



# Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning

Richard D. Riley<sup>1</sup> | Thomas P.A. Debray<sup>2</sup> | David Fisher<sup>3</sup> | Miriam Hattle<sup>1</sup> | Nadine Marlin<sup>4</sup> | Jeroen Hoogland<sup>2</sup> | Francois Gueyffier<sup>5</sup> | Jan A. Staessen<sup>6</sup> | Jiguang Wang<sup>7</sup> | Karel G.M. Moons<sup>2</sup> | Johannes B. Reitsma<sup>2</sup> | Joie Ensor<sup>1</sup>

<sup>1</sup>Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK

<sup>2</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>3</sup>MRC Clinical Trials Unit, Institute of Clinical Trials & Methodology, Faculty of Population Health Sciences, University College London, London, UK

<sup>4</sup>Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

<sup>5</sup>Inserm, Lyon, France

<sup>6</sup>Department of Cardiovascular Sciences, Research Unit Hypertension and Cardiovascular Epidemiology, Studies Coordinating Centre, KU Leuven, Leuven, Belgium

<sup>7</sup>Centre for Epidemiological Studies and Clinical Trials, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

## Correspondence

Richard D. Riley, Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire ST5 5BG, UK.  
Email: r.riley@keele.ac.uk

## Funding information

NIHR Health Technology Assessment Programme, Grant/Award Number:

Precision medicine research often searches for treatment-covariate interactions, which refers to when a treatment effect (eg, measured as a mean difference, odds ratio, hazard ratio) changes across values of a participant-level covariate (eg, age, gender, biomarker). Single trials do not usually have sufficient power to detect genuine treatment-covariate interactions, which motivate the sharing of individual participant data (IPD) from multiple trials for meta-analysis. Here, we provide statistical recommendations for conducting and planning an IPD meta-analysis of randomized trials to examine treatment-covariate interactions. For conduct, two-stage and one-stage statistical models are described, and we recommend: (i) interactions should be estimated directly, and not by calculating differences in meta-analysis results for subgroups; (ii) interaction estimates should be based solely on within-study information; (iii) continuous covariates and outcomes should be analyzed on their continuous scale; (iv) nonlinear relationships should be examined for continuous covariates, using a multivariate meta-analysis of the trend (eg, using restricted cubic spline functions); and (v) translation of interactions into clinical practice is nontrivial, requiring individualized treatment effect prediction. For planning, we describe first why the decision to initiate an IPD meta-analysis project should not be based on between-study heterogeneity in the overall treatment effect; and second, how to calculate the power of a potential IPD meta-analysis project in advance of IPD collection, conditional on characteristics (eg, number of participants, standard deviation of covariates) of the trials (potentially) promising their IPD. Real IPD meta-analysis projects are used for illustration throughout.

## KEYWORDS

effect modifier, individual participant data (IPD), meta-analysis, subgroup effect, treatment-covariate interaction

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Statistics in Medicine* published by John Wiley & Sons, Ltd.

programme 12/01; National Institute for Health Research (NIHR) Clinical Trials Unit Support Funding; TOP grant of the Netherlands Organisation for Health Research and Development (ZonMw), Grant/Award Number: 91215058; NIHR Doctoral Fellowship, Grant/Award Number: DRF-2018-11-ST2-077; Netherlands Organisation for Health Research and Development; Keele University

## 1 | INTRODUCTION

Precision medicine and stratified healthcare tailors treatment decisions to individuals based on their particular characteristics.<sup>1,2</sup> The goal is to optimize treatment decisions and reduce unnecessary costs for each individual, by selecting treatments most likely to benefit (or least likely to harm) them based on their participant-level covariate values. For example, trastuzumab is only given to the subgroup of breast cancer patients who are human epidermal growth factor receptor 2 (HER-2) positive, as it is known to lock on to the HER-2 protein, block the receptor, and stop the cells from dividing and growing.<sup>3</sup> It is therefore considered unnecessary for those who are HER-2 negative. In this situation, there is a so-called treatment-covariate interaction, such that the treatment effect (measured on a scale such as a mean difference, risk ratio, odds ratio, or hazard ratio) changes according to an individual's HER-2 status. Participant-level covariates that interact with treatment effect are also known as effect modifiers, moderators, predictors of treatment effect and, mainly in the cancer literature, predictive markers.

Though some treatment-covariate interactions, such as HER-2, are suspected in advance due to strong biological rationale, others are only identified following secondary investigations of existing data. Single randomized trials are typically powered on the overall treatment effect (ie, the treatment effect averaged across all individuals), and so do not usually have sufficient power to detect differences in treatment effect between individuals. Powering a single trial to detect a genuine treatment-covariate interaction will typically require at least four times the sample size needed to test the overall treatment effect;<sup>4</sup> thus the funding of such trials is expensive and often infeasible. When individual participant data (IPD) from multiple randomized trials are available, meta-analysis provides the opportunity to increase power to detect true treatment-covariate interactions.<sup>5</sup> For this reason, many IPD meta-analyses of randomized trials are initiated specifically to examine one or more potential treatment-covariate interactions, often with the aim to reduce or explain observed between-study heterogeneity in a previous meta-analysis of (published) aggregate data. For pharmaceutical companies, a driver may be to rescue a treatment that previously showed no overall benefit on all patients combined, but which may still benefit a select group of individuals.

Over the last decade we have been involved in a number of IPD meta-analysis projects aiming to examine treatment-covariate interactions at the participant-level, and learnt important lessons and pitfalls from a statistical point of view. In this tutorial article, we share our experience to help other researchers in this situation, building on previous work in both IPD meta-analysis and single study settings.<sup>4,6-16</sup> Our main goal is to provide statistical recommendations for the meta-analysis part of the project; in particular, how to estimate treatment-covariate interactions using either a one-stage or two-stage IPD meta-analysis framework. A further aim is to highlight two statistical topics that are relevant to consider before embarking on an IPD meta-analysis project to examine treatment-covariate interactions. The outline of the article is as follows. Section 2 describes two motivating examples used for illustration throughout the whole article. Section 3 describes a framework for two-stage and one-stage IPD meta-analysis models for estimating treatment-covariate interactions, and Section 4 highlights five recommendations in this context: (i) calculating differences in meta-analysis results for subgroups is misleading; (ii) interaction estimates should be based solely on within-study information; (iii) continuous covariates (and outcomes) should be analyzed on their continuous scale; (iv) nonlinear relationships should be examined for continuous covariates, using a multivariate meta-analysis of the trend (eg, using restricted cubic spline functions); and (v) translation of interactions into clinical practice requires individualized treatment effect prediction. Section 5 then describes two statistical issues in guiding decisions to initiate an IPD meta-analysis project to examine interactions. First, how the decision to initiate such IPD meta-analysis projects should not be based on the amount of between-study heterogeneity in the overall treatment effect. Second, how to calculate, in advance of IPD collection, the

power of a potential IPD meta-analysis conditional on the characteristics of those trials promising their IPD. Section 6 concludes with discussion.

## 2 | MOTIVATING EXAMPLES

We now introduce two examples, which will be used to illustrate some of the key issues in later sections.

### 2.1 | An IPD meta-analysis examining the effect of antihypertensive treatment

Wang et al<sup>17</sup> performed an IPD meta-analysis of trials in participants with hypertension to investigate to what extent lowering of systolic blood pressure (SBP) contributed to the prevention of cardiovascular events. They selected randomized trials that tested active antihypertensive drugs against control. IPD was sought from trials in the Individual Data ANalysis of Antihypertensive intervention trials (INDANA) dataset or at the Studies Coordinating Centre in Leuven (Belgium). Ten trials (with a parallel group design) were ultimately included, and these provided IPD for a total of 28 581 patients. Key outcomes of interest by end of follow-up were SBP, cardiovascular disease (CVD), and all-cause mortality. An important focus is on whether the effect of antihypertensive treatment on these outcomes is modified by (ie, interacts with) participant-level covariates such as gender and baseline blood pressure.

### 2.2 | An IPD meta-analysis examining the effect of interventions to reduce gestational weight gain in pregnancy

The International Weight Management in Pregnancy (i-WIP) Collaborative Group obtained IPD from 36 trials (12 447 women), to investigate whether diet and lifestyle interventions improve outcomes during pregnancy. A primary outcome was whether interventions reduced gestational weight gain, for which 33 studies and a total of 9320 women were available. Weight was recorded as a continuous outcome (in kg), available at baseline (ie, confirmation of pregnancy) and follow-up (ie, last available weight recorded before delivery). Other outcomes include pre-eclampsia and stillbirth. A key aim was to examine if the intervention effect on these maternal and fetal outcomes is modified by (ie, interacts with) participant-level covariates such as the mother's age and body mass index (BMI) at baseline.

## 3 | TWO-STAGE AND ONE-STAGE IPD META-ANALYSIS MODELS FOR ESTIMATING TREATMENT-COVARIATE INTERACTIONS

In this section, we introduce the framework of two-stage and one-stage models for conducting an IPD meta-analysis to examine treatment-covariate interactions based on within-study information.

### 3.1 | The two-stage approach

A two-stage IPD meta-analysis for summarizing treatment-covariate interactions is a straightforward application of a traditional meta-analysis framework. In the first stage the treatment-covariate interactions are estimated using the IPD from each trial separately; then in the second stage these interaction estimates are pooled using a traditional (eg, inverse-variance weighted) meta-analysis model.

Consider an IPD meta-analysis of multiple randomized trials, each comparing the effect of a particular treatment relative to a control using a simple parallel group design. Let  $i$  denote trial ( $i = 1$  to  $S$ ),  $n_i$  denote the number of participants in the  $i$ th trial,  $j$  denote participant ( $j = 1$  to  $n_i$ ), and  $x_{ij}$  denote allocation to either the treatment ( $x_{ij} = 1$ ) or control ( $x_{ij} = 0$ ) group for the  $j$ th participant in the  $i$ th trial. Let  $z_{ij}$  be a key participant-level covariate of interest (eg, the sex of participant  $j$  in trial  $i$ ), observed for all participants in each trial. Then, the two-stage approach can be detailed as follows.

### 3.1.1 | First stage

The analysis to apply in the first stage depends on the outcome type. For example, assume the aim is to evaluate the treatment effect on a continuous outcome ( $y_{ij}$ ), such as SBP, then the first stage of the IPD meta-analysis should apply a linear regression *in each trial separately*, adjusting for the baseline SBP ( $y_{0ij}$ ) and the covariate ( $z_{ij}$ ), whilst including the treatment ( $x_{ij}$ ) and the treatment-covariate interaction ( $x_{ij}z_{ij}$ ):

$$\begin{aligned} y_{ij} &= \alpha_i + \beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}z_{ij} + e_{ij} \\ e_{ij} &\sim N(0, \sigma_i^2). \end{aligned} \quad (1)$$

The key parameter in model (1) is the treatment-covariate interaction ( $\gamma_{Wi}$ ), which represents the change in the treatment effect (ie, the change in the mean difference in outcome value for treatment compared to control) for a 1-unit increase in  $z_{ij}$  after adjusting for the prognostic effects of  $z_{ij}$  and  $y_{0ij}$ . Other terms are  $\alpha_i$  (the intercept, that is, the expected outcome value for participants in the control group with zero values of  $z_{ij}$  and  $y_{0ij}$ ),  $\beta_{1i}$  and  $\beta_{3i}$  (the expected change in the outcome value for a 1-unit increase in  $z_{ij}$  and  $y_{0ij}$ , respectively) and  $\beta_{2i}$  (the treatment effect for those with a  $z_{ij}$  of zero after adjusting for baseline). The same residual variance ( $\sigma_i^2$ ) is assumed for the treatment and control groups, but this can be relaxed if considered appropriate.<sup>18</sup>

If a binary outcome was rather of interest (ie,  $y_{ij} = 0$  or 1), then alternatively a binomial regression could be fitted in each trial separately, such as a logistic regression,

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(p_{ij}) \\ \ln\left(\frac{p_{ij}}{1-p_{ij}}\right) &= \alpha_i + \beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}z_{ij}, \end{aligned} \quad (2)$$

where  $p_{ij}$  is the probability of an outcome event (ie,  $y_{ij} = 1$ ) for the  $j$ th participant in the  $i$ th trial conditional on their covariate values. The parameters are defined similar to those for model (1), except now we model the log-odds of the outcome event (ie,  $\ln\left(\frac{p_{ij}}{1-p_{ij}}\right)$ ) such that the treatment effect is measured by a log odds ratio. Hence, the treatment-covariate interaction ( $\gamma_{Wi}$ ) represents the change in the log odds ratio for a 1-unit increase in  $z_{ij}$  after adjusting for the prognostic effects of  $z_{ij}$  and  $y_{0ij}$ .

