Received ; | Revised ; | Accepted

DOI: xxx/xxxx

ARTICLE TYPE

An extended mixed-effects framework for meta-analysis

Francesco Sera*[1,2] | Benedict Armstrong[1,2] | Marta Blangiardo[3] | Antonio Gasparrini[1,2]

[1]Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, United Kingdom

[2]Centre for Statistical Methodology, London School of Hygiene & Tropical Medicine, London, United Kingdom

[3]Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom

Correspondence
*Francesco Sera, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH, United Kingdom. Email: francesco.sera@lshtm.ac.uk

Summary

Standard methods for meta-analysis are limited to pooling tasks in which a single effect size is estimated from a set of independent studies. However, this setting can be too restrictive for modern meta-analytical applications. In this contribution, we illustrate a general framework for meta-analysis based on linear mixed-effects models, where potentially complex patterns of effect sizes are modelled through an extended and flexible structure of fixed and random terms. This definition includes, as special cases, a variety of meta-analytical models that have been separately proposed in the literature, such as multivariate, network, multilevel, dose-response, and longitudinal meta-analysis and meta-regression. The availability of a unified framework for meta-analysis, complemented with the implementation in a freely-available and fully documented software, will provide researchers with a flexible tool for addressing non-standard pooling problems.

KEYWORDS:
meta-analysis, mixed-effects models, dose-response, longitudinal

1 | INTRODUCTION

Meta-analysis has become a standard method to summarize evidence in various scientific fields.[1] Traditional applications require a set of single effect size estimates that are collected from multiple independent studies. However, extensions to deal with more complex meta-analytical problems have been presented. These include, potentially among others, multivariate models for pooling multiple outcomes or multi-parameter associations,[2,3] network meta-analysis for indirect mixed-treatment comparison,[4] multilevel versions for hierarchically-structured studies,[5] dose-response meta-analysis,[6,7] and longitudinal meta-analysis for studies reporting multiple estimates at different times.[8] Although these extensions were presented separately, all of them can be described as cases where multiple observations are collected within each study, and their dependence within and/or between studies creates more complex correlation structures that need to be modelled or accounted for.

In this contribution, we review and bring together these different developments into a coherent unified framework, built on the known link between meta-analysis and linear mixed-effects (LME) models, where patterns of effect sizes are modelled through a flexible structure of fixed and random terms.[9,10,11,12,13,14,15,16,17] The manuscript is organized as follows: the analytic formulation of the unified framework is introduced in Section 2, followed by estimation and inferential procedures in Section 3. Specific applications are presented in Section 4, including analytic definitions linked to the general framework, and illustrations through real-data examples. Section 5 describes the software implementation of the modelling framework in the new R package mixmeta, while Section 6 presents the results of a simulation study. Section 7 draws some conclusions. R code and data for replicating

[0]**Abbreviations:** LME, linear mixed-effects; ML, maximum likelihood; REML, restricted maximum likelihood; GLS, generalized least squares; AIC, Akaike information criteria; BIC, Bayesian information criteria; BLUP, best linear unbiased prediction; BCG, Bacillus Calmette-Guerin; TB, tubercolosis; OR, odds ratio; RR, incidence relative rate; (R)IGLS, (restricted) iterated generalized least squares; RMSE, root mean square error

examples and simulation results are added as supplementary material, with an updated version available at the personal website and GitHub page of the last author.

2 | A MIXED-EFFECTS FRAMEWORK FOR META-ANALYSIS

A unified modelling framework can be defined by casting the meta-analytical problem as a LME model. In general terms, we assume that there is a set of n total measures effect sizes (observations) of k different outcomes, representing *units* of analysis aggregated in $i = 1, \dots, m$ *groups* that are considered independent. Additional $L - 1$ inner levels of grouping could exist within each of the m outer groups, for a total of L grouping levels. Grouping levels can be represented by studies themselves, as in standard meta-analysis, or be defined either between or within studies. An extended mixed-effects meta-regression model for the $\mathbf{y}_i$ effect sizes (outcomes) in group i can be generally written as:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \ , \ i = 1, \dots, m \ , \\ \mathbf{b}_i &\sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Psi}_i\right) \ , \boldsymbol{\epsilon}_i \sim \mathrm{N}\left(\mathbf{0}, \mathbf{S}_i\right) \ . \end{aligned} \tag{1}$$

Here, $\mathbf{X}_i\boldsymbol{\beta}$ defines the fixed effects that represent the population-averaged outcomes in terms of p unit-level meta-predictors in the design matrix $\mathbf{X}_i$, with fixed-effects coefficients $\boldsymbol{\beta}$. The random part of the model, $\mathbf{Z}_i\mathbf{b}_i$, describes the deviation from the population averages in terms of q predictors defined at different grouping levels and composing the random-effects design matrix $\mathbf{Z}_i$, with coefficients $\mathbf{b}_i$. The vector $\boldsymbol{\epsilon}_i$ defines the unit-level sampling errors. The model have marginal distribution $\mathbf{y}_i \sim \mathrm{N}\left(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i\right)$, where the marginal (co)variance matrix $\boldsymbol{\Sigma}_i = \mathbf{S}_i + \mathbf{Z}_i\boldsymbol{\Psi}_i\mathbf{Z}_i^{\mathrm{T}}$ is given by the sum of within-group errors (assumed known) and between-group random effects, defined by (co)variance matrices $\mathbf{S}_i$ and $\boldsymbol{\Psi}_i$, respectively. The latter is composed of a block-diagonal form of level-specific matrices $\boldsymbol{\Psi}^{(1)}, \dots, \boldsymbol{\Psi}^{(L)}$ (from outer to inner levels), defined by a set of parameters $\boldsymbol{\xi}$ dependent on their specific form (e.g. unstructured, (heterogeneous) compound symmetry, (heterogeneous) autoregressive of first order)[18] and on constraints for ensuring positive-definiteness. These matrices are expanded consistently with the inner structure of each group, similarly to $\mathbf{Z}_i$ (see Section 4.3 for algebraic details).

3 | ESTIMATION

Likelihood functions

The unknown parameters of the model in Equation 1 are the vector $\boldsymbol{\beta}$ of fixed effects, and the vector $\boldsymbol{\xi}$ that characterises the set of level-specific (co)variance matrices of random effects composing $\boldsymbol{\Psi}_i$. These can be estimated through (restricted) maximum likelihood (ML and REML) estimators, with the marginal (restricted) log-likelihood functions derived from the LME framework[19,20] as:

$$\begin{aligned} l\left(\boldsymbol{\beta}, \boldsymbol{\xi} \mid \mathbf{y}\right) &= -\tfrac{1}{2}n\log(2\pi) - \tfrac{1}{2}\sum_{i=1}^{m}\log\left|\boldsymbol{\Sigma}_i\right| - \tfrac{1}{2}\sum_{i=1}^{m}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \ , \\ l_R\left(\boldsymbol{\xi} \mid \mathbf{y}\right) &= -\tfrac{1}{2}(n-p)\log(2\pi) + \tfrac{1}{2}\log\left|\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\right| - \tfrac{1}{2}\log\left|\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right| \\ &\quad - \tfrac{1}{2}\sum_{i=1}^{m}\log\left|\boldsymbol{\Sigma}_i\right| - \tfrac{1}{2}\sum_{i=1}^{m}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \ . \end{aligned} \tag{2}$$

Note that the REML version l_R only depends on $\boldsymbol{\xi}$, as it is obtained by re-projecting $\mathbf{y}$ using a set of $n - p$ orthogonal error contrasts, a transformation which is reflected algebraically in the inclusion of the two additional determinant terms in Equations 2. The estimators derived from REML are generally considered superior, in particular regarding the estimation of random components, as they account for the loss of degrees of freedom in the estimation of $\boldsymbol{\beta}$ that induces a downward bias in the ML counterpart. However, they pose limitations for hypothesis testing, as discussed below.

