthesis Methods* 2015; 6(August): 293–309. doi: 10.1002/jrsm.1160

11. Legha A, Riley RD, Ensor J, Snell KI, Morris TP, Burke DL. Individual participant data meta-analysis of continuous outcomes: A comparison of approaches for specifying and estimating one-stage models. *Statistics in Medicine* 2018; 37(4): 4404–4420.

12. Turner R, Rumana R, Yang M, Goldstein H, Thompson S. Multilevel models for meta analysis of clinical trials with binary outcomes. *Statistics in Medicine* 1999.

13. Whitehead A, Omar R, Higgins J, Savaluny E, Turner R, Thompson S. Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* 2001; 20(15): 2243–2260.

14. Guyot P, Ades A, Ouwens M, Welton N. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC medical research methodology* 2012; 12(1): 9.

15. Stijnen, T., Hamza, T H., and Ozdemir, P . Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine* 2010; 29: 3046–3067.

16. Papadimitropoulou K, Stijnen T, Dekkers OM, Cessie lS. One-stage random effects meta-analysis using linear mixed models for aggregate continuous outcome data. *Research Synthesis Methods* 2019; 10(3): 360–375.

17. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *BMJ: British Medical Journal* 2001; 323(7321): 1123–1124.

18. Deeks J, Higgins J, Altman D. on behalf of the Cochrane Statistical Methods Group. Chapter 9.4. 5.2. Meta-analysis of change scores. *Cochrane Handbook for Systematic Reviews of Interventions Version*; 5(1.0).

19. Rosenberger WF, Lachin JM. *Randomization in clinical trials: theory and practice*. John Wiley & Sons . 2015.

20. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *The Lancet* 2002; 359(9306): 614–618.

21. Riley RD, Kauser I, Bland M, et al. Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. *Statistics in Medicine* 2013; 32(16): 2747–2766.

22. Balk EM, Moorthy D, Obadan NO, et al. *Diagnosis and treatment of obstructive sleep apnea in adults*. Comparative Effectiveness Review No. 32. (Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-1). AHRQ Publication No. 11-EHC052-EF. Rockville, MD: Agency for Healthcare Research and Quality. July 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

23. Riley R, Price M, Jackson D, et al. Multivariate meta-analysis using individual participant data. *Research Synthesis Methods* 2015; 6(2): 157–174.

24. Wang JG, Staessen JA, Franklin SS, Fagard R, Gueyffier F. Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome. *Hypertension* 2005; 45(5): 907–913.

25. Gueyffier F, Boutitie F, Boissel J, et al. INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results.. *Therapie* 1995; 50(4): 353–362.

26. Liu L, Wang JG, Gong L, Liu G, Staessen JA, others . Comparison of active treatment and placebo in older Chinese patients with isolated systolic hypertension. *Journal of hypertension* 1998; 16(12): 1823–1829.

27. Staessen JA, Fagard R, Thijs L, et al. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. *The Lancet* 1997; 350(9080): 757–764.

28. Amery A, Brixko P, Clement D, et al. Mortality and morbidity results from the European Working Party on High Blood Pressure in the Elderly trial. *The Lancet* 1985; 325(8442): 1349–1354.

29. Hua H, Burke DL, Crowther MJ, Ensor J, Tudur Smith C, Riley RD. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. *Statistics in Medicine* 2017; 36(5): 772–789.

30. Riley RD, Lambert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in medicine* 2008; 27(11): 1870–1893.

31. R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2013.

32. Base 9.3 Procedures Guide . SAS Institute Inc. 2011.

33. Higgins J, Deeks J, Altman D. on behalf of the Cochrane Statistical Methods Group, editors. Chapter 16.1. 3.2. Imputing standard deviations for changes from baseline. *Cochrane handbook for systematic reviews of interventions. Version*; 5(0).

34. Wiebe N, Vandermeer B, Platt RW, Klassen TP, Moher D, Barrowman NJ. A systematic review identifies a lack of standardization in methods for handling missing variance data. *Journal of clinical epidemiology* 2006; 59(4): 342–353.