If a time-to-event outcome was of interest, then a Cox proportional hazards regression model could be fitted in each trial separately, such as

$$\lambda_{ij}(t) = \lambda_{0i}(t) \exp(\beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}z_{ij}), \quad (3)$$

where  $\lambda_{ij}(t)$  is the hazard rate of the outcome for the  $j$ th participant in the  $i$ th trial conditional on their covariate values, whilst  $\lambda_{0i}(t)$  is the baseline hazard (ie, the hazard rate for a participant in the control group with zero values of  $z_{ij}$  and  $y_{0ij}$ ). Now the treatment effect is measured by a log hazard ratio, such that the treatment-covariate interaction ( $\gamma_{Wi}$ ) represents the change in the log hazard ratio for a 1-unit increase in  $z_{ij}$  after adjusting for the prognostic effects of  $z_{ij}$  and  $y_{0ij}$ .

Extension of models (1) to (3) to adjust for further baseline covariates (ie, in addition to  $z_{ij}$  and  $y_{0ij}$ ) is recommended when there are other known prognostic factors.<sup>19</sup> Ideally, the researcher should predefine a core set of strong prognostic factors to be adjusted for in each trial, particularly focusing on those routinely recorded in the trials available for IPD meta-analysis. For example, in many disease fields, age and stage of disease are key prognostic factors and routinely recorded at baseline, such that they can be adjusted for in the analysis. Other more complex trial designs (eg, cluster trials, or multicenter or multiarm trials) would also require the models to be modified, as would allowing for nonproportional hazards in model (3).

Estimation of models (1) to (3) in each study, for example using (restricted) maximum likelihood estimation, produces a treatment-covariate interaction estimate,  $\hat{\gamma}_{Wi}$ , and its variance,  $\text{var}(\hat{\gamma}_{Wi})$ . The “W” is used to emphasize that the interaction,  $\gamma_{Wi}$ , in each study is based solely on within-study information; that is, it is only based on differences in the treatment effect across participant-level covariate values observed within each study. This is an important point (as it avoids using across-study information and potentially introducing aggregation bias), which we return to in the next section. Each  $\gamma_{Wi}$  indicates for trial  $i$  the change in treatment effect for a one-unit increase in the participant-level covariate  $z_{ij}$ . For a continuous covariate, this assumes the effect of the interaction is linear (although extension to nonlinear trends is important, as described later in the article).

### 3.1.2 | Second stage

In the second stage of a two-stage IPD meta-analysis, we simply combine the  $\hat{\gamma}_{wi}$  values across trials in a traditional meta-analysis model, such as a common-effect (sometimes known as a fixed-effect) model,

$$\hat{\gamma}_{wi} \sim N(\gamma_w, \text{var}(\hat{\gamma}_{wi})), \quad (4)$$

or a random-effects model:

$$\begin{aligned} \hat{\gamma}_{wi} &\sim N(\gamma_{wi}, \text{var}(\hat{\gamma}_{wi})) \\ \gamma_{wi} &\sim N(\gamma_w, \tau^2). \end{aligned} \quad (5)$$

After estimation of the chosen meta-analysis model (eg, using restricted maximum likelihood estimation), the estimate of  $\gamma_w$  summarizes the difference in the treatment effect for two individuals who differ in  $z_{ij}$  by one-unit. Based on the linear, logistic, and Cox models used in the first-stage,  $\gamma_w$  represents a difference in mean difference for a continuous outcome, a difference in log odds ratios for a binary outcome (ie,  $\exp(\hat{\gamma}_w)$  gives a ratio of odds ratios), and a difference in log hazard ratios for a time-to-event outcome (ie,  $\exp(\hat{\gamma}_w)$  is a ratio of hazard ratios).

Note that random-effects meta-analysis model (5) allows for between-study heterogeneity in the true treatment-covariate interaction. It may arise due to differences across studies in, for example, the dose of the treatment, the length of follow-up, the way the covariate has been measured, and the magnitude of any interaction. It may also be due to case-mix differences in the study populations, for example leading to between-study differences in the distribution of within-study confounders and even the covariate itself. For example, if a treatment-covariate interaction is nonlinear, and the covariate distribution is narrow in some studies and wide in others, then this will induce between-study heterogeneity in the treatment-covariate interaction, unless the nonlinear association is modeled directly (see later). The magnitude and impact of heterogeneity can be summarized by providing estimates of  $\tau^2$  and using 95% prediction intervals.<sup>20,21</sup> To account for uncertainty in the estimate of  $\tau^2$  when deriving 95% confidence intervals for  $\gamma_w$ , we recommend the Hartung-Knapp Sidik-Jonkman (HKSJ) approach,<sup>22,23</sup> or alternatively using a Bayesian framework.

If some studies do not agree to share their IPD but do provide (either directly or in a publication) the required treatment-covariate interaction estimate and its corresponding standard error (SE), these can be incorporated in the second stage. That is, the  $\hat{\gamma}_{wi}$  derived directly from IPD trials are combined with the  $\hat{\gamma}_{wi}$  extracted (or provided) from non-IPD trials.

### 3.2 | The one-stage approach

A one-stage IPD meta-analysis can also be used to summarize treatment-covariate interactions. This approach analyzes the IPD from all trials in a single step using a general or generalized linear (mixed) model framework,<sup>24,25</sup> or a survival (frailty) model.<sup>26</sup> This allows a more exact likelihood specification than that used in the second stage of the two-stage approach, and thus avoids assuming study-specific treatment-covariate interaction estimates are normally distributed with known variances.<sup>24</sup> Hence one-stage models may be most advantageous when the studies in the IPD meta-analysis have small numbers of participants and/or events.<sup>27,28</sup>

Inclusion of a treatment-covariate interaction term in a one-stage model is not as straightforward as it may seem. Depending on the model specification, the apparently simple inclusion of a global treatment-covariate interaction term may allow across-study information to contribute toward the summary interaction estimate, in combination with within-study information.<sup>8,29,30</sup> This may lead to aggregation bias (also known as ecological bias), which refers to when the information across studies distorts the interaction estimate compared to when using only within-study information.

Let  $\bar{z}_i$  represent the study-specific mean of the covariate  $z_{ij}$ , and let us assume there is potential between-study heterogeneity in the treatment-covariate interaction. Then, to disentangle within-study and across-study information in a one-stage model, there are two key options:

- (i) Center the covariate  $z_{ij}$  about its study-specific mean,  $\bar{z}_i$  and add an additional term which allows the covariate means ( $\bar{z}_i$ ) to explain between-study heterogeneity in the overall treatment effect<sup>7,31,32</sup>;

and/or.

- (ii) Stratify by trial other parameters outside the interaction term, including the parameter representing the treatment effect (ie, the treatment effect at the covariate's reference value, typically  $z_{ij} = 0$  or if the covariate is centered  $z_{ij} = \bar{z}_i$ ).

Implementing either approach should give very similar summary estimates of the treatment-covariate interaction,  $\gamma_W$ , from a one-stage IPD meta-analysis. Approach (i) leads to a one-stage model for continuous, binary, and time-to-event outcomes of the following format<sup>7,8,29</sup>:

$$\begin{aligned} y_{ij} &= \alpha_i + \beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}(z_{ij} - \bar{z}_i) + \varepsilon_{ij} \\ \beta_{2i} &\sim N(\varphi + \gamma_A\bar{z}_i, \tau_1^2) \quad \gamma_{Wi} \sim N(\gamma_W, \tau_2^2), \end{aligned} \quad (6)$$

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(p_{ij}) \\ \text{logit}(p_{ij}) &= \alpha_i + \beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}(z_{ij} - \bar{z}_i) \\ \beta_{2i} &\sim N(\varphi + \gamma_A\bar{z}_i, \tau_1^2) \quad \gamma_{Wi} \sim N(\gamma_W, \tau_2^2), \end{aligned} \quad (7)$$

$$\begin{aligned} \lambda_{ij}(t) &= \lambda_{0i}(t) \exp(\beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}(z_{ij} - \bar{z}_i)) \\ \beta_{2i} &\sim N(\varphi + \gamma_A\bar{z}_i, \tau_1^2) \quad \gamma_{Wi} \sim N(\gamma_W, \tau_2^2). \end{aligned} \quad (8)$$

Parameters in models (6) to (8) are similar to those defined for models (1) to (3) in Section 3.1.1. In addition, the study-specific treatment-covariate interactions ( $\gamma_{Wi}$ ) are assumed normally distributed with a mean  $\gamma_W$  and between-study variance  $\tau_2^2$ , as for random-effects model (5) in the second stage of the two-stage approach. Of key interest is an estimate of  $\gamma_W$ , to summarize the expected change in the treatment effect (eg, log hazard ratio in model (8)) for each one unit increase in  $z_{ij}$ . As previously, “W” indicates that the interaction will be based solely on within-study information.

Models (6) to (8) also include a meta-regression component ( $\varphi + \gamma_A\bar{z}_i$ ), where  $\varphi$  denotes the summary treatment effect for participants with zero values of  $z_{ij}$  and  $y_{0ij}$  in trials with  $\bar{z}_i$  equal to zero; and  $\gamma_A$  is the change in the summary treatment effect for each 1-unit increase in  $\bar{z}_i$ . The models also allow for residual between-study variance ( $\tau_1^2$ ) in the summary treatment effect (ie, that not explained by  $\bar{z}_i$ ). Inclusion of the  $\gamma_A\bar{z}_i$  term, together with the centering of  $z_{ij}$  within the interaction term, disentangles  $\gamma_W$  and  $\gamma_A$  such that they are uncorrelated with each other, and thus  $\hat{\gamma}_W$  will be based solely on within-trial information.<sup>8,32</sup> Note that if the  $\gamma_A\bar{z}_i$  term is not included in model (8), then the interaction term will represent some weighted average of  $\hat{\gamma}_W$  and the magnitude of aggregation bias ( $\hat{\gamma}_E = \hat{\gamma}_A - \hat{\gamma}_W$ ). Other parameters (ie, intercepts, baseline hazards, prognostic effects of  $z_{ij}$  and  $y_{0ij}$ ) are stratified by trial (ie, estimated separately for each trial).