For known $\boldsymbol{\xi}$, maximum likelihood estimates of the fixed effects coefficients and their associated (co)variance matrix can be easily obtained by generalized least squares (GLS) estimators, expressed in closed form as:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{m} \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{m} \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i \ , \\ V\left(\hat{\boldsymbol{\beta}}\right) &= \left(\sum_{i=1}^{m} \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \ . \end{aligned} \tag{3}$$

Fixed-effects meta-analytical models can be simply estimated using Equation 3 by setting $\boldsymbol{\Sigma}_i = \mathbf{S}_i$. For random-effects models, when the random part is unknown, the joint estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ requires iterative methods for maximizing the likelihood functions in Equations 2. For computational convenience, a *profiled* approach is preferable, where iterative algorithms are defined in terms of parameters $\boldsymbol{\xi}$ only, and values of $\hat{\boldsymbol{\beta}}$ are obtained by Equation 3 and plugged in at each iteration, until convergence. Alternative algorithms have been proposed, such as Newton-Raphson, expectation-maximization (EM), and (restricted) iterative generalized least squares (IGLS adn RIGLS), each of them with different properties.[18,21,22] See Section 5 for additional details.

Hypothesis testing and model comparison

Inferential procedures follow standard LME theory, and concern the fixed-effects parameter vector $\boldsymbol{\beta}$ and the set of random-effects (co)variance matrices $\boldsymbol{\Psi}^{(\ell)}$, with $\ell = 1, \dots, L$. Regarding the fixed effects, under the marginal model and replacing $\boldsymbol{\Sigma}_i$ in Equation 3 with its ML or REML estimate through $\hat{\boldsymbol{\xi}}$, the vector $\hat{\boldsymbol{\beta}}$ follows a multivariate normal distribution with (co)variance matrix $V(\hat{\boldsymbol{\beta}})$. These results can be used to derive approximated confidence intervals and (multivariate) Wald tests for specific coefficients or their linear combinations. If the Wald test gives significant results, a common question is which particular linear combinations of the coefficients are significantly different from zero. The common example is where we find a difference on the k effect sizes and we wish to perform all possible comparisons. A simultaneous comparisons procedure which maintain the overall type I error were proposed by Goldstein.[23] Comparison between nested models can be performed through likelihood ratio (LR) tests, or more generally using fit statistics such as the Akaike or Bayesian information criteria (AIC and BIC), each of which is easily computed using the (restricted) maximum likelihood values from Equations 2.

LR tests and AIC/BIC can also be used for hypothesis testing and model selection, for instance by comparing alternative structures for random-effects (co)variance matrices $\boldsymbol{\Psi}^{(\ell)}$, or by assessing the presence of heterogeneity at each grouping level ℓ. However, it must be noted that the chi-square distribution is a poor approximation to the actual distribution of the LRT statistic when applied to a large number of parameters, and when testing heterogeneity, with the null hypothesis $\boldsymbol{\Psi}^{(\ell)} = \mathbf{0}$ lying on the boundary of the parameters space. More importantly, the REML log-likelihood function is not invariant to one-to-one re-parametrisation of the fixed effects, as this changes the specification of the error contrasts, and therefore LR tests and AIC/BIC can only be used to compare REML models with the same fixed-effects specification.

In addition to inferential tools borrowed directly from LME models, other statistics traditionally used in meta-analysis to assess the presence and amount of heterogeneity can be easily extended in this more general mixed-effects framework, for instance the Cochran Q and I^2.[24,25] These can be defined as:

$$\begin{aligned} Q &= \sum_{i=1}^{m} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}\right)^{\mathrm{T}} \mathbf{S}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}\right) \\ I^2 &= \max\left\{ \frac{Q - n + p}{Q} \ , \ 0 \right\} \end{aligned} \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ are estimated by the correspondent fixed-effect model with no random term. The Cochran Q statistic follows a χ^2_{n-p} distribution under the hypothesis of no heterogeneity, and can be used to define the related test, while the I^2 statistic quantifies the amount of heterogeneity as the proportion of total variation above that related to sampling error.

Prediction

In this complex meta-analytic setting, inferential procedures can be complemented with prediction tools that inform about potentially complex relationships that are pooled across studies, including for example multivariate and non-linear associations.[3] In this context, predictions offer a method to link specific values of meta-regressors defined at any grouping level with effect size

expectations. Given a set of unit-level meta-predictors $\mathbf{x}_0$ that form the design matrix $\mathbf{X}_0$ depending on the specific model (see Section 4), the (marginal) predicted mean $\hat{\mathbf{y}}_0$ with (co)variance matrix $V(\hat{\mathbf{y}}_0)$ are obtained by:

$$\begin{aligned} \hat{\mathbf{y}}_0 &= \mathbf{X}_0\hat{\boldsymbol{\beta}} \,, \\ V\left(\hat{\mathbf{y}}_0\right) &= \mathbf{X}_0 V\left(\hat{\boldsymbol{\beta}}\right)\mathbf{X}_0^{\mathrm{T}} \,. \end{aligned} \tag{5}$$

In addition to the marginal level, improved study-specific estimates can be obtain as *best linear unbiased predictions* (BLUPs). These are interpreted as trade-off between $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$, with estimates of effect sizes borrowing information within and/or between studies. BLUPs can be defined as conditional expectations given the random effects, and its empirical version $\hat{\mathbf{y}}_{b_i}$ and (co)variance matrix $V\left(\hat{\mathbf{y}}_{b_i}\right)$ are provided as:

$$\begin{aligned} \hat{\mathbf{y}}_{b_i} &= \mathbf{X}_0\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\boldsymbol{\Psi}}_i\mathbf{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_i^{-1}\left(\hat{\mathbf{y}}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right) \,, \\ V\left(\hat{\mathbf{y}}_{b_i}\right) &= \mathbf{X}_0 V(\hat{\boldsymbol{\beta}})\mathbf{X}_0^{\mathrm{T}} + \mathbf{Z}_i\hat{\boldsymbol{\Psi}}_i\mathbf{Z}_i^{\mathrm{T}} - \mathbf{Z}_i\hat{\boldsymbol{\Psi}}_i\mathbf{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_i^{-1}\mathbf{Z}_i\hat{\boldsymbol{\Psi}}_i\mathbf{Z}_i^{\mathrm{T}} \,, \end{aligned} \tag{6}$$

It is interesting to note that, in a multilevel context, BLUPs can be defined also as predictions at higher levels of grouping. For instance, BLUPs at level $\ell \leq L$ are derived by including in $\hat{\boldsymbol{\Psi}}_i$ and $\mathbf{Z}_i$ only the random-effects components corresponding to the grouping levels in ℓ and above (see Section 4.3 for an algebraic definition of levels).

4 | SPECIFIC APPLICATIONS

Different models for meta-analysis can be expressed as special versions of the general framework in Equation 1. These includes the standard methods, extensions mentioned above, and their combinations, among potentially other models. In this section, we describe the most common cases, graphically represented in Figure 1 , highlighting their distinctive aspects and their link with the general framework.

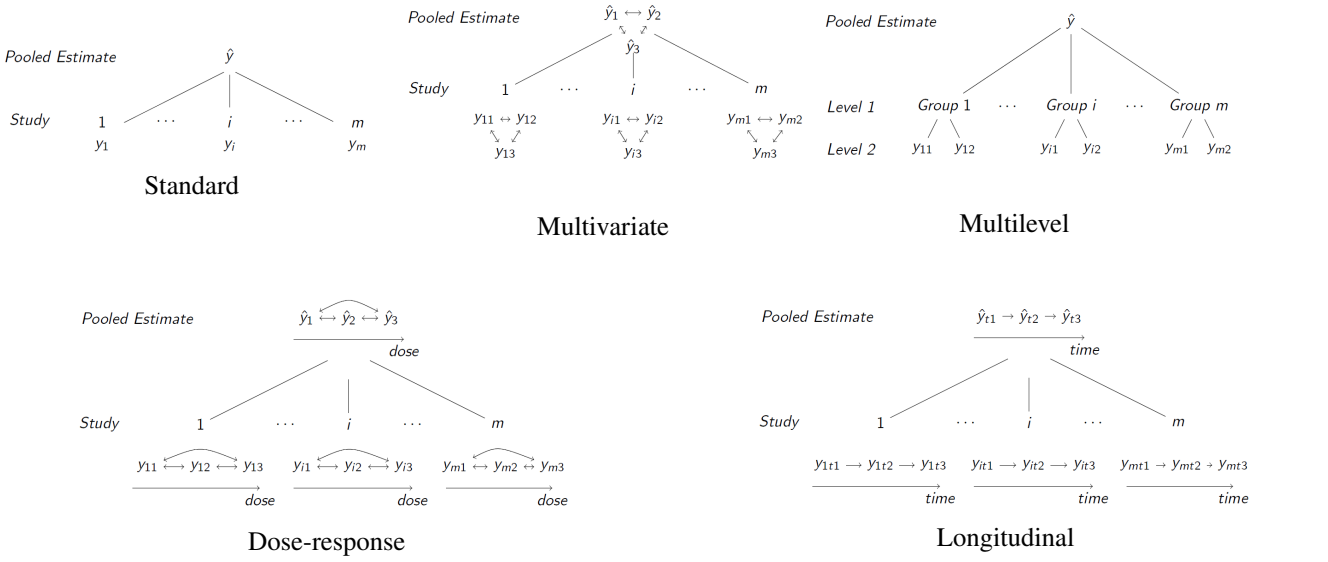

FIGURE 1 Graphical illustration of data structures in specific applications of the extended framework for meta-analysis.