35. Balk EM, Earley A, Patel K, Trikalinos TA, Dahabreh IJ. *Empirical assessment of within-arm correlation imputation in trials of continuous outcomes*. Methods Research Report. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 12(13)-EHC141-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

36. Weir CJ, Butcher I, Assi V, et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. *BMC medical research methodology* 2018; 18(1): 25.

37. Paccagnella O. Sample size and accuracy of estimates in multilevel models. *Methodology* 2011.

38. Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics bulletin* 1946; 2(6): 110–114.

39. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team . *nlme: Linear and Nonlinear Mixed Effects Models*. 2020. R package version 3.1-145.

40. Akaike H. Information theory and an extension of the maximum likelihood principle, in Proceeding of 2nd International Symposium on Information Theory. 1973.

41. DerSimonian R, Laird N. Meta-Analysis in Clinical Trials. *Statistics in medicine* 1986; 188: 177–188. doi: 10.1016/0197-2456(86)90046-2

42. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York . 2016.

43. Bonofiglio F, Schumacher M, Binder H. Recovery of original individual person data (IPD) inferences from empirical IPD summaries only: Applications to distributed computing under disclosure constraints. *Statistics in Medicine* 2020; 39(8): 1183–1198.

44. Sutton AJ, Abrams KR, Jones DR, Jones DR, Sheldon TA, Song F. *Missing Data. in: Methods for meta-analysis in medical research*. 348. Wiley Chichester . 2000.

45. Whitehead A. Dealing with non-standard data sets. *Meta-analysis of controlled clinical trials* 2002: 215–240.

46. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*. 4. John Wiley & Sons . 2011.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

TABLE S1 Aggregate data of the 8 trials included in the meta-analysis of Balk *et al.* [22]

		Number of subjects		**AHI index at baseline**		**AHI index at follow-up**		**Reported correlation**	**Calculated correlation using Eq. (7)**
				Treatment	Control	Treatment	Control		Equal between treatment and control groups
ID	Trial name	Treatment	Control	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)		
1	Egea 2008	27	29	43.7 (22.9)	35.3 (16.7)	10.8 (11.4)	28.0 (24.8)	-	0.4979
2	Haensel 2007	25	25	65.9 (28.6)	57.5 (32.1)	3.5 (3.4)	53.4 (32.9)	-	0.4981
3	Loredo 1999	23	18	56.4 (24.1)	44.2 (25.3)	3.3 (3.8)	28.3 (22.7)	-	0.4442
4	Mills 2006	17	16	65.0 (34.0)	61.2 (41.0)	2.6 (2.4)	57.3 (41.0)	-	0.4969
5	Loredo 2006	22	19	65.9 (28.6)	57.5 (32.1)	3.0 (4.7)	52.5 (37.5)	-	0.5704
6	Norman 2006	18	15	66.1 (29.1)	53.9 (29.8)	3.4 (3.0)	50.1 (32.1)	-	0.4967
7	Becker 2003	16	16	62.5 (17.8)	65.0 (26.7)	3.4 (3.1)	33.4 (29.2)	-	0.5025
8	Spicuzza 2006	15	10	55.3 (11.9)	59.2 (17.3)	2.1 (0.3)	57.0 (8.6)	-	0.5052

AHI: Apnea-hypopnea index, SD: standard deviation

SUPPORTING INFORMATION

The following supporting information is available as part of the online article:

dataWang. Data of meta-analysis of Wang *et al.*[24] used in the application.

dataSleepApnea. Data of review of Balk *et al.*[22] used in the application.

pseudo IPD SAScode. Implementation of the algorithm and model fitting in SAS using proc mixed.

pseudo IPD Rcode. Implementation of the algorithm and model fitting in R using nlme.

How to cite this article: Papadimitropoulou K., T. Stijnen, R.D Riley, O.M. Dekkers, and S. le Cessie (2020), Meta-analysis of continuous outcomes: using pseudo IPD created from aggregate data to adjust for baseline imbalance and assess treatment-by-baseline modification, *Research Synthesis Methods*.