Approach (ii) leads to one-stage models for continuous, binary and time-to-event outcomes of the following format:

$$\begin{aligned} y_{ij} &= \alpha_i + \beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}z_{ij} + \varepsilon_{ij} \\ \gamma_{Wi} &\sim N(\gamma_W, \tau^2), \end{aligned} \quad (9)$$

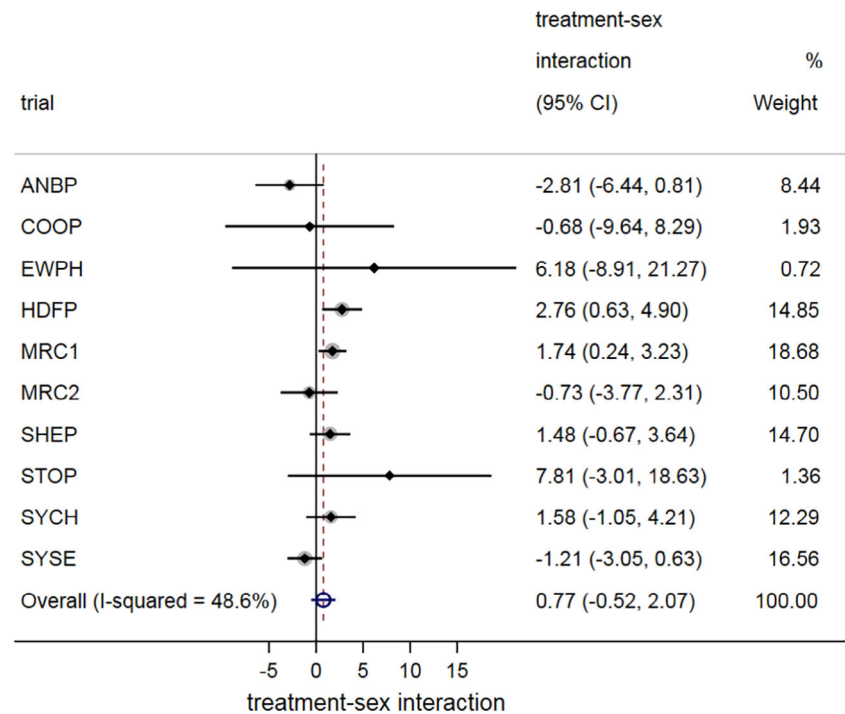
$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(p_{ij}) \\ \text{logit}(p_{ij}) &= \alpha_i + \beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}z_{ij} \\ \gamma_{Wi} &\sim N(\gamma_W, \tau^2), \end{aligned} \quad (10)$$

$$\begin{aligned} \lambda_{ij}(t) &= \lambda_{0i}(t) \exp(\beta_{1i}z_{ij} + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + \gamma_{Wi}x_{ij}z_{ij}) \\ \gamma_{Wi} &\sim N(\gamma_W, \tau^2). \end{aligned} \quad (11)$$

In models (9) to (11) each of the “nuisance” parameters (ie, the  $\alpha_i$ ,  $\lambda_{0i}(t)$ ,  $\beta_{1i}$ ,  $\beta_{2i}$ , and  $\beta_{3i}$  parameters which are not of primary interest) are stratified by trial, and random effects are placed only on the within-study interaction. This stratification of all nuisance parameters by trial, in particular  $\beta_{2i}$  representing the treatment effect at the covariate's reference value, ensures that  $\gamma_W$  only contains within-trial information. It also closely reflects the two-stage approach, where all nuisance parameters are naturally stratified by trial as they are estimated in each trial separately in the first stage.

As mentioned for the two-stage approach, all these one-stage models (6) to (11) need to be modified for other trial designs (eg, cluster trials) and when adjusting for further prognostic factors (ie, in addition to  $z_{ij}$  and  $y_{0ij}$ ). The latter is a sensible strategy (see discussion in Section 3.1.1), but again requires each prognostic factor's effect to be stratified by study. This will increase the number of parameters and so, for estimation reasons, centering each included prognostic factor by

**FIGURE 1** A two-stage IPD meta-analysis of treatment-sex interactions, summarizing the difference in the effect of antihypertensive treatment for males compared to females. *Note:* The interactions refer to the difference between males and females in the treatment effect (ie, their difference in the mean difference in final systolic blood pressure for treatment vs control, after adjusting for baseline), with negative values indicating that the treatment effect is better in females than males [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



its study-specific mean is a sensible default approach (especially when unrestricted maximum likelihood estimation is used to fit the one-stage model).

### 3.3 | Applied example: Is the effect of antihypertensive treatment different for males and females?

The two-stage and one-stage approaches are now illustrated using the IPD from the 10 randomized trials examining the effect of antihypertensive treatment on SBP, as introduced in Section 2.1. The question is whether the treatment effect is different for males compared to females; that is, whether there is a treatment-sex interaction.

#### Two-stage approach

In the first stage, restricted maximum likelihood (REML) estimation was used to fit model (1) with final SBP as the response, and sex (males = 1, females = 0) as the covariate,  $z_{ij}$ , of interest. Figure 1 provides a forest plot of the treatment-sex interaction estimates,  $\hat{\gamma}_{W_i}$ , plotted as circles. Circles are recommended by Fisher et al to help distinguish a forest plot of interaction estimates from a standard forest plot of treatment effect estimates,<sup>6,9,33</sup> for which squares are typically used. In the second stage, REML estimation was used to fit random-effects meta-analysis model (5) and this gave a summary interaction estimate of  $\hat{\gamma}_W = 0.77$  (95% CI:  $-0.52$  to  $2.07$ ). Hence there is no evidence of an important difference in treatment effect for males compared to females; the summary interaction estimate is close to zero and clinically unimportant, the confidence interval overlaps zero, and there is also between-study heterogeneity. Using the HKSJ approach, the 95% confidence is slightly wider ( $-0.73$  to  $2.27$ ).

#### One-stage approach

For the one-stage approach, we used REML estimation to fit models (6) and (9). This gave the same summary interaction estimate of  $\hat{\gamma}_W = 0.77$  (95% CI:  $-0.52$  to  $2.06$ ), practically identical to that from the two-stage approach when REML and a Wald-based confidence interval are used. When we rather used the Satterthwaite approach to derive confidence intervals following estimation of the one-stage model (as recommended for continuous outcomes<sup>34,35</sup>) the obtained 95% confidence interval for  $\gamma_W$  was slightly wider ( $-0.72$  to  $2.26$ ), and practically identical to that derived using the HKSJ method in the two-stage approach (as they both utilize the  $t$ -distribution rather than the standard normal distribution). Thus, one-stage and two-stage IPD meta-analysis results for the treatment-sex interaction are almost identical. This agrees with Burke et al who describe in detail that when the same assumptions and the same estimation approaches are applied, one-stage and two-stage approaches will closely agree unless most studies are small.<sup>28</sup> In conclusion, this example does not support there being an important treatment-sex interaction.

## 4 | STATISTICAL MODELING RECOMMENDATIONS WHEN CONDUCTING AN IPD META-ANALYSIS TO EXAMINE TREATMENT-COVIARIATE INTERACTIONS

Some published IPD meta-analysis projects that examine treatment-covariate interactions use inappropriate statistical approaches that ignore or incorrectly modify the models described in Section 3.<sup>9</sup> We now make five recommendations to address the key concerns.

### 4.1 | Do not make inferences about interactions using the summary treatment effect derived in each subgroup separately

When the participant-level covariate is categorical, it may be tempting to perform a one-stage or two-stage IPD meta-analysis of the overall treatment effect in each category (subgroup) separately. For example, an IPD meta-analysis might be conducted for males and females separately, to obtain the summary treatment effect for males and females. However, Fisher et al and Belias et al show that it is dangerous to use the subsequent results to make inferences about whether an interaction exists.<sup>9,36</sup> In particular, a common mistake is to conclude a treatment-covariate interaction exists if the summary treatment effect estimate is statistically significant in one subgroup but not the other. In this situation the actual treatment-covariate interaction (difference between subgroups based on within-study information) may not be statistically significant. Altman and Bland consider this eloquently.<sup>37</sup>

It is even flawed to compare the summary treatment effects for each subgroup, for example via a statistical test or by calculating their difference. Although this is apparently simple, it amalgamates within-study and across-study information, and so should be avoided. The larger the differences in mean covariate values across studies, the larger the potential contribution of the across-study information. One way to understand this is to consider an extreme situation, where a treatment-sex interaction is of interest, but some trials contain only males. Such trials cannot contribute any within-study information about the interaction between treatment effect and sex at the individual level, as there are no females. However, the trial would still contribute toward the subgroup result for males, and thus subsequently toward the difference between meta-analysis results for male and female subgroups. This issue is avoided by meta-analyzing the treatment-covariate interaction estimates observed within trials, by using the two-stage and one-stage approaches outlined in Section 3.

#### *Application to the hypertension example*

Recall that a two-stage IPD meta-analysis using only within-study information gave a summary treatment-sex interaction estimate of  $\hat{\gamma}_W = 0.77$  (95% CI:  $-0.52$  to  $2.07$ ). When we do a separate IPD meta-analysis for males and females, we obtain summary treatment effect estimates of  $-9.02$  (95% CI:  $-10.47$  to  $-7.57$ ) and  $-10.59$  (95% CI:  $-12.60$  to  $-8.58$ ), respectively. The estimated difference in these subgroup results is  $1.57$ , which is an amalgamation of within-trial and across-trial information, and about twice the size of the summary interaction estimate of  $0.77$  based solely on within-study information. The across-trial information arises because the proportion of males varies considerably across trials, from about  $0.23$  to  $0.70$ .

### 4.2 | Separate within-trial and across-trial information in one-stage models

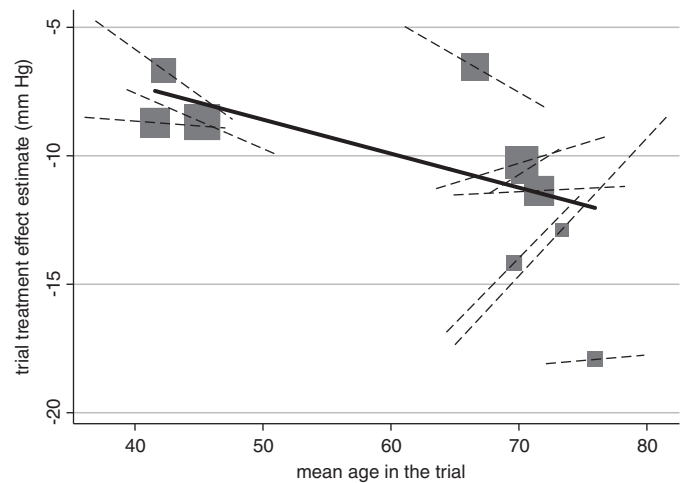
One-stage IPD meta-analysis models are appealing to statisticians, as they eloquently synthesize all the IPD in a single statistical model as described in Section 3.2. However, compared to the two-stage approach, it is easier to make modeling errors when estimating  $\gamma_{WA}$ . One-stage models (6) to (11) show how to separate within-study and across-study information, and ensure that  $\gamma_{WA}$  is estimated based solely on within-study information. As such, they are labeled as “def” by Fisher et al.<sup>9</sup> However, many researchers wrongly fit models (6) to (8) without the  $\gamma_A \bar{z}_i$  term, and with  $\gamma_W x_{ij} (z_{ij} - \bar{z}_i)$  replaced by  $\gamma_{WAi} x_{ij} z_{ij}$ . For example, a flawed one-stage linear regression model for a continuous outcome is:

$$\begin{aligned} y_{ij} &= \alpha_i + \beta_{1i} z_{ij} + \beta_{2i} x_{ij} + \beta_{3i} y_{0ij} + \gamma_{WAi} x_{ij} z_{ij} + e_{ij} \\ \gamma_{WAi} &\sim N(\gamma_{WA}, \tau_{\gamma_{WA}}^2) \quad \beta_{2i} \sim N(\beta_2, \tau_{\beta_2}^2) \quad e_{ij} \sim N(0, \sigma_i^2). \end{aligned} \quad (12)$$