Figure 1 shows how extensions of the standard model are generally characterized by repeated measures and groupings that induce patterns of correlation across effect sizes. As in LME models, these potentially complex structures can be flexibly modelled by a combination of fixed and random terms, optionally including meta-predictors with alternative parameterizations, for instance indicators and continuous smooth functions. In the sub-sections below, we illustrate each case by replicating and extending real-data meta-analyses from published studies, reproduced in the R scripts provided in the Supplementary Material. Details on the data and substantive context can be found in the references or in the help pages of the R package `mixmeta`.

4.1 | Standard meta-analysis

The objective of a *standard meta-analysis* is to obtain a summary (pooled) estimate from single effect sizes estimated separately in independent studies. The basic estimation procedures are based on the computation of weighted averages across studies, with weights proportional to the precision of the estimates.[26] The pooled estimate can be derived under *fixed-effects* or *random-effects* assumptions,[27,28] with random-effects models incorporating the underlying between-study variation into the weights.[24]

Analytic formulation

Several authors have already pointed out that the random-effects meta-analysis can be expressed as a LME model in a regression context.[29,14] Specifically, the standard model for a set of effect sizes y_i can be defined as:

$$\begin{aligned} y_i &= \beta_0 + b_i + \epsilon_i \ , \ i = 1, \dots, m \ , \\ b_i &\sim \mathrm{N}\left(0, \tau^2\right) \ , \ \epsilon_i \sim \mathrm{N}\left(0, s_i^2\right) \ , \end{aligned} \tag{7}$$

where β_0 is the pooled effect, b_i are the study-specific random effects distributed with between-study variance τ^2, and ϵ_i is the error term distributed with known within-study variance s_i^2. This standard model represents the simplest case of the general extended framework in Equation (1), with $n = m$ (a single estimate from separate studies), and scalar quantities $\mathbf{X}_i = \mathbf{Z}_i = 1$, $\boldsymbol{\beta} = \beta_0$, $\mathbf{b}_i = b_i$, $\boldsymbol{\epsilon}_i = \epsilon_i$, $\boldsymbol{\Psi} = \tau^2$, and $\mathbf{S}_i = s_i^2$. In fixed-effects models, the term b_i does not exist. The model in Equation (7) can be extended to *meta-regression* by defining a set of study-level predictors $\mathbf{x}_i = \left[x_{i1}, \dots, x_{ip}\right]^{\mathrm{T}}$ and setting $\mathbf{X}_i = \mathbf{x}_i^T$, where usually $x_{i1} = 1$ specifies the intercept term.

Illustrative example

In this first example we consider a meta-analysis and meta-regression performed by Colditz and colleagues that evaluate the efficacy of the Bacillus Calmette-Guerin (BCG) vaccine for preventing tuberculosis (TB).[30] The dataset was used by several authors to illustrate their random-effects regression models.[29,15,17] The data refers to 13 prospective clinical trials that estimated the odds ratio (OR) of TB between groups vaccinated with the (BCG) vaccine and non-vaccinated control populations.

We apply the general framework to estimate the parameters for the log-OR β_0 and between-study variance τ^2 in Equation (7) using a maximum likelihood estimator (see Section 3), replicating the results reported by Van Houwelingen and colleagues.[17] Consistently, the estimated OR is 0.476 (95%CI: 0.336 to 0.675), with a clear indication of a protective effect of BCG vaccine, and the estimated τ^2 is 0.302, with suggestions of a large heterogeneity ($I^2 = 92.6\%$). Similarly to the original analysis, we can investigate the influence on vaccine efficacy of various meta-predictors such as study location, year of publication and method of treatment allocation. For instance, adding latitude in a meta-regression model reduces the between-study variance and residual heterogeneity ($\tau^2 = 0.004$, and $I^2 = 56.2\%$). The coefficient for latitude is -0.033 (95%CI: -0.039 to -0.026), indicating an improved efficacy of the vaccine at higher latitudes.

4.2 | Multivariate meta-analysis

An important extension of the standard univariate model in Equation (7) is *multivariate meta-analysis*, in which each study still reports single estimates, but for multiple effect sizes referring to different outcomes, such as disease free and overall survival risks in cancer patients.[31,2] The same model has been extended to other contexts, for instance to pool results from multi-parameter functions defining non-linear relationships,[3] or meta-analysis of diagnostic accuracy tests.[32] A common application of multivariate models is for network meta-analysis applied in mixed-treatment comparisons, where efficiency can be gained by exploiting the correlation among effect sizes that measure relative effects across different treatments.[33,4]

Analytic formulation

In all these applications, the k-dimensional vector $\mathbf{y}_i$ contains estimates for multiple (potentially correlated) effect sizes in each study. A model for random-effect multivariate meta-analysis can be represented as follows:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{b}_i + \boldsymbol{\epsilon}_i \ , \ i = 1, \dots, m \ , \\ \mathbf{b}_i &\sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Psi}\right) \ , \boldsymbol{\epsilon}_i \sim \mathrm{N}\left(\mathbf{0}, \mathbf{S}_i\right) \ . \end{aligned} \tag{8}$$

This can be written in terms of the general framework in Equation (1) by setting $\mathbf{X}_i = \mathbf{Z}_i = \mathbf{I}_k$, $\boldsymbol{\beta} = \left[\beta_1, \dots, \beta_k\right]^{\mathrm{T}}$, $\mathbf{b}_i = \left[b_{i1}, \dots, b_{ik}\right]^{\mathrm{T}}$, and $k \times k$ between-study $\boldsymbol{\Psi}$ and within-study $\mathbf{S}_i$ (co)variance matrices. Here each of the m studies represents a group with multiple estimates, with a total of $n = k \cdot m$ units in the balanced (full-outcome) case. Missing outcomes (unbalanced

case) can be accounted for by excluding related rows in the matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$. Similarly to the univariate case in sub-Section 7, the model can extended to *multivariate meta-regression* by setting $\mathbf{X}_i = \mathbf{x}_i^T \otimes \boldsymbol{I}_k$, with $\otimes$ as the Kronecker product operator and $\boldsymbol{\beta}$ being the kp-dimensional coefficient vector that defines the association of the k outcomes with the p predictors. This notation defines the same set of meta-predictors for all outcomes, but it allows different associations for each of them. Alternative parameterizations can be used to define outcome-specific sets of meta-predictors, or to impose the same effects across outcomes.

Illustrative example

As an example of multivariate models, we consider an application of network meta-analysis on 24 trials that compare four alternative interventions to promote smoking cessation, labelled here A–D (see Table 1).[34] Each trial compares only two or three interventions, and the joint meta-analysis allows to gain information through indirect comparisons. Treatment A is used here as the reference, and trials without an arm A were augmented with 0.01 individuals and 0.001 successes. Here the $\mathbf{y}_i$ and $\mathbf{S}_i$ represent the log-OR of cessation and associated (co)variance error matrix of treatments B, C, and D versus A estimated in each trial, including missing values. Below we replicate results previously presented in the article by White.[35]

The first model is formulated under the assumption of consistency, that is allowing heterogeneity between studies but with no systematic variation across trial designs (defined by groups of trials reporting the same comparisons).[36] This model can be fitted using the general framework in Equation 1 expressed as Equation 8, with $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \beta_3\right]^{\mathrm{T}}$ representing the three comparisons of treatments B, C and D versus A. Following White,[35] we impose a parsimonious structure to the random-effects (co)variance matrix $\boldsymbol{\Psi}$, assuming the same variance τ^2 for all the comparisons and fixing their correlation to 0.5. The results of the consistency models are reported in Table 1 : treatments C and D are effective with respect treatment A, with substantial heterogeneity among studies ($\tau^2 = 0.454$ and $I^2 = 86.3\%$). The consistency assumption that direct and indirect evidence agree with each other can be relaxed by defining and testing design-by-treatment interactions.[33,36] This *inconsistency* model has 10 fixed-effects coefficients compared to the three of the simpler version, and a global Wald test with a p-value of 0.646 fails to reject the consistency assumption.