Here, the interaction  $\gamma_{WA}$  is an amalgamation of within-study information and across-study information; essentially a weighted average of  $\gamma_W$  and  $\gamma_A$ .<sup>8</sup> The estimate of  $\gamma_{WA}$  from model (12) will be more precise than the estimate of  $\gamma_W$  from



**FIGURE 2** Does age interact with the effect of antihypertensive treatment on systolic blood pressure? Findings from an IPD meta-analysis of 10 trials, showing difference in results based on across-study (solid line) and within-study (dashed lines) information. Across-study relationship (from meta-regression of trial's treatment effect estimates vs mean age) denoted by gradient of solid line. Participant-level relationship using within-study information (ie, treatment-sex interaction within each trial) denoted by gradient of dashed lines and the summary gradient ( $\hat{\gamma}_W$ ) is  $-0.036$  (95% CI:  $-0.19$  to  $0.12$ ). Each block represents one trial, and the block size is proportional to the size of the trial.



model (6), but at the expense of the across-study information distorting the estimate and interpretation compared to  $\gamma_W$ . This motivates Fisher et al to call it a “deluded” analysis.<sup>9</sup>

#### Application to the hypertension example

Returning to the hypertension example, let us consider whether there is a treatment-age interaction whilst assuming that the effect of age is linear. REML estimation of one-stage model (6) gives  $\hat{\gamma}_W = -0.036$  (95% CI:  $-0.19$  to  $0.12$ ), and thus no clear evidence of a treatment-age interaction. When rather fitting model (12), we obtain  $\hat{\gamma}_{WA} = -0.079$  (95% CI:  $-0.14$  to  $-0.02$ ), which is almost twice the size of that based solely on within-study information and now strongly suggests a treatment-age interaction. The difference is because  $\hat{\gamma}_{WA}$  is strongly influenced by the across-study information, which is suggesting a different magnitude of effect than suggested by the within-study information in half of the trials. This is illustrated in Figure 2, which also suggests that there may be a nonlinear interaction with age, as the direction of the interactions appears to be different in trials with younger and older ages. Extension to nonlinear trends is considered later.

### 4.3 | Do not dichotomize continuous covariates or outcomes

Categorization, and in particular dichotomization, of continuous covariates (eg, such as age, blood pressure, and most biomarkers) using cut-points is best avoided when examining their interaction with treatment effect.<sup>38-40</sup> The usual argument for categorization is to aid clinical interpretation and maintain simplicity. However it can rarely, if ever, be justified that an individual whose value is just below the cut-point is completely different from an individual whose value is just above it. Moreover, categorization reduces the power to detect genuine covariates that interact with treatment effect. Ensor et al show that, in the aforementioned IPD meta-analysis to examine exercise interventions to reduce gestational weight gain in pregnancy, the loss of information by dichotomizing BMI (rather than keeping BMI as continuous) is equivalent to throwing away about one-third of the IPD available.<sup>41</sup>

Categorization of continuous *outcomes* should also be avoided, as it can mislead researchers into thinking there are differential responses to treatment. In particular, classifying individuals as either responders or nonresponders, based on an arbitrary cut-point value for a continuous outcome (eg, SBP  $< 120$  mmHg at follow-up), will lead to those just above the threshold being classed differently to those just below it, which is nonsense. Such dichotomization also leads to misclassification when there is measurement error.<sup>42</sup>

#### Application to the pregnancy example

In the pregnancy example an outcome of interest was caesarean section and, across all participants in 32 trials, the intervention reduces the odds of caesarean section by about 10% (summary OR = 0.91; 95% CI: 0.83 to 0.99). It is of interest whether the mother's age interacts with this intervention effect. If we arbitrarily dichotomize age at 35 years, so that it becomes a binary variable (ie, 0 if age  $\leq 35$ , and 1 if age  $> 35$ ), then a two-stage IPD meta-analysis (model (2) followed by model (5)) shows no clear evidence of an interaction between age and treatment effect (summary ratio of ORs = 1.14, 95% CI: 0.90 to 1.45). However, if age is kept as a continuous variable, then there is stronger evidence of a treatment-age interaction. The intervention becomes less effective at reducing the odds of caesarean section as a women's age increases. For every 5-year increase in age the odds ratio increases by about 10% (summary ratio of ORs = 1.10, 95% CI: 1.00 to 1.20). When analyzed on the risk (rather than odds scale), using a binomial regression with a log-link, this translates to about a 6% increase in the risk ratio for every 5-year increase (95% CI: 0.98 to 1.15).

#### 4.4 | Allow for potential nonlinear relationships when modeling interactions with continuous covariates

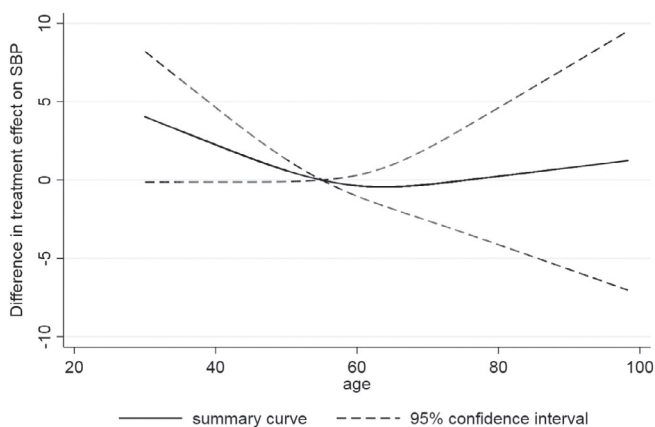
In the previous example we assumed a linear trend for the interaction of treatment and a continuous covariate. However, sometimes the interaction may be nonlinear, as emphasized by Royston and Sauerbrei,<sup>15</sup> and considered in detail by Kasenda et al.<sup>43,44</sup> This implies the change in treatment effect for every 1-unit increase in the covariate may vary across the distribution of the covariate. Therefore, nonlinear interactions should be evaluated when the interaction of a continuous covariate and treatment effect is of interest. We now illustrate this with an example, and explain how it can be implemented using a multivariate meta-analysis of spline functions.

##### *Application to the hypertension example*

Wang et al consider whether there is an interaction between age and the effect of hypertensive treatment, and identify a nonlinear relationship, with older patients between 60 and 80 years old seeming to benefit more than younger patients and those older than 80.<sup>17</sup> Their original analysis categorized age, but we now update their analysis to rather display the change in treatment effect as a smooth, nonlinear function of age (Figure 3), with the reference group being the treatment effect for a 55 year old. A J-shaped relationship is visible; in particular, there is strong evidence that younger individuals have the smallest treatment effect. For example, compared to an individual aged 40 years, an individual aged 55 years has about a 3 mmHg greater reduction in SBP due to the treatment (compared to control). In very old ages the treatment effect also appears to reduce slightly, but there is large uncertainty (wide confidence intervals). Interestingly, if we just include an interaction with age (ie, assume a linear interaction term) then there is no evidence of a treatment-age interaction. The change in treatment effect for each year increase of age is  $-0.05$  (95% CI:  $-0.19$  to  $0.12$ ). The linear assumption hides the more J-shaped interaction revealed by the nonlinear modeling approach.

##### *Conducting a two-stage multivariate IPD meta-analysis of restricted cubic splines*

Figure 3 was obtained by conducting a two-stage IPD meta-analysis, where in the first stage a restricted cubic spline function was estimated in each study separately and in the second stage a multivariate meta-analysis was fitted. Splines are a flexible way of modeling smooth nonlinear relationships.<sup>45</sup> Briefly, a restricted cubic spline is obtained by fitting a series of cubic functions and forcing them to join (and be smoothed) at certain points (called internal knots), whilst constraining the function to be linear in the tails (ie, before the first internal knot and after the last internal knot). The magnitude and shape of the curve are defined by multiple parameters depending on the number of knots chosen; as explained in Figure 4. The knot locations are forced to be the same in every study, to ensure the study-specific curves can be synthesized meaningfully. Rather than using a reference group whose covariate value is 0, it helps to center the covariate at a meaningful value (eg, 55 years was chosen as the reference group in Figure 3), which is the same in every study. In our example, three internal knots (at the same locations in every study) were chosen for the restricted cubic function representing the association between age and final SBP in the control group. Hence, in the first stage three parameters defining the spline function are estimated in each study (an intercept and two slope terms), plus the interaction of this spline function with the treatment effect. The latter provides the within-study treatment-covariate interaction defined by  $\gamma_{W1i}$  and  $\gamma_{W2i}$  (see Figure 5A), which represent the change (difference) in treatment effect across covariate values, relative to the chosen reference group of an individual aged 55 years.



**FIGURE 3** Summary of the nonlinear interaction between age and effect of hypertension treatment on final SBP value. Figure created by fitting an analysis of covariance model in each study separately, with the interaction between age and treatment modeled via a restricted cubic spline function (with knot positions of 39, 60, and 75), and then the study-specific parameter estimates (relating to the interaction) pooled in a multivariate random-effects meta-analysis

Let the original continuous covariate (e.g. age) be denoted by  $z$ . Rather than simply assuming a linear association with the outcome of interest (e.g.  $Y = \delta_0 + \delta_1 z$ ), a restricted spline function, denoted by  $f(z)$ , allows for a potential non-linear association.<sup>45</sup> It is obtained by fitting a series of cubic functions and forcing them to join (and be smoothed) at certain points (called internal knots), whilst constraining the function to be linear in the tails (i.e. before the first internal knot and after the last internal knot). The magnitude and shape of the curve is defined by multiple parameters depending on the number of knots chosen. If we assume there are  $k$  internal knots in total at locations  $t_1, t_2, \dots, t_k$ , then the restricted cubic spline is:

$$f(z) = \delta_0 + \delta_1 z_1 + \delta_2 z_2 + \dots + \delta_{k-1} z_{k-1}$$

where  $z_1 = z$  (and thus there is an assumed linear association between  $z$  and the outcome before the first internal knot (i.e. when  $z < t_1$ ), and for  $c = 1, \dots, k - 2$ ,

$$z_{c+1} = (z - t_c)_+^3 - (z - t_{k-1})_+^3 \frac{(t_k - t_c)}{(t_k - t_{k-1})} + (z - t_k)_+^3 \frac{(t_{k-1} - t_c)}{(t_k - t_{k-1})}$$

where the  $A_+$  notation means that  $A = A$  if  $A > 0$ , and  $A = 0$  if  $A \leq 0$ . This specification of  $z_{c+1}$  also forces a linear association between  $z$  and the outcome after the last internal knot (i.e. when  $z \geq t_k$ ).