TABLE 1 Example of multivariate (network) meta-analysis of 24 trials comparing alternative treatments of smoking cessation, using a consistency model and a structured between-study (co)variance matrix. Previously reported by White.[35]

Comparison	Estimated log-OR	Standard Error	p-value
Self-help versus no contact (B vs A)	0.398	0.330	0.227
Individual counselling versus no contact (C vs A)	0.702	0.196	0.000
Group counselling versus no contact (D vs A)	0.866	0.373	0.020

4.3 | Multilevel meta-analysis

The previous models work under the assumption that studies independently provide single estimates of one or multiple outcomes. This setting can be too simplistic for some applications of meta-analysis. For example, some studies can report multiple estimates of the same effect size, either at different stages or for separate groups. Similarly, studies can exhibit nested levels of hierarchy, with higher grouping factors being represented for instance by geographical areas, administrative units, or study characteristics.[5] This configuration of repeated measures and/or hierarchical structures creates a potentially complex pattern of dependence across effect sizes that must be accounted for. *Multilevel meta-analysis* has been proposed, in different forms, to extend the model defined in Section 4.1 by modelling the dependence through structured random effects.[5,9,10,11,13]

Analytic formulation

Here we provide a general definition of multilevel random-effect meta-analysis that is applicable in various settings. The pattern of correlation is defined by aggregating effect sizes in *groups*, which can be defined both between and within studies. Nested

grouping levels are used to express a hierarchical structure of random effects. For the sake of clarity, we start from a model for n effect sizes (units) aggregated in two nested grouping levels, written as:

$$\begin{gathered} y_{ijr} = \beta_0 + b_i + b_{ij} + \epsilon_{ijr} \ , \ i = 1, \dots, m \ , \ j = 1, \dots, m_i \ , \ r = 1, \dots, n_{ij} \ , \\ b_i \sim \mathrm{N}\left(0, \tau_1^2\right) \ , \ b_{ij} \sim \mathrm{N}\left(0, \tau_2^2\right) \ , \ \epsilon_{ijr} \sim \mathrm{N}\left(0, s_{ijr}^2\right) \ . \end{gathered} \tag{9}$$

Here τ_1^2 is the variance of the random effects at the outer grouping level i, which includes m independent groups. In contrast, τ_2^2 is the random-effects variance within each of the m_i inner level groups nested in each outer-level group i. The units indexed by r represent measured effect sizes from the n_{ij} studies in the inner group j nested within the outer group i, each with known within-study variance s_{ijr}^2.

The model in Equation (9) can be extended to include an indefinite number L of grouping levels, with $\ell = 1, \dots, L$, and generally written in terms of the unified framework described in Section 2. First, we define $j = 1, \dots, g_i^\ell$ as the number of groups at level ℓ within each outer level i, with $g_i^1 = 1$ by definition. Each group includes $r = 1, \dots, n_{ij}^\ell$ units, with $\sum_{j\ell} n_{ij}^\ell = n_i$. The definition of the various elements in Equation 1 requires block-diagonal expansions and column binding consistent with repeated measures and grouping levels, respectively, with $\bigoplus_v a_v$ representing an operator that creates a block-diagonal matrix of elements a_v. We first define the design matrix for the fixed effects as $\mathbf{X} = \mathbf{1}_{n_i}$, and the known error structure at the outermost level as $\boldsymbol{\epsilon}_i = \bigoplus_{rj} \epsilon_{ijr}$. The random-effects part can be written by first defining design matrices for each group at various levels of random effects as $\mathbf{Z}_{ij}^\ell = \mathbf{1}_{n_{ij}^\ell}$, then expanding them at each level as $\mathbf{Z}_i^\ell = \bigoplus_j \mathbf{Z}_{ij}^\ell$, and finally binding them as $\mathbf{Z}_i = \left[\mathbf{Z}_i^1, \dots, \mathbf{Z}_i^L\right]$. Consistently, the between-group (co)variance matrix is defined as $\boldsymbol{\Psi}_i = \bigoplus_\ell \mathbf{I}_{g_i^\ell} \otimes \tau_\ell^2$. The model can be extended further to meta-regression by replacing $\mathbf{X}_i$ with a $n_i \times p$ design matrix including p fixed-effects predictors. Similarly, q_ℓ random-effects predictors at any level ℓ can be included by replacing $\mathbf{Z}_{ij}^\ell$ with a $n_{ij}^\ell \times q_\ell$ design matrix, and τ_ℓ^2 with a random-effects (co)variance matrix $\boldsymbol{\Psi}^\ell$.

Illustrative examples

In a first example, we consider a meta-analysis of 56 studies that evaluate the effect of a modified school calendar on standardized reading achievement.[37] The studies were performed in 11 separate school districts, with at least three studies in each district, therefore providing a classic example of multilevel structure. Using the notation in Equation 9, in this example, the outer grouping level i are the school districts, which define $m = 11$ independent groups. Within each school district, a variable number of studies were performed, e.g. four studies were performed in the first school district, that is $m_1 = 4$. The study j, nested within the school district i, is the inner level in the multilevel structure with one single observation $r = 1$, with a single effect size in each inner group j nested within the outer group i. We fitted three models with different random-effects structures using a maximum likelihood estimator: first, a traditional meta-analysis using the model in Section 4.1, with a single level of random effects assigned to each study; second, a single-level meta-analysis with random effects by district, therefore including repeated measures within each group; third, a full two-level meta-analysis with nested random effects by study and district. The results, partly replicating the analysis of Konstantopoulos,[13] are reported in Table 2 . The comparison makes clear the advantage of recognising the multilevel structure of the data, with the pooled effect size β_0 increasing from 0.128 in the standard model to 0.184 in the two-level model. The latter, in addition, shows a better fit, as suggested by the lower AIC, and indicates the presence of heterogeneity at both district and study levels, with $\tau_1^2 = 0.058$ and $\tau_2^2 = 0.033$, respectively.

TABLE 2 Example of multilevel meta-analysis of 56 studies that evaluate changes in standardized reading achievement after the implementation of a modified school calendar, with studies clustered within school districts. Previously reported by Konstantopoulos.[13]

Model	Grouping levels	Pooled estimate (Std error)	Random-effects variances		
		β_0	τ_1^2 (district)	τ_2^2 (study)	AIC
One-level (standard)	Study	0.128 (0.043)	-	0.087	37.292
One-level (repeated measures)	District	0.196 (0.086)	0.075	-	69.432
Two-level	Study within district	0.184 (0.080)	0.058	0.033	22.790

A second example of multilevel meta-analysis considers 20 randomized trials of thrombolytic therapy, which evaluated short-term mortality risks after a myocardial infarction.[38] The hypothesis is that the thrombolytic therapy reduces the risk, and that the benefit is particularly substantial for very early treatment. Some of the trials report separate estimates of absolute risk change for sub-groups of treatment times, leading to a multilevel structure with 38 (potentially repeated) observations within 20 trials. We applied alternative models fitted by REML, partly reproducing the analysis by Thompson and colleagues.[10] Specifically: a standard meta-analysis that ignores clustering by trial; a two-level meta-analysis; a two-level meta-regression that includes treatment delay as a meta-predictor. The results, shown in Table 3 , show that both standard and two-levels random-effects meta-analyses produce an estimate of absolute risk difference of -0.02600, suggesting a protective effects of thrombosis treatment, and that the second model indicates presence of heterogeneity within but not between the higher level of grouping represented by trials ($\tau_2^2 = 0.00747$ and $\tau_1^2 < 0.0000$). However, the inclusion of treatment delay in a meta-regression explains most of the variability at the inner level ($\tau_2^2 = 0.00006$), with a residual heterogeneity between trials of $\tau_1^2 = 0.00216$. These models (and other specifications) can be compared through AIC (when defined using the same fixed-effects structure) or using Wald tests for meta-predictors.