Therefore, using a restricted cubic spline in a regression analysis will include the original covariate ( $z$ ) as linear and  $k - 2$  piecewise cubic variables. So with  $k$  internal knots, there are  $k - 1$  parameters to estimate which define the spline function, plus an 'intercept' term  $\delta_0$ . We can examine whether the use of the spline function (i.e.  $f(z)$ ) adds value over and above assuming a linear trend by comparing the change in the model fit (e.g. the likelihood ratio statistic). In terms of how to choose the number of knots, Harrell states that "for many datasets,  $k = 4$  offers an adequate fit of the model and is a good compromise between flexibility and loss of precision caused by overfitting a small sample".<sup>45</sup> If the sample size is small, three knots should be used in order to have enough observations in between the knots to be able to fit each polynomial. The location of the knots are best pre-specified, based on the quantiles of the continuous variable, with the following suggested by Harrell:<sup>45</sup>

Number of internal knots, $k$	Knots locations in terms of quantiles of the $z$ variable						
3	0.1	0.5	0.9				
4	0.05	0.35	0.65	0.95			
5	0.05	0.275	0.5	0.725	0.95		
6	0.05	0.23	0.41	0.59	0.77	0.95	
7	0.025	0.1833	0.3417	0.5	0.6583	0.8167	0.975

**FIGURE 4** Introduction to modeling nonlinear relationships in a single study using restricted cubic splines; for further details we recommend the reader refer to Harrell<sup>45</sup>

For summarising a treatment-covariate interaction across multiple studies, where the covariate ( $z$ ) is continuous and has a potentially non-linear association with the outcome, we can apply a two-stage IPD multivariate random-effect meta-analysis utilising a restricted cubic spline function. The key steps are now described, in relation to the hypertension example of Figure 3 which examined a treatment-age interaction.

- **Create the restricted spline transformation of  $z$ , which requires choosing the number of knots and their location (as described in Figure 4).** The number and location of the knots must be identical for each study. Usually 3 or 4 knots will suffice. Their location could be defined by quantiles of the covariate distribution observed within the entire IPD from all studies.<sup>45</sup> However, if the distribution of the covariate varies considerably across studies, then knot locations might be modified so that they fall at relevant places.

For example, in the hypertension example, some studies included only older patients (> 60 years). Hence, we considered it important that one of the knots for age was located at 60. We chose 3 knots, and placed these at knot positions of 39 (the 0.1 quantile value), 60 (the 0.48 quantile value) and 75 (the 0.9 quantile value).

- **Estimate the treatment-covariate interaction in each study.** This requires the specification and estimation of a suitable regression model, followed by storing the parameter estimates (and corresponding variance matrix) that define the treatment-covariate interaction. The module *mvmeta\_make* in Stata allows the user to perform a particular regression analysis in each study, and automatically store the relevant estimates and corresponding variance matrix.<sup>52,53</sup>

For example, in the hypertension example we fitted the following model in each study separately:

$$y_{ij} = \alpha_i + f(\text{age}_{ij}) + \beta_{2i}x_{ij} + \beta_{3i}y_{0ij} + f(x_{ij}\text{age}_{ij}) + e_{ij}$$

$$e_{ij} \sim N(0, \sigma_i^2)$$

with the restricted cubic spline function defined by,

$$f(\text{age}_{ij}) = \delta_{1i}\text{age}_{1ij} + \delta_{2i}\text{age}_{2ij}$$

and  $\text{age}_{1ij}$  and  $\text{age}_{2ij}$  denoting the first and second spline transformations of age (see Figure 4), respectively. The interaction between the spline function and treatment is defined by,

$$f(x_{ij}\text{age}_{ij}) = \gamma_{W1i}x_{ij}\text{age}_{1ij} + \gamma_{W2i}x_{ij}\text{age}_{2ij}$$

Of key interest are estimates of  $\gamma_{W1i}$  and  $\gamma_{W2i}$ , as these define the treatment-covariate interaction function in study  $i$ , together with their variances ( $\text{var}(\gamma_{W1i})$  and  $\text{var}(\gamma_{W2i})$ ) and covariance ( $\text{cov}(\hat{\gamma}_{W1i}, \hat{\gamma}_{W2i})$ ).

**FIGURE 5** (A): Overview of the first stage of a two-stage multivariate IPD meta-analysis to summarize a nonlinear treatment-covariate interaction using a restricted cubic spline. The steps are described in relation to the hypertension example of Figure 3 which examined a treatment-age interaction.

To aid interpretation, before model estimation it may be helpful to centre the spline transformations by a useful reference value. In the hypertension example, we repeated the above analysis in each study after centering  $age_{1ij}$  and  $age_{2ij}$  by values that corresponded to an age of 55 years. These led to  $(age_{1ij} - 55)$  and  $(age_{2ij} - 3.16)$ .

Note that if some parameters cannot be estimated in some studies (e.g. due to a narrow distribution of  $z_{ij}$ , or perfect prediction), then data augmentation can be used (for example, via *mvmeta\_make*<sup>52</sup>). Essentially this leads to an arbitrary estimate of the problematic parameter but with a very large variance (e.g. 1000000000) and any associated covariances set to zero, such that the estimates will receive barely any weighting in the subsequent multivariate meta-analysis.

- **Perform a multivariate random-effects meta-analysis of the treatment-covariate interaction estimates, to produce a summary of the treatment-covariate interaction.** The multivariate approach allows the joint synthesis of multiple parameter estimates, whilst accounting for their correlation.<sup>54,55</sup> It can be fitted, for example, using REML estimation.

In the hypertension example we fitted a bivariate random-effects meta-analysis,

$$\begin{pmatrix} \hat{\gamma}_{W1i} \\ \hat{\gamma}_{W2i} \end{pmatrix} \sim N \left( \begin{pmatrix} \gamma_{W1i} \\ \gamma_{W2i} \end{pmatrix}, \begin{pmatrix} \text{var}(\hat{\gamma}_{W1i}) & \text{cov}(\hat{\gamma}_{W1i}, \hat{\gamma}_{W2i}) \\ \text{cov}(\hat{\gamma}_{W1i}, \hat{\gamma}_{W2i}) & \text{var}(\hat{\gamma}_{W2i}) \end{pmatrix} \right)$$

$$\begin{pmatrix} \gamma_{W1i} \\ \gamma_{W2i} \end{pmatrix} \sim N \left( \begin{pmatrix} \gamma_{W1} \\ \gamma_{W2} \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{pmatrix} \right)$$

where  $\tau_1^2$  and  $\tau_2^2$  define the between-study variances of  $\gamma_{W1i}$  and  $\gamma_{W2i}$ , respectively, and  $\tau_{12}$  defines their between-study correlation.

The summary estimates of  $\hat{\gamma}_{W1}$  and  $\hat{\gamma}_{W2}$  define the summary treatment-age interaction of  $\gamma_{W1}x_{ij}(age_{1ij} - 55) + \gamma_{W2}x_{ij}(age_{2ij} - 3.16)$ .

- **Plot the summary treatment-covariate interaction and its confidence interval across the distribution of covariate values.** After estimation of the multivariate model, the summary treatment-covariate interaction function can be applied (e.g. via the *predict* post-estimation command in Stata) to each individual in the original IPD, to obtain their predicted treatment-covariate interaction; i.e. the difference in their treatment effect compared to that for the reference individual (in our example, this individual was 55 years old). This predicted value can then be plotted (on the y-axis) against the original covariate value (on the x-axis). The standard error (s.e.) of the predicted value can also be estimated for each individual, and then a confidence interval calculated. See Figure 3.

**FIGURE 5** (B): Overview of the second stage of a two-stage multivariate IPD meta-analysis to summarize a nonlinear treatment-covariate interaction using a restricted cubic spline. The steps are described in relation to the hypertension example of Figure 3 which examined a treatment-age interaction

In the second stage, the study estimates of the difference in slope parameters of the spline function (ie,  $\hat{\gamma}_{W1i}$  and  $\hat{\gamma}_{W2i}$ ) can be meta-analyzed using a multivariate random-effects meta-analysis.<sup>46-50</sup> This allows the synthesis of multiple parameter estimates whilst accounting for their within-study and between-study correlation, and produces a summary estimate for each parameter, from which the summary spline (nonlinear) function can be derived. It can handle missing parameter estimates (ie, for missing parts of the spline function in studies with a narrow distribution of covariate values) as described in Figure 5A,B. This summary curve describes the association between values of the covariate and the change in treatment effect, relative to the reference group (age 55 years in the example). This can then be plotted graphically to aid interpretation. The study-specific estimated curves from the first stage (or study-specific empirical Bayes curves obtained postestimation from the second stage) might also be presented, as shown by Gasparrini et al.<sup>46</sup>

A summary of the two-stage multivariate IPD meta-analysis of restricted spline functions is given in Figure 5A,B. A similar approach is a multivariate IPD meta-analysis of a polynomial function<sup>51</sup>; in the first stage the differences (ie, treatment minus control) of the parameters of a chosen polynomial function are estimated in each study (eg, differences in a quadratic shape, such as  $\gamma_{W1}x_{ij}z_{ij} + \gamma_{W2}x_{ij}z_{ij}^2$ ) and the parameter estimates ( $\hat{\gamma}_{W1}$  and  $\hat{\gamma}_{W2}$ ) are jointly synthesized to produce a summary function.<sup>51</sup> In particular, fractional polynomial functions provide a flexible set of power transformations to describe a potentially nonlinear association.<sup>52-54</sup> In order for the parameters of the function to be combinable across studies in a multivariate meta-analysis, the same powers of the fractional polynomial function must be specified in each study (ie, the shape of the nonlinear association is fixed across studies).<sup>51</sup> In contrast, as long as the same number and location of knots is used in each study, a restricted cubic spline function is more flexible (ie, the shape of the nonlinear association can vary across studies) and it can handle different covariate distributions across studies by strategic placement of knots. Sauerbrei and Royston show that it is possible to combine different fractional polynomial power transformations across studies, if the predicted values of the function are pooled (rather than the parameters defining the function),<sup>55</sup> with the pooling of predictions done at each value of the covariate separately.

A one-stage IPD meta-analysis model can also be used to examine nonlinear treatment-covariate interactions, but difficulties can arise. First, the extra parameters required to model nonlinear functions, and potentially multiple random effects, may cause estimation and convergence problems. Second, when centering covariates by their study-specific means (ie, to avoid aggregation bias), by extending models (6) to (8), the interpretation of the spline function becomes problematic. Unless all studies have the same mean covariate value, the change in treatment effect for a 1-unit increase in a covariate from its mean will have a different interpretation in each study; this will make the summary spline function uninterpretable. Therefore, it is preferable to examine nonlinear trends by extending one-stage models (9) to (11), stratifying by trial parameters outside the interaction term, to remove aggregation bias.

## 4.5 | To personalize decision-making, individualized predictions of treatment effect are required

For clinical practice, translation of identified treatment-covariate interactions is required for the individual patient, in order to tailor their treatment decisions. This is nontrivial. A fundamental error is to assume that if a treatment-covariate interaction exists then the treatment is effective in some individuals but not in others. Actually, even when the magnitude of treatment effect varies across individuals, the direction of effect may consistently suggest the treatment is beneficial for everyone.

To better guide decision-making, a predicted treatment effect is required for each individual. This requires a robust prediction model equation, developed and validated using methodology principles for clinical prediction models.<sup>2,45,56</sup> In particular, to reduce overfitting (too extreme predictions) the model development might require penalization and shrinkage approaches,<sup>57-59</sup> to reduce the variability of predictions in new datasets, thereby reducing the mean-square error of the predictions. A step further is to predict an individual's absolute outcome value or risk conditional on their prognostic factors and expected treatment effect. These and related issues are discussed in detail elsewhere.<sup>60-63</sup> Crucially, aggregation bias should also be avoided when making predictions, and we now illustrate this with an example.