TABLE 3 Example of multilevel meta-analysis of 20 randomized trials of thrombolytic therapy for myocardial infarction, with multiple estimates of absolute risk change at different times of treatment. Previously reported by Thompson and colleagues.[10]

Model	Fixed effects (Std error)		Random-effects variances	
	β_0 (intercept)	β_1 (treatment delay in hours)	τ_1^2 (trials)	τ_2^2 (times)
Standard meta-analysis	-0.02600 (0.00314)	-	-	0.00747
Two-level meta-analysis	-0.02600 (0.00314)	-	<0.00001	0.00747
Two-level meta-regression	-0.03494 (0.00421)	0.00161 (0.00049)	0.00216	0.00006

4.4 | Dose-response meta-analysis

Dose-response meta-analysis has been used to summarise linear and non-linear health associations across epidemiological studies.[39] The standard approach consists of a two-stage procedure. In the first stage, study-specific associations are determined using a set of parameters that represent estimates at different doses, usually retrieved from published data and relying on various methods to approximate their (co)variance matrix accounting for within-study correlations.[6] These estimates are then pooled in the second stage using standard meta-analytical models (see Section 4.1) for linear dose-response relationships, or multivariate methods (see Section 4.2) for multiple parameters of functions representing non-linear associations.[40,41] Recently, Crippa and colleagues have proposed a one-stage approach for dose-response meta-analysis that provides important advantages and allows defining dose-response meta-analysis within the general framework proposed in Equation 1.[7]

Analytic formulation

The one-stage model for a linear dose-response random-effects meta-analysis can be written as follow:

$$\begin{aligned} y_{ij} &= \beta x_{ij} + b_i x_{ij} + \epsilon_{ij} \ , \ i = 1, \dots, m \ j = 1, \dots, n_i \ , \\ b_i &\sim \mathrm{N}\left(0, \tau^2\right) \ , \ \left[\epsilon_{i1}, \dots, \epsilon_{in_i}\right] \sim \mathrm{N}\left(\mathbf{0}, \mathbf{S}_i\right) \ . \end{aligned} \tag{10}$$

Here, for each study i, the n_i units y_{ij} represent the association measures (*e.g.*, log odds ratios or log risk ratios) at different doses x_{ij}, commonly retrieved by using published estimates. The fixed and random-effects parameters β and b_i represent the pooled linear dose-response association and its study specific deviations, respectively. Note the absence of an intercept for this model. Similarly to the standard model in Equation 7, τ^2 represents the variance of the random effects. Following Crippa and colleagues,[7] this model can be written as a special version of the general framework in Section 2 by setting $\mathbf{X}_i = \mathbf{Z}_i =$

$\left[x_{i1}, \ldots, x_{ij}\right]^{\mathrm{T}}$, $\boldsymbol{\beta} = \beta$, $\mathbf{b}_i = b_i$, and $\boldsymbol{\Psi} = \tau^2$. The within-study error structure is represented by a $n_i \times n_i$ matrix $\mathbf{S}_i$, usually approximated using alternative methods.[6,42] This model can be easily extended to the pooling of non-linear dose-responses by applying functions to transform x_{ij}, for instance quadratic terms or splines, thus obtaining a $n_i \times q$ matrices $\mathbf{X}_i$ and/or $\mathbf{Z}_i$, and a $q \times q$ matrix $\boldsymbol{\Psi}$.

Illustrative example

As an example of dose-response meta-analysis, we consider the data on eight cohort studies participating in the Pooling Project of Prospective Studies of Diet and Cancer.[43] Each study estimated the incidence relative rate (RR) of colorectal cancer in various categories of alcohol intake while controlling for a set of potential confounders, using non-drinkers as the reference. The categories were then converted in a dose by assigning the median value of individual consumptions, reporting log-RR estimates at multiple levels in a continuous scale. We fitted alternative models using a maximum likelihood estimator, exploiting the flexibility of the extended framework in defining fixed effects and within and between-study correlations. Specifically, we specified linear and non-linear terms in both fixed and random parts, the latter by using natural cubic splines with internal knots at approximately the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of alcohol consumption. Within-study correlations were optionally reconstructed using the method of Greenland and Longnecker.[39]

The models are presented in Table 4 , including number of parameters and AIC, and partly replicate and extend results presented by Crippa, Orsini and colleagues.[6,42] Consistently, findings show that accounting for within-study correlation significantly improves the fit of the model, as indicated by the lower AIC of model L2 *vs* L1. The RR corresponding to 12 grams/day of alcohol intake in the two models changes to 1.080 (95%CI: 1.047–1.115) from 1.048 (1.016–1.080), respectively. The inclusion of non-linear terms in both fixed and random parts does not improve the fit (NL1 *vs* L2). However, the simplification to linear random effects in NL2, allowed by the flexibility of the unified framework, indicates evidence of non-linearity (NL2 *vs* L2). The Cochran Q test for models NL1-NL2 suggests little evidence of heterogeneity (p-value = 0.25, not shown), as confirmed by the better fit of the fixed-effects non-linear dose-response meta-analysis in model NL3. The predicted RR for different doses obtained through linear and non-linear meta-analytic models L2 and NL1 are represented in Figure2 , with similar shapes to graphs previously presented.[6,42]

TABLE 4 Example of dose-response meta-analysis of eight cohort studies on alcohol and colorectal cancer, with alternative model specifications. Previously partly reported by Crippa, Orsini and colleagues.[6,42]

Model	Within-study correlations	Fixed effects	Random effects	Degrees of freedom	AIC
L1	Zero correlation	Linear	Linear	2	-2.06
L2	Greenland and Longnecker	Linear	Linear	2	-6.13
NL1	Greenland and Longnecker	Non Linear	Non Linear	9	1.28
NL2	Greenland and Longnecker	Non Linear	Linear	4	-7.88
NL3	Greenland and Longnecker	Non Linear	None	3	-9.88

4.5 | Longitudinal meta-analysis

Another example of recent extensions of meta-analytical methods is for applications with studies where the same outcome is measured at several time points. *Longitudinal meta-analysis* have been proposed in this context to account for the intrinsic within and between-study correlations.[8,44] The common procedure is to apply meta-analytical methods for multivariate meta-analysis (see Section 4.2), treating effect sizes estimated at different times as separate outcomes,[8,45,46] although this representation poses important constraints, as explained below. Alternative approaches define the longitudinal design as a special case of the multilevel setting described in Section 4.3, with repeated measures within each study.[47,44] This provides a way to formulate longitudinal meta-analysis within the unified framework in Equation 1, offering a general, flexible, and efficient modelling structure.

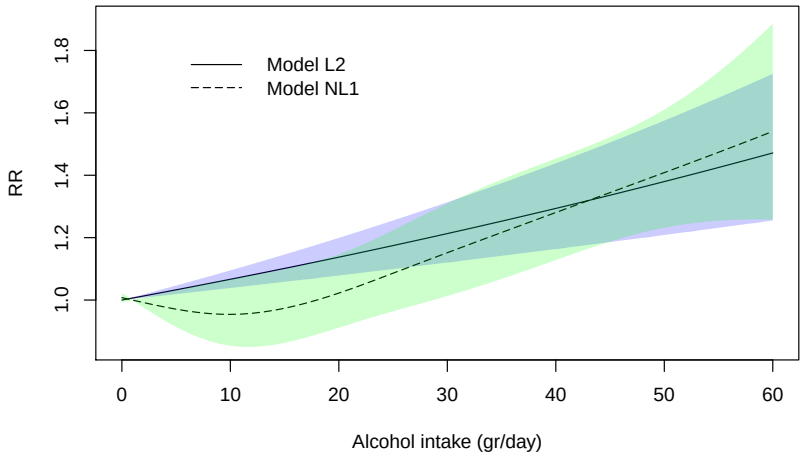

FIGURE 2 Dose-response relationships between alcohol intake and incidence relative rates of colorectal cancer assuming a linear and non-linear association (Models L2 and NL1 in Table 4), with 95% confidence intervals.

Analytic formulation

As mentioned above, traditional methods defines longitudinal meta-analysis as a multivariate model, where the effect sizes $\mathbf{y}_i = (y_{i,t_1}, \dots, y_{i,t_k})^{\mathrm{T}}$ measured at k times in study i are treated as separate outcomes, and modelled as in Equation 8. However, this approach requires that the measurements are taken at common time points across studies, and while it may account for the sequential order, for instance by imposing autoregressive structures to the within and/or between-study correlations for evenly-spaced measures, it ignores most of the information provided by the longitudinal setting. A more flexible definition for a set of effect sizes measures at n_i times in study i is derived directly from LME models as:

$$\begin{gathered} y_{it} = \left(\alpha + a_i\right) + \left(\beta + b_i\right) t_{ij} + \epsilon_{it} \ , \ i = 1, \dots, m \ , \ j = 1, \dots, n_i \ , \\ \left[a_i, b_i\right] \sim \mathrm{N}\left(\mathbf{0}, \mathbf{\Psi}\right) \ , \ \left[\epsilon_{it_1}, \dots, \epsilon_{it_{n_i}}\right] \sim \mathrm{N}\left(\mathbf{0}, \mathbf{S}_i\right) \ , \end{gathered} \tag{11}$$

with α, β, a_i, and b_i as fixed and random coefficients for intercepts and slopes. This formulation treats time as a continuous predictor that can be modelled through both fixed and random terms, and allows studies to report estimates at different times. The traditional multivariate approach can be defined as a special case by using indicators for a common set of time points. The model in Equation 11 can be written as the general framework in Equation 1 by setting $\mathbf{t}_i = \left[t_{i1}, \dots, t_{in_i}\right]^{\mathrm{T}}$, and $\mathbf{X}_i = \mathbf{Z}_i = \left[\mathbf{1}_{n_i}, \mathbf{t}_i\right]$. $\mathbf{\Psi}$ and $\mathbf{S}_i$ define the random-effects and within-study error (co)variance matrices, respectively, optionally with specific structures, such as diagonal or (continuous) auto-regressive of fist order (AR_1). The model can allow non-linear trends by specifying smooth functions of time (see Section 4.4), or include additional meta-predictors, in both cases either as fixed or random effects by extending $\mathbf{X}_i$ and $\mathbf{Z}_i$, respectively.

Example of longitudinal meta-analysis

We illustrate an application of longitudinal meta-analysis using data on 17 randomized controlled trials comparing treatments of malignant gliomas. Each study measured the survival OR at 6, 12, 18, and 24 months since the start of the treatment.[48] Musekiwa and colleagues have previously analysed the data using multivariate models fitted with REML,[45] defining various longitudinal meta-analytic models with different specification of the within and between-study (co)variance structures. Here we replicate and extend the results using the more flexible general framework and adopting alternative specifications.

The first set of results using multivariate models with indicators for the four time points are reported in Table 5 . Consistently with the original analysis,[45] the first three options (Models 1-3) do not allow correlations in the within-study errors, while the other options (Models 4-6) assume a heterogeneous AR_1 structure with correlation fixed at 0.61. Different structures were used for the random-effects (co)variance, leading to different total degrees of freedom. The best-fitting option in terms of AIC is Model 4, with independent random effects and AR_1 within-study errors. The analysis can be extended by defining time as a

TABLE 5 Example of longitudinal meta-analysis of 17 randomized controlled trials comparing treatments of malignant gliomas, reporting survival odds ratio (OR) at multiple times after treatment. Previously partly reported by Musekiwa and colleagues.[45]

Model	(Co)variance structures		Degrees of freedom	AIC
	Within-study errors	Random effects		
Model 1	Diagonal	Diagonal	8	121.6
Model 2	Diagonal	Compound symmetry	5	117.0
Model 3	Diagonal	Heterogeneous AR_1	9	120.9
Model 4	Heterogeneous AR_1	Diagonal	8	107.5
Model 5	Heterogeneous AR_1	Heterogeneous AR_1	9	107.7
Model 6	Heterogeneous AR_1	Unstructured	14	117.3

continuous variable, specifying an additional Model 7 as in Equation 11. This random-slope model specifies a diagonal structure for intercept and (centered) time as random effects, and keeps the AR_1 within-study errors, using only four degrees of freedom. Models 7 and 4 were (re)fitted using maximum likelihood, that allows comparison between different fixed-effects specifications (in this case, linear and through indicators, respectively). The results are graphically illustrated in Figure 3 , showing the pooled OR along time after treatment. AIC indicates a better fit of Model 7 (101.4 *vs* 107.5, respectively), suggesting a linear trend and actually little evidence of changes in survival along time, with a p-value of 0.92 for the coefficient β of time (not shown). This example highlights the advantages offered by the modelling flexibility of the general modelling framework.

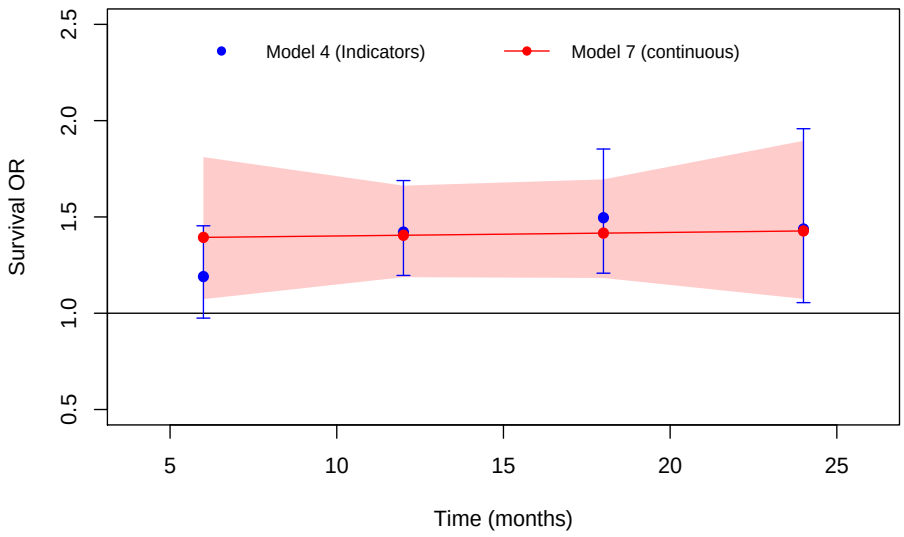

FIGURE 3 Survival odds ratio after start of the treatment of gliomas (Models 4 and 7 fitted using maximum likelihood), with 95% confidence intervals.

5 | SOFTWARE

The unified random-effect framework for meta-analysis and the frequentist inferential procedure described in the previous sections are implemented in the R package `mixmeta`. The main function of the program is `mixmeta()`, which uses a simple syntax to fit a wide range of meta-analytical models. For instance, the following code:

```
mixmeta(cbind(y1 + y2) ~ x1 + x2, S, data, random = list(~ z1 | g1, ~ 1 | g2), method="reml")
```

performs a bivariate two-level meta-regression using a REML estimator. In this example, effect sizes for two outcomes `y1` and `y2`, with unit-level errors `S`, are modelled in terms of fixed-effect predictors `x1` and `x2`. Random effects are specified by intercept plus variable `z1` and intercept only, for nested grouping levels `g1` (outer) and `g2` (inner), respectively. The flexible formula syntax, similar to that applied in the R package `nlme` for LME models, allows the definition of all the various versions illustrated in Section 4. Other functions are available for hypothesis testing, predictions, model assessment, and simulations, among other regression tasks.

At the time of writing, the package implements a hybrid estimation procedure, with few runs of a (R)IGLS algorithm followed by quasi-Newton iterations. The former is robust to initial values and quickly moves close to the maximum likelihood, while the latter provides a fast convergence within this region. As mentioned in Section 3, these algorithms adopt a profiled approach, where the likelihood functions in Equations 2 are defined in terms of random-effect parameters $\boldsymbol{\xi}$ only, with a parameterization that ensures positive-definiteness and allows different structures for any of the (co)variance matrices $\boldsymbol{\Psi}^{(\ell)}$. Computationally, the estimation algorithms exploit the block-diagonal form of the design and (co)variance matrices defined in Equations 1–6, which is particularly convenient in the presence of a high number of studies or outer groups. A QR decomposition is applied internally in the GLS routine, providing numerical stability even in not well-conditioned least squares problems.