### *Application to the hypertension example*

Consider the example within Figure 3 that suggests a potential nonlinear interaction between age and the effect of antihypertensive treatment, with younger patients seemingly having less benefit (in terms of reduction in SBP) than older individuals. The smooth curve reflects the treatment-age interaction; it reveals the summary estimate of the *difference* in treatment effect for an individual with a particular age compared to an individual aged 55 years. However, it does *not* tell us the predicted (expected) treatment effect for an individual with a particular age. To obtain this, we performed a

multivariate IPD random-effects meta-analysis, where we estimate and then meta-analyze the two slopes ( $\gamma_{W1i}$  and  $\gamma_{W2i}$ ) of the spline function (as defined in Figure 5A), and also the reference treatment group ( $\beta_{2i}$ , the study-specific treatment effect for individuals aged 55 years, our reference group). Then, the summary estimates ( $\hat{\beta}_2, \hat{\gamma}_{W1}, \hat{\gamma}_{W2}$ ) obtained were used to define the prediction model for an individual's treatment effect conditional on their age:

$$\begin{aligned} \text{Predicted treatment effect for } j\text{th individual} &= \hat{\beta}_2 + \hat{\gamma}_{W1}(\text{age}_{1j} - 55) + \hat{\gamma}_{W2}(\text{age}_{2j} - 3.16) \\ &= -10.66 - (0.251 \times (\text{age}_{1j} - 55)) \\ &\quad + (0.156 \times (\text{age}_{2j} - 3.16)). \end{aligned} \quad (13)$$

Here,  $-10.66$  is the predicted treatment effect for an individual aged 55 years, and  $\text{age}_{1j}$  and  $\text{age}_{2j}$  represent, respectively, the individual's values of the first and second spline transformation of their age. For example, for an individual aged 40, their corresponding  $\text{age}_{1j}$  is 40 and  $\text{age}_{2j}$  is 0.00077, and thus their predicted treatment effect is

$$-10.66 - (0.251 \times (40 - 55)) + (0.15 \times (0.00077 - 3.16)) = -7.39.$$

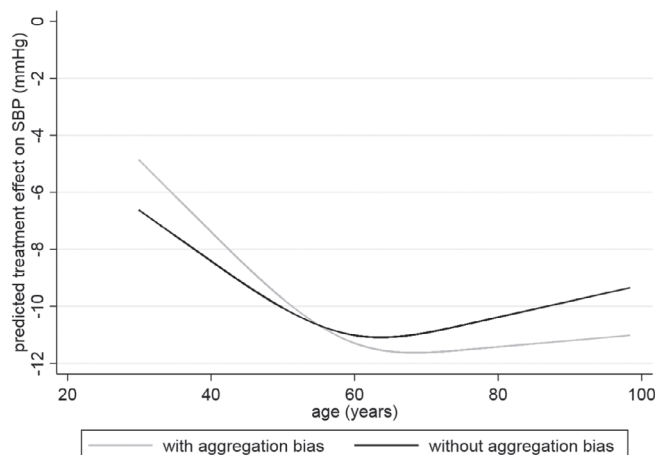
Overfitting is potentially of limited concern for prediction Equation (13), as it was derived using over 28 000 participants with only two parameters (excluding the intercept) estimated. The predicted treatment effect across the age range is summarized in Figure 6. Across all ages the predicted treatment effect is at least 5 mmHg. Hence, on average, all patients are predicted to benefit by a clinically useful amount, even though the magnitude of effect varies due to the treatment-age interaction.

A concern is that a multivariate meta-analysis of  $\hat{\beta}_{2i}, \hat{\gamma}_{W1i}$ , and  $\hat{\gamma}_{W2i}$  accounts for the correlation amongst these parameters, both within and between studies. Yet again this allows the potential for aggregation bias to influence the summary interaction terms,  $\hat{\gamma}_{W1}$  and  $\hat{\gamma}_{W2}$ , as their correlation with  $\hat{\beta}_2$  allows the borrowing of across-trial information. Though it is perhaps sensible for  $\hat{\beta}_2$  (as it is our reference treatment effect averaged across studies), we should want our summary interactions terms to be based solely on within-trial information, as previously argued. To address this, we replace the summary  $\hat{\gamma}_{W1}$  and  $\hat{\gamma}_{W2}$  estimates in our prediction Equation (13) with  $-0.178$  and  $0.133$ , their respective estimates from a multivariate meta-analysis ignoring any correlation with  $\hat{\beta}_{2i}$  terms. Then, the modified prediction equation is:

$$\begin{aligned} \text{Predicted treatment effect for } j\text{th individual} &= \hat{\beta}_2 + \hat{\gamma}_{W1}(\text{age}_{1j} - 55) + \hat{\gamma}_{W2}(\text{age}_{2j} - 3.16) \\ &= -10.66 - (0.178 \times (\text{age}_{1j} - 55)) \\ &\quad + (0.133 \times (\text{age}_{2j} - 3.16)). \end{aligned} \quad (14)$$

This allows us to produce predicted treatment effects conditional on age, summarized across all studies and removing aggregation bias. Using Equation (14) for an individual aged 40 years, their predicted treatment effect is,

$$-10.66 - (0.178 \times (40 - 55)) + (0.133 \times (0.00077 - 3.16)) = -8.41 \text{ mmHg}$$



**FIGURE 6** The predicted effect of antihypertensive treatment on SBP conditional on an individual's age, based on a multivariate meta-analysis either with or without aggregation bias in the summary treatment-age interactions

which is larger than the  $-7.39$  predicted from the previous equation. Indeed, the predicted curve across the entire age range is noticeably shallower after removing aggregation bias (Figure 6), because the aggregation bias (arising from incorporating across trial information) makes interaction estimates larger than when based solely on within-trial information. Note that the predictions shown here are averaged across studies. If possible, incorporation of study-level covariates (eg, country) that explain between-study heterogeneity in parameter estimates might help tailor predictions further.

## 5 | TO IPD OR NOT TO IPD? STATISTICAL RECOMMENDATIONS WHEN PLANNING AN IPD META-ANALYSIS PROJECT TO EXAMINE TREATMENT-COVIARIATE INTERACTIONS

Statistical considerations are also important when deciding whether to initiate an IPD meta-analysis project (ie, before IPD collection), and two key issues are now discussed in the context of treatment-covariate interactions.

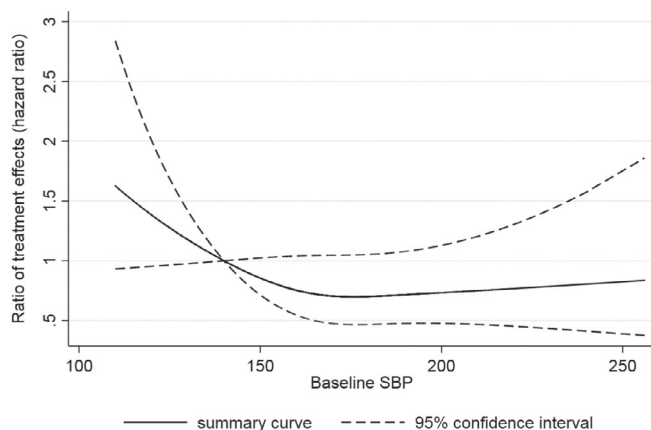
### 5.1 | The decision to initiate an IPD meta-analysis project to examine treatment-covariate interactions should not be based on the amount of between-study heterogeneity in the overall treatment effect

Without IPD, most meta-analyses of randomized trials will summarize the treatment effect based on all trial participants (ie, without consideration of treatment-covariate interactions). If such a meta-analysis does not find evidence of between-study heterogeneity in the treatment effect, it may be tempting to conclude that the treatment effect is the same for all individuals. However, this logic is flawed: the absence of heterogeneity in the overall effect (across all participants) does not necessarily imply an interaction does not exist. First, if there is a genuine treatment-covariate interaction, but the distribution of the covariate is very similar across studies, then (assuming no other patient-level or study-level effect modifiers differ across studies), the overall treatment effect will be the same in each study (ie, there will be no heterogeneity). Second, even if the distribution of the covariate does change across studies, the overall treatment effect may still be homogenous; for example this may arise due to chance, or because the covariate has a nonlinear (eg, U-shaped) interaction with treatment (so that the overall treatment effect may still be the same in two studies with very different mean covariate values), or even due to multiple effect modifiers acting in combination and different directions.

Conversely, if there is heterogeneity in the overall treatment effect this does not imply that an interaction exists at the participant-level. Heterogeneity can arise due to changes in study-level characteristics such as dose, follow-up length, and setting, even when there is no interaction at the participant-level. Hence, the decision to investigate treatment-covariate interactions should not be driven by the presence of between-study heterogeneity in treatment effect, and should be rather motivated by the supposed (biological) mechanism of treatment response.

#### *Application to the hypertension example*

As an illustration of why heterogeneity is a poor indicator of whether treatment-covariate interactions exist, consider a two-stage IPD meta-analysis examining the effect of antihypertensive treatment on rate of CVD, for which the summary hazard ratio was 0.74 (95% CI: 0.67 to 0.81) with no observed heterogeneity ( $\hat{\tau}^2 = 0$ ).<sup>64</sup> However, there is some suggestion



**FIGURE 7** Evidence of a potential nonlinear interaction between baseline SBP and the effect of hypertension treatment on the rate of CVD, even though there was no between-study heterogeneity in the overall treatment effect



of a nonlinear interaction between baseline SBP and the treatment effect on CVD, with the treatment effect gradually reducing as the baseline SBP moves from about 170 mmHg toward 120 mmHg (Figure 7), as obtained using a two-stage multivariate meta-analysis of restricted cubic splines as described in Figure 5A,B.

## 5.2 | Calculate the power to identify a treatment-covariate interaction prior to collection of IPD

IPD meta-analysis projects are more likely to be funded if they have sufficient power to answer the research question at hand. Before IPD collection, power calculations for IPD meta-analysis projects can be made conditional on the number of trials promising their IPD, using known trial characteristics such as the number of participants and standard deviation (SD) of covariate values. Closed-form solutions are difficult to obtain, especially for noncontinuous outcomes unless approximations are made.<sup>65</sup> A pragmatic starting point is to consider whether the total sample size (and overall outcome proportion or rate for binary and time-to-event outcomes) of the promised IPD has adequate power if naively considered to come from a single study. Then standard statistical software for estimating the power of an interaction in a single randomized trial could be used.

To mirror the IPD meta-analysis setting more exactly, a simulation-based approach has been proposed.<sup>66-69</sup> In particular, Ensor et al suggest generating IPD containing the same number of studies (and same number of participants and events therein) as promising their IPD,<sup>41</sup> with covariate values simulated based on assumed true treatment effects and treatment-covariate interaction effect size. Then a two-stage IPD meta-analysis is applied to this simulated dataset, and the estimated treatment-covariate interaction and its confidence interval are stored. This is repeated  $m$  times (ideally thousands). Based on a traditional frequentist paradigm, power can then be estimated by calculating the proportion of times the summary estimate was statistically significant (eg, as defined by the associated 95% confidence interval excluding the null value). This process might also be repeated assuming different values for the size of the assumed treatment-covariate interaction, and different values for the assumed magnitude of heterogeneity in the interaction across studies (starting from zero).