6 | SIMULATIONS

We performed a simulation study to explore the validity and inferential properties of the software implementation of the unified framework. We considered a complex case represented by a multivariate multilevel meta-analysis, combining the applications described in Sections 4.2 and 4.3 within the general model defined in Equation 1. Specifically, we simulated $k = 3$ outcomes and $L = 2$ grouping levels, with m groups at the (outer) level 1, each including $g_i^2 = 10$ groups at the (inner) level 2. All the fixed-effects $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \beta_3\right]^{\mathrm{T}}$, representing the three pooled intercepts, were simulated as 0. We assumed a compound-symmetry structure for both the 3×3 random-effects (co)variance matrices $\boldsymbol{\Psi}^1$ and $\boldsymbol{\Psi}^2$, and a heterogeneous compound-symmetry structure for the residual error matrix $\mathbf{S}_i$. Random-effects variances τ_1^2 and τ_2^2 were set to 1, while the residual error variances s_{ijr}^2 were sampled from a uniform distribution with range $[0.1, 2]$. Various simulation scenarios are represented by combinations of number of outer-level groups m (10 or 50), correlation ρ_{b1} and ρ_{b2} for each level of random effects (0 or 0.8), and residual correlation ρ_w (0 or 0.8). For each combination, we simulated 10,000 datasets using the function `mixmetaSim()`, and fitted the general model with `mixmeta()` assuming the correct random-effects (co)variance structures.

Results are reported in Table 6 , showing the bias, root mean square error (RMSE), and coverage for the (first) fixed-effect coefficient, and the bias for the four random-effects parameters. Simulations indicate a negligible amount of bias for both the fixed and random-effects parameters in all scenarios. As expected, the RMSE decreases when increasing number of outer-level units. The coverage is slightly below the nominal value, especially for scenarios with lower number ($m = 10$) of outer-level groups. Inferential properties do not seem affected by the presence of within or between units correlation.

7 | DISCUSSION

In this contribution, we have presented an extended mixed-effects framework that provides a common modelling and inferential setting for meta-analysis. It includes traditional applications, but also non-standard extensions for which common meta-analytical methods are not appropriate. The unified approach proposed here generally characterises these extensions as patterns of dependence between effect sizes, modelled through fixed and random effects defined by meta-predictors and grouping structures. This modelling approach allows a flexible specification of variety of meta-analytical models and facilitates the design and implementation of non-standard pooling studies.

The LME structure adopted in the definition of the general model in Equation 1 provides substantial modelling flexibility, through which important constraints in design and modelling aspects can be relaxed. For instance, analyses of longitudinal data are traditionally performed using models for multivariate meta-analysis that consider repeated measurements from the same study as multiple outcomes.[8,6] However, this approach requires a limited set of measurements to be taken at the same doses/times

TABLE 6 Simulation study: multivariate multilevel meta-analysis with $k = 3$ outcomes and $L = 2$ grouping levels, with m groups at the (outer) level 1, each including $g_i^2 = 10$ groups at the (inner) level 2. Eight simulation scenarios are defined by the number of outer-level groups m and correlations ρ_{b1} and ρ_{b2} for each level of random effects.

Parameters			β_1			τ_1	τ_2	ρ_{b1}	ρ_{b2}
m	ρ_w	ρ_b	Bias	RMSE	Coverage	Bias	Bias	Bias	Bias
10	0.00	0.00	0.000	0.349	0.940	-0.006	0.004	0.004	-0.004
50	0.00	0.00	-0.001	0.157	0.948	-0.001	0.001	0.002	-0.001
10	0.80	0.00	-0.005	0.351	0.930	-0.005	0.005	-0.045	-0.003
50	0.80	0.00	0.001	0.154	0.951	-0.000	0.001	-0.007	0.000
10	0.00	0.80	0.003	0.357	0.934	-0.007	0.004	-0.019	-0.030
50	0.00	0.80	-0.004	0.155	0.949	0.001	0.000	-0.002	-0.007
10	0.80	0.80	-0.006	0.347	0.928	-0.006	-0.001	-0.058	-0.015
50	0.80	0.80	0.001	0.153	0.951	0.000	0.000	-0.008	-0.003

across studies, and prevents their analysis as continuous variables. In the examples in Sections 4.4 and 4.5, we showed how more flexible models can be defined within our general framework, allowing studies to provide an indefinite number of measurements taken at any point and the modelling of continuously dose-response shapes and trends through linear or smooth functions. Similarly, the flexible definition of multilevel models in Section 4.3 allows the specification of complex hierarchical structures and the inclusion of random-effects meta-predictors.

The extended framework presented in this paper is well placed for a two-stage analytical setting, where the estimated effect sizes are derived from published studies or previously obtained from separate study-specific analyses. One-stage formulations have been proposed for individual patient data (IPD) meta-analysis, when data from the original studies are available and can be directly modelled.[49] However, in many applications the one-stage approach provides little advantages, and two-stage procedures offer a valid, computationally stable, and efficient alternative.[9,50] In addition, the flexible framework proposed here allows extensions of the two-stage design to address specific limitations, for example with the pooling of multiple study-specific parameters of main and interaction terms to evaluates effect modification from participant-level variables. The two-stage approach relies on the assumption of normal distribution of estimated effect sizes and random effects, thus requiring approximations, in particular for outcomes measured in a binary scale. One-stage methods based on generalized linear mixed models (GLMM) have been developed in this setting, including versions with alternative distributional assumptions.[11,51,52] While these have theoretical and inferential advantages, they present considerable computational problems, and simulations show improvements only in the presence of small and sparse data.[53,54] An additional requirement of the two-stage procedure, when applied in multi-parameter meta-analyses, is the knowledge of the within-study covariances. Methods for estimating them from published data were developed in multivariate meta-analysis,[5,31,55] and in dose-response meta-analysis, and can be applied in this general model.[56,57,58] In addition, interestingly, the unknown correlations can alternatively be merged in a marginal random-effects structure that includes within and between-study dependencies.[16,59] One of the advantage of using an LME formulation in the extended framework is that it does not require balanced data where the full set of effect sizes are measured (or reported) for each study. The extended framework can in fact deal with unbalanced data and more generally deal with the presence of missing effect sizes. However, the analysis requires the assumption of missing at random (MAR) to provide unbiased estimates.[60]

The methodology is implemented in the freely-available and fully documented R package `mixmeta`, which complements standard software for meta-analysis and additional tools for specific extensions. For instance, some analysts have proposed the use of general LME programs for fitting complex meta-analytical models, such as the procedure `PROC MIXED` in SAS,[17] the program `GLLAMM` in Stata,[15] MLwiN,[11] or the package `nlme` in R.[61] However, the use of general LME software requires advanced knowledge of statistical and computational aspects, and can be difficult for more applied users. Dedicated routines are available for specific meta-analytical extensions, such as the Stata command and R package `mvmeta` for multivariate meta-analysis,[35,3] or the R package `dosresmeta` and `drmeta` Stata module for dose-response meta-analysis,[42,62] while `metafor` in R can offer a set of general tools for standard models and various extensions.[63] Our implementation in `mixmeta` offers a flexible platform where the full range of models presented in Section 4, and their combinations, can be defined through a simple syntax,

fitted using an efficient computational structure, and estimated following a common underlying statistical theory. This software can complement existing packages and modules for the specific meta-analytic extensions presented in Section 4.