For an IPD meta-analysis of  $S$  randomized trials with continuous outcomes, closed-form solutions are obtainable for the power to evaluate a treatment-covariate interaction.<sup>10</sup> Simmonds and Higgins provide the following analytic solution for the maximum likelihood estimate of a treatment-covariate interaction for a continuous outcome in a single randomized trial with two parallel groups:<sup>10</sup>

$$\hat{\gamma}_{Wi} = \frac{2}{\sum_{j=1}^{n_i} z'_{ij}{}^2} \left( \sum_{j \in T_i} y_{ij} z'_{ij} - \sum_{j \in C_i} y_{ij} z'_{ij} \right).$$

Here,  $T_i$  denotes the treatment group and  $C_i$  the control group in study  $i$ , and  $z'_{ij}$  denotes that each  $z_{ij}$  is centered about the study-specific mean  $z_{ij}$  value (ie,  $z'_{ij} = z_{ij} - \bar{z}_i$ ). Simmonds and Higgins use this to derive subsequent power calculations,<sup>10</sup> assuming common residual variances across trials. If we extend their work by allowing for different residual variances in each trial ( $\sigma_i^2$ ), the variance (var) of the interaction estimate in a particular study  $i$  is:

$$\begin{aligned} \text{var}(\hat{\gamma}_{Wi}) &= \text{var} \left( \frac{2}{\sum_{j=1}^{n_i} z'_{ij}{}^2} \left( \sum_{j \in T_i} y_{ij} z'_{ij} - \sum_{j \in C_i} y_{ij} z'_{ij} \right) \right) \\ &= \frac{4}{\left( \sum_{j=1}^{n_i} z'_{ij}{}^2 \right)^2} \left( \sum_{j \in T_i} z'_{ij}{}^2 \text{var}(y_{ij}) + \sum_{j \in C_i} z'_{ij}{}^2 \text{var}(y_{ij}) \right) \\ &= \frac{4\sigma_i^2}{\left( \sum_{j=1}^{n_i} z'_{ij}{}^2 \right)^2} \left( \sum_{j \in T_i} z'_{ij}{}^2 + \sum_{j \in C_i} z'_{ij}{}^2 \right). \end{aligned} \quad (15)$$

Let us assume an equal number of participants in the treatment and control groups, and that the variance of the covariate ( $\sigma_{z_i}^2$ ) is the same in each treatment group (ie,  $\sigma_{T_i z_i}^2 = \sigma_{C_i z_i}^2 = \sigma_{z_i}^2$ ). Then  $\sum_{j=1}^{n_i} z'_{ij}{}^2 = n_i \sigma_{z_i}^2$ , and Equation (15) simplifies to:

$$\text{var}(\hat{\gamma}_{Wi}) = \frac{4\sigma_i^2}{\left(\sum_{j=1}^{n_i} z_{ij}^{\prime 2}\right)^2} \left(\sum_{j=1}^{n_i} z_{ij}^{\prime 2}\right) = \frac{4\sigma_i^2}{\sum_{j=1}^{n_i} z_{ij}^{\prime 2}} = \frac{4\sigma_i^2}{n_i\sigma_{z_i}^2}. \quad (16)$$

Using the solution for  $\text{var}(\hat{\gamma}_{Wi})$  from either Equation (15) or (16), we can derive a closed-form solution for the variance of the summary interaction estimate from the second stage of a two-stage IPD meta-analysis. Assuming common-effect model (4),  $\text{var}(\hat{\gamma}_W)$  is simply the sum of the inverse of the variances from each study,

$$\text{var}(\hat{\gamma}_W) = \left(\sum_{i=1}^S \frac{1}{\text{var}(\hat{\gamma}_{Wi})}\right)^{-1}, \quad (17)$$

and the subsequent power to estimate a treatment-covariate interaction of size  $\gamma_W$  using the IPD meta-analysis is approximately,

$$\text{Power} = \text{Prob}\left(\frac{\hat{\gamma}_W}{\sqrt{\text{var}(\hat{\gamma}_W)}} > 1.96\right) + \text{Prob}\left(\frac{\hat{\gamma}_W}{\sqrt{\text{var}(\hat{\gamma}_W)}} < -1.96\right). \quad (18)$$

For example, assuming a common interaction and that trials have equally sized groups, we can use Equations (16) and (17) to give,

$$\begin{aligned} \text{Power} &= \text{Prob}\left(\frac{\hat{\gamma}_W}{\sqrt{\text{var}(\hat{\gamma}_W)}} > 1.96\right) + \text{Prob}\left(\frac{\hat{\gamma}_W}{\sqrt{\text{var}(\hat{\gamma}_W)}} < -1.96\right) \\ &= \Phi\left(-1.96 + \gamma_W \sqrt{\sum_{i=1}^S \frac{n_i\sigma_{z_i}^2}{4\sigma_i^2}}\right) \\ &\quad + \Phi\left(-1.96 - \gamma_W \sqrt{\sum_{i=1}^S \frac{n_i\sigma_{z_i}^2}{4\sigma_i^2}}\right), \end{aligned} \quad (19)$$

where  $\Phi(x)$  is the probability of sampling a value  $< x$  from the standard normal distribution. Allowing for different sized groups in a trial (ie, using Equation (15) rather than (16)), the power calculation becomes,

$$\text{Power} = \Phi\left(-1.96 + \gamma_W \sqrt{\sum_{i=1}^S \left(\frac{\left(\sum_{j=1}^{n_i} z_{ij}^{\prime 2}\right)^2}{4\sigma_i^2 \left(\sum_{j \in T_i} z_{ij}^{\prime 2} + \sum_{j \in C_i} z_{ij}^{\prime 2}\right)}\right)}\right) + \Phi\left(-1.96 - \gamma_W \sqrt{\sum_{i=1}^S \left(\frac{\left(\sum_{j=1}^{n_i} z_{ij}^{\prime 2}\right)^2}{4\sigma_i^2 \left(\sum_{j \in T_i} z_{ij}^{\prime 2} + \sum_{j \in C_i} z_{ij}^{\prime 2}\right)}\right)}\right), \quad (20)$$

and as before we could replace  $\sum_{j=1}^{n_i} z_{ij}^{\prime 2}$  with  $n_i\sigma_{z_i}^2$ , and similarly  $\sum_{j \in T_i} z_{ij}^{\prime 2}$  and  $\sum_{j \in C_i} z_{ij}^{\prime 2}$  could be replaced with  $n_{T_i}\sigma_{T_i z_i}^2$  and  $n_{C_i}\sigma_{C_i z_i}^2$ , respectively.

Equations (19) and (20) require the user to specify a minimally important value for  $\gamma_W$ , and also the values of  $n_i$ ,  $\sigma_i^2$ , and  $\sigma_{z_i}^2$  based on published study information (eg, from the baseline characteristics table for a trial) or provided directly by study authors. The residual variance ( $\sigma_i^2$ ) might be unavailable, and so approximated by the variance of outcome values.

Extension can be made to allow for potential heterogeneity in the treatment-covariate interaction, but the amount of heterogeneity is difficult to know in advance. Thus we suggest to focus on the power in an ideal situation where there is no between-study heterogeneity in the size of the interaction.

### 5.2.1 | Application to the pregnancy example

In the IPD meta-analysis conducted by the i-WIP collaboration introduced in Section 2.2, a primary objective was to examine a potential interaction between baseline BMI and intervention effect on gestational weight gain. The prior

hypothesis was that those with high baseline BMI may benefit most from weight management interventions. No formal power calculation was performed in advance of IPD collection but ultimately IPD from 24 trials. For illustration, here we reconstruct the power calculation for this project assuming it was known that IPD could be obtained from these 24 trials. The values of  $n_i$ ,  $\sigma_i^2$ , and  $\sigma_{z_i}^2$  were extracted from the publications of the 24 trials, and are shown by Ensor et al<sup>41</sup>; occasionally  $\sigma_i^2$  and  $\sigma_{z_i}^2$  were missing and for these we used a weighted average of values from other studies. These values were input into Equation (16) to derive  $\text{var}(\hat{\gamma}_{w_i})$  for each trial, which then were used within Equation (17) to give  $\text{var}(\hat{\gamma}_w) = 0.0014$ .

A minimally important interaction size of  $-0.1$  was suggested by clinical experts, such that the reduction in weight is at least 1 kg larger for a 10-unit increase in BMI. Thus inputting  $\gamma_w = -0.1$  and  $\text{var}(\hat{\gamma}_w) = 0.0014$  into Equation (19), we obtain:

$$\begin{aligned} \text{power} &= \Phi\left(-1.96 - 0.1\sqrt{\frac{1}{0.0014}}\right) + \Phi\left(-1.96 + 0.1\sqrt{\frac{1}{0.0014}}\right) \\ &= \Phi(-4.917) + \Phi(0.997) \\ &= 0 + 0.84. \end{aligned}$$

Thus the estimated power is 84%. Ensor et al estimate a comparable 83% power using a simulation-based power calculation,<sup>41</sup> but our closed-form solution is obtained in a much quicker time-frame. Had this power been known at the time of the i-WIP grant application, it would have given further merit to undertaking and funding the IPD meta-analysis project. Although, even if power was deemed low, there are often many other potential benefits of undertaking an IPD meta-analysis project (eg, obtaining additional follow-up, standardizing inclusion criteria, and so on).

## 6 | DISCUSSION

There are multiple reasons for pursuing the collection and synthesis of IPD,<sup>70,71</sup> but personalized medicine is driving many IPD meta-analyses to search for treatment-covariate interactions. To guide such IPD meta-analysis projects, this article has outlined key statistical methods for their conduct and planning, whilst drawing attention to issues and pitfalls. We focused on interactions in randomized trials, but key points also apply to modeling interactions in IPD meta-analyses of study types, such as those evaluating the accuracy of diagnostic tests or the association of prognostic factors to subsequent outcomes.<sup>72</sup>

The actual identification, validation, and successful implementation of treatment-covariate interaction is rare.<sup>2</sup> A sensible starting point—for funders, researchers, health professionals, and patients—is that the relative treatment effect is similar for all individuals, unless there is strong justification otherwise (eg, from previous findings, and especially biological rationale).<sup>11,73</sup> We highlighted many issues and pitfalls that if ignored may produce misleading conclusions, and we suspect many will affect the 102 IPD meta-analyses identified by Schuit et al as providing a significant interaction estimate.<sup>74</sup>

Our work focuses on statistical models for estimating an interaction. However if we are willing to make clinical decisions (and even withhold treatment) based on an individual's predicted treatment effect, then there should be a firm understanding of the causal mechanism and pathway, and the cost-effectiveness of the approach. Sun et al suggest criteria to assess the credibility of a treatment-covariate interaction (which they call a subgroup effect),<sup>12</sup> including an item about the need for evidence to be based on within-study rather than across-study information. A key role is to disentangle different sources of variation,<sup>75</sup> and Senn recommends we need more  $N$ -of-1 trials,<sup>76</sup> to repeatedly test multiple treatments in the same person, including the same treatment multiple times. For further consideration of whether interactions are likely to be genuine, we refer to two excellent tutorials which discuss intricate issues such as dose-response relationships, measurement error, adjusting for confounders, dealing with multiple covariates, and multiplicative vs additive scales.<sup>72,77</sup> Scale of the analysis is an important issue. For example, a treatment-covariate interaction may occur on the odds ratio scale but not the risk ratio scale, when the covariate is also a prognostic factor, as superbly illustrated by Shrier and Pang.<sup>78</sup> Conversely the risk ratio scale may also be problematic in some particular situations as, unlike the odds ratio, its value may be bounded.<sup>45</sup> Nonlinear trends may also exist on one scale but not another. Therefore consideration of multiple

scales may be important, as illustrated in the example of Section 4.3. Allowing for nonproportional hazards may also be important for similar reasons.

More research is needed to help address the issue of multiple testing and exploratory research when investigating multiple interactions in IPD meta-analysis settings, extending recommendations for a single trial.<sup>79</sup> Mistry et al propose a tree-based recursive partitioning algorithm (called “IPD-SIDES”) to identify subgroup effects in an IPD meta-analysis when there are many covariates of interest (ie, in an exploratory analysis).<sup>80</sup> A limitation of their proposal is that it requires the use of cut-points to dichotomize continuous covariates, rather than leaving them as continuous, and—as far as we can tell—amalgamated within-study and across-study information. Extension of their work to address these issues would be welcome.

Treatment-covariate interactions correspond to changes across individuals in the treatment effect as measured on the relative scale (eg, risk ratio, odds ratio, hazard ratio). However, another approach to personalizing or stratifying the use of treatments is to consider their impact on *absolute* risks. Those people with the highest absolute risk will derive the largest absolute benefit from a treatment (eg, greatest reduction in probability of the outcome) when the treatment effect expressed as a risk ratio is the same for all patients. Therefore, even in the absence of treatment-covariate interactions, treatment decisions might be tailored conditional on absolute outcome risk.<sup>1,2</sup> Lastly, we note that those considering IPD meta-analysis projects should also consider other practicalities aside from the analysis, such as obtaining, cleaning and harmonizing the IPD.<sup>17,18</sup>

## ACKNOWLEDGEMENTS

R.D.R. and J.E. were supported by funding to Keele University for an awarded National Institute for Health Research (NIHR) Clinical Trials Unit Support Funding Opportunity entitled “Supporting efficient/innovative delivery of NIHR research”. T.P.A.D., J.H., K.G.M.M., and J.B.R. were supported by a TOP grant of the Netherlands Organisation for Health Research and Development (ZonMw) entitled “Promoting tailored healthcare: improving methods to investigate subgroup effects in treatment response when having multiple individual participant datasets” (Grant number: 91215058). N.M. is funded by an NIHR Doctoral Fellowship (DRF-2018-11-ST2-077). The i-WIP project was funded by NIHR Health Technology Assessment (programme 12/01). This publication presents independent research funded by the NIHR. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

## DATA AVAILABILITY STATEMENT

The IPD used to provide examples is not routinely available for sharing

## ORCID

Richard D. Riley  <https://orcid.org/0000-0001-8699-0735>

Thomas P.A. Debray  <https://orcid.org/0000-0002-1790-2719>

Jeroen Hoogland  <https://orcid.org/0000-0002-2397-6052>

Joie Ensor  <https://orcid.org/0000-0001-7481-0282>

## REFERENCES

1. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ*. 2013;346:e5793.
2. Riley RD, van der Windt D, Croft P, Moons KGM, eds. *Prognosis Research in Healthcare: Concepts, Methods and Impact*. Oxford, UK: Oxford University Press; 2019.
3. Hudis CA. Trastuzumab—mechanism of action and use in clinical practice. *N Engl J Med*. 2007;357(1):39-51.
4. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol*. 2004;57(3):229-236.
5. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? *Lancet*. 2005;365(9456):341-346.
6. Fisher DJ, Copas AJ, Tierney JF, Parmar MK. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. *J Clin Epidemiol*. 2011;64(9):949-967.
7. 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. *Stat Med*. 2017;36(5):772-789.
8. Riley RD, Lambert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med*. 2008;27(11):1870-1893.

9. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? *BMJ*. 2017;356:j573.
10. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med*. 2007;26(15):2982-2999.
11. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JP. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. *JAMA Intern Med*. 2017;177(4):554-560.
12. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 2010;340:c117.
13. Koopman L, van der Heijden GJ, Hoes AW, Grobbee DE, Rovers MM. Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. *Int J Technol Assess Health Care*. 2008;24(3):358-361.
14. Koopman L, van der Heijden GJ, Glasziou PP, Grobbee DE, Rovers MM. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. *J Clin Epidemiol*. 2007;60(10):1002-1009.
15. Royston P, Sauerbrei W. Interactions between treatment and continuous covariates: a step toward individualizing therapy. *J Clin Oncol*. 2008;26(9):1397-1399.
16. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Stat Med*. 2004;23(16):2509-2525.
17. 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.
18. Papadimitropoulou K, Stijnen T, Dekkers OM, le Cessie S. One-stage random effects meta-analysis using linear mixed models for aggregate continuous outcome data. *Res Synth Methods*. 2019;10(3):360-75.
19. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med*. 2011;154(10):680-683.
20. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549.
21. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J Royal Stat Soc Ser A*. 2009;172:137-159.
22. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med*. 2001;20(24):3875-3889.
23. van Aert RCM, Jackson D. A new justification of the Hartung-Knapp method for random-effects meta-analysis based on weighted least squares regression. *Res Synth Methods*. 2019;10(4):515-527.
24. Debray TP, Moons KG, Abo-Zaid GM, Koffijberg H, Riley RD. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLoS One*. 2013;8(4):e60650.
25. Stewart GB, Altman DG, Askie LM, Duley L, Simmonds MC, Stewart LA. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. *PLoS One*. 2012;7(10):e46042.
26. Crowther MJ, Look MP, Riley RD. Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Stat Med*. 2014;33(22):3844-3858.
27. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 2010;29:3046-3067.
28. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med*. 2017;36(5):855-875.
29. Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Res Synth Methods*. 2010;1:2-9.
30. Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Stat Med*. 2001;20(15):2219-2241.
31. Simmonds MC. Statistical methodology for individual patient data meta-analysis [PhD thesis]. University of Cambridge; 2006.
32. Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Stat Med*. 2003;22(16):2591-2602.
33. Fisher DJ. Two-stage individual participant data meta-analysis and generalized forest plots. *Stata J*. 2015;15(2):369-396.
34. Legha A, Riley RD, Ensor J, KIE S, Morris TP, Burke DL. Individual participant data meta-analysis of continuous outcomes: a comparison of approaches for specifying and estimating one-stage models. *Stat Med*. 2018;37(29):4404-4420.
35. Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics*. 1946;2(6):110-114.
36. Belias M, Rovers MM, Reitsma JB, Debray TPA, Int'Hout J. Statistical approaches to identify subgroups in meta-analysis of individual participant data: a simulation study. *BMC Med Res Methodol*. 2019;19(1):183.
37. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219.
38. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst*. 1994;86(11):829-835.
39. Altman DG, Royston P. Statistics notes: the cost of dichotomising continuous variables. *BMJ*. 2006;332:1080.
40. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25(1):127-141.
41. Ensor J, Burke DL, Snell KIE, Hemming K, Riley RD. Simulation-based power calculations for planning a two-stage individual participant data meta-analysis. *BMC Med Res Methodol*. 2018;18(1):41.
42. Senn S. Individual response to treatment: is it a valid assumption? *BMJ*. 2004;329(7472):966-968.

43. Kasenda B, Sauerbrei W, Royston P, et al. Multivariable fractional polynomial interaction to investigate continuous effect modifiers in a meta-analysis on higher versus lower PEEP for patients with ARDS. *BMJ Open*. 2016;6(9):e011148.
44. Kasenda B, Sauerbrei W, Royston P, Briel M. Investigation of continuous effect modifiers in a meta-analysis on higher versus lower PEEP in patients requiring mechanical ventilation—protocol of the ICEM study. *Syst Rev*. 2014;3:46.
45. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. New York, NY: Springer; 2015.
46. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med*. 2012;31:3821-3839.
47. Riley RD, Jackson D, Salanti G, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ*. 2017;358:j3932.
48. Jackson D, Riley RD, White IR. Multivariate meta-analysis: potential and promise. *Stat Med*. 2011;30:2481-2498.
49. White IR. Multivariate random-effects meta-regression: updates to mvmeta. *Stata J*. 2011;11:255-270.
50. White IR. Multivariate meta-analysis. *Stata J*. 2009;9:40-56.
51. White IR, Kaptoge S, Royston P, Sauerbrei W, Emerging Risk Factors Collaboration. Meta-analysis of non-linear exposure-outcome relationships using individual participant data: a comparison of two methods. *Stat Med*. 2019;38(3):326-38.
52. Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Soc Ser A*. 1999;162:71-94.
53. Royston P, Sauerbrei W. *Multivariable Model-Building—A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Chichester, UK: Wiley; 2008.
54. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat*. 1994;43(3):429-467.
55. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. *Stat Med*. 2011;30(28):3341-3360.
56. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer; 2009.
57. Copas JB. Using regression models for prediction: shrinkage and regression to the mean. *Stat Methods Med Res*. 1997;6(2):167-183.
58. Steyerberg EW, Harrell FEJ, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781.
59. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586.
60. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245.
61. van Klaveren D, Balan TA, Steyerberg EW, Kent DM. Models with interactions overestimated heterogeneity of treatment effects and were prone to treatment mistargeting. *J Clin Epidemiol*. 2019;114:72-83.
62. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol*. 2016;45(6):2075-2088.
63. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298(10):1209-1212.
64. Riley RD, Price MJ, Jackson D, et al. Multivariate meta-analysis using individual participant data. *Res Synth Methods*. 2015;6:157-174.
65. Kovalchik SA, Cumberland WG. Using aggregate data to estimate the standard error of a treatment-covariate interaction in an individual patient data meta-analysis. *Biom J*. 2012;54(3):370-384.
66. Arnold BF, Hogan DR, Colford JM Jr, Hubbard AE. Simulation methods to estimate design power: an overview for applied research. *BMC Med Res Methodol*. 2011;11:94.
67. Landau S, Stahl D. Sample size and power calculations for medical studies by simulation when closed form expressions are not available. *Stat Methods Med Res*. 2013;22(3):324-345.
68. Feiveson AH. Power by simulation. *Stata J*. 2002;2(2):107-124.
69. Browne WJ, Lahi MG, Parker RMA. A guide to sample size calculation for random effects models via simulation and the MLPowSim software package. *MLPowSim Manual*. Bristol: University of Bristol; 2009.
70. Tierney JF, Vale C, Riley R, et al. Individual Participant Data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med*. 2015;12(7):e1001855.
71. Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM, Cochrane IPD Meta-analysis Methods group. Individual Participant Data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med*. 2015;12(10):e1001886.
72. Greenland S. Basic problems in interaction assessment. *Environ Health Perspect*. 1993;101(suppl 4):59-66.
73. Malottki K, Biswas M, Deeks JJ, et al. Stratified medicine in European Medicines Agency licensing: a systematic review of predictive biomarkers. *BMJ Open*. 2014;4(1):e004188.
74. Schuit E, Li AH, Ioannidis JPA. How often can meta-analyses of individual-level data individualize treatment? A meta-epidemiologic study. *Int J Epidemiol*. 2019;48(2):596-608.
75. Senn S. Mastering variation: variance components and personalised medicine. *Stat Med*. 2016;35(7):966-977.
76. Senn S. Statistical pitfalls of personalized medicine. *Nature*. 2018;563(7733):619-621.
77. VanderWeele Tyler J, Knol Mirjam J. A tutorial on interaction. *Epidemiol Methods*. 2014;3(1):33.
78. Shrier I, Pang M. Confounding, effect modification, and the odds ratio: common misinterpretations. *J Clin Epidemiol*. 2015;68(4):470-474.

79. Lipkovich I, Dmitrienko A, D'Agostino RB. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med*. 2017;36(1):136-196.
80. Mistry D, Stallard N, Underwood M. A recursive partitioning approach for subgroup identification in individual patient data meta-analysis. *Stat Med*. 2018;37(9):1550-1561.

**How to cite this article:** Riley RD, Debray TP, 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;1–23. <https://doi.org/10.1002/sim.8516>