The simulation study in Section 6 demonstrates the validity and good performance of the modelling framework and software, even in a relatively complex application represented by a trivariate multilevel meta-analysis. However, some limitations of the inferential approaches described in Section 3 must be acknowledged. The Wald test procedure for fixed effects is based on asymptotic distributional approximations, and it ignores the uncertainty related to the estimation of the random-effects components. This explains the small undercoverage of confidence intervals in Table 6 , which however can be non-negligible in small-sample studies. Similarly, hypotheses on random effects are evaluated through LR tests and AIC/BIC, but these can have poor performances and problems with boundary conditions. Solutions can be found in the LME models literature, such as the use of t or F distributions,[18] adjustments for standard errors and degrees of freedom,[64] and use of mixture distributions.[23,18] Some of these have also been defined for meta-analytical models,[65,29,27,66,67,68] but still need to be fully developed for this extended framework. Alternative methods can also be developed in a Bayesian framework, which offers advantages in accounting for various sources of uncertainty, although requiring appropriate parameterizations and priors specification.[69,27,10,70,71]

There is an increasing interest in developing meta-analysis for applications in more complex pooling studies, beyond the now established extensions described in Section 4.[4,72] Emerging areas include investigations that apply two-stage designs for the analysis of large datasets, where either the complexity of the first-stage regression or the computational demand prevent the definition of a one-stage model, and the partition of the analysis in two steps provides a feasible and efficient approach.[3,73,50] However, the limitations of traditional meta-analytical methods, requiring the estimation of single independent parameters from each subset, poses important constraints in this setting. In contrast, the model in Equation 1 offers flexibility in the definition of the two-stage analysis, allowing for instance repeated measurements in time or sub-groups, hierarchies and spatial or temporal clustering, and complex multi-parameter effect estimates. The definition of a unified framework for meta-analysis, complemented with a full software implementation, provides researchers with a flexible tool for defining and applying flexible meta-analytical models in a variety of pooling problems.

ACKNOWLEDGMENTS

This research was supported by funding from the Medical Research Council UK (Grants ID: MR/M022625/1 and MR/R013349/1).

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

SUPPORTING INFORMATION

The following supporting information is available as part of the online article: R code and data for replicating examples and simulation results in Sections 4 and 6, with an updated version available at the personal website and GitHub page of the last author.

References

1. Borenstein M, Hedges LV, Higgins J, Rothstein HR. *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons; 2009.

2. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med.* 2011;30(20):2481–2498.

3. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med.* 2012;31(29):3821–3839.

4. 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.

5. Stevens JR, Taylor AM. Hierarchical dependence in meta-analysis. *J Educ Behav Stat.* 2009;34(1):46–73.

6. Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. *Am J Epidemiol.* 2011;175(1):66–73.

7. Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. One-stage dose–response meta-analysis for aggregated data. *Stat Methods Med Res.* 2019;28(5):1579–1596.

8. Ishak K, Platt RW, Joseph L, Hanley JA, Caro JJ. Meta-analysis of longitudinal studies. *Clin Trials.* 2007;4(5):525–539.

9. Goldstein H, Yang M, Omar R, Turner R, Thompson S. Meta-analysis using multilevel models with an application to the study of class size effects. *J R Stat Soc Ser C Appl Stat.* 2000;49(3):399–412.

10. Thompson SG, Turner RM, Warn DE. Multilevel models for meta-analysis, and their application to absolute risk differences. *Stat Methods Med Res.* 2001;10(6):375–392.

11. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med.* 2000;19(24):3417–3432.

12. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Stat Med.* 1998;17(22):2537–2550.

13. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. *Res Synth Methods.* 2011;2(1):61–76.

14. Stram DO. Meta-analysis of published data using a linear mixed-effects model. *Biometrics.* 1996;52(2):536–544.

15. Bagos PG. Meta-analysis in Stata using gllamm. *Res Synth Methods.* 2015;6(4):310–332.

16. Noortgate W, López-López José A, Marín-Martínez F, Sánchez-Meca J. Three-level meta-analysis of dependent effect sizes. *Behav Res Methods.* 2013;45(2):576–594.

17. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med.* 2002;21(4):589–624.

18. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. New York, NY; Springer Verlag; 2000.

19. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc.* 1977;72(358):320–338.

20. Verbeke G, Molenberghs G.. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer; 1997.

21. Goldstein H. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika.* 1986;73(1):43–56.

22. Goldstein H. Restricted unbiased iterative generalized least-squares estimation. *Biometrika.* 1989;76(3):622–623.

23. Goldstein H. *Multilevel Statistical Models*. Hoboken, NJ: John Wiley & Sons; 2011.

24. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7(3):177–188.

25. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21(11):1539–1558.

26. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ.* 2011;342:d549.

27. Higgins J, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc.* 2009;172(1):137–159.

28. Rice K, Higgins J, Lumley T. A re-evaluation of fixed effect (s) meta-analysis. *J R Stat Soc Ser A Stat Soc.* 2018;181(1):205–227.

29. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Stat Med.* 1995;14(4):395–411.

30. Colditz GA, Brewer TF, Berkey CS, et al. Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. *JAMA.* 1994;271(9):698–702.

31. Olkin I, Gleser L. *Stochastically dependent effect sizes*. New York, NY: Russell Sage Foundation; 2009.

32. Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. *Stat Methods Med Res.* 2016;25(4):1596–1619.

33. White IR, Barrett JK, Jackson D, Higgins J. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Methods.* 2012;3(2):111–125.

34. Fiore MC, Bailey W C, Cohen S.J, et al. *Smoking Cessation. Clinical Practice Guideline No. 18.* AHCPR Publication No. 96-0692: Agency for Health Care Policy and Research, U.S. Department of Health and Human Services; 1996.

35. White IR. Multivariate random-effects meta-regression: updates to mvmeta. *Stata J.* 2011;11(2):255.

36. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods.* 2012;3(2):98–110.

37. Cooper H, V Jeffrey C, Charlton K, Melson A. The effects of modified school calendars on student achievement and on school and community attitudes. *Rev Educ Res.* 2003;73(1):1–52.

38. Boersma E, Maas ACP, Deckers JW, Simoons ML. Early thrombolytic treatment in acute myocardial infarction: reappraisal of the golden hour. *Lancet.* 1996;348(9030):771–775.

39. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology.* 1993;4(3):218–228.

40. Liu Q, Cook NR, Bergström A, Hsieh CC. A two-stage hierarchical regression model for meta-analysis of epidemiologic nonlinear dose–response data. *Comput Stat Data Anal.* 2009;53(12):4157–4167.

41. Rota M, Bellocco R, Scotti L, et al. Random-effects meta-regression models for studying nonlinear dose–response relationship, with an application to alcohol and esophageal squamous cell carcinoma. *Stat Med.* 2010;29(26):2679–2687.

42. Crippa A, Orsini N. Multivariate dose-response meta-analysis: The dosresmeta R package. *J Stat Softw.* 2016;72(1):1–15.

43. Cho E, Smith-Warner SA, Ritz J, et al. Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. *Ann Intern Med.* 2004;140(8):603–613.

44. Peters JL, Mengersen KL. Meta-analysis of repeated measures study designs. *J Eval Clin Pract.* 2008;14(5):941–950.

45. Musekiwa A, Manda SOM, Mwambi HG, Chen DG. Meta-analysis of effect sizes reported at multiple time points using general linear mixed model. *PloS One.* 2016;11(10):e0164898.

46. Trikalinos TA, Olkin I. Meta-analysis of effect sizes reported at multiple time points: a multivariate approach. *Clin Trials.* 2012;9(5):610–620.

47. Ahn JE, French JL. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *J Pharmacokinet Pharmacodyn.* 2010;37(2):179–201.

48. Fine HA, Dear KBG, Loeffler JS, Mc Black PL, Canellos GP. Meta-analysis of radiation therapy with and without adjuvant chemotherapy for malignant gliomas in adults. *Cancer.* 1993;71(8):2585–2597.

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

50. 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.

51. Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. *Stat Med.* 1999;18(6):643–654.

52. 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(29):3046–3067.

53. Jackson D, Law M, Stijnen T, Viechtbauer W, White IR. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Stat Med.* 2018;37(7):1059–1085.

54. Bakbergenuly I, Kulinskaya E. Meta-analysis of binary outcomes via generalized linear mixed models: a simulation study. *BMC Med Res Methodol.* 2018;18(1):70.

55. Wei Y, Higgins J. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Stat Med.* 2013;32(7):1191–1205.

56. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol.* 1992;135(11):1301–1309.

57. Orsini N, Bellocco R, Greenland S. Generalized least squares for trend estimation of summarized dose–response data. *Stata J.* 2006;6(1):40–57.

58. Hamling J, Lee P, Weitkunat R, Ambuhl M. Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category. *Stat Med.* 2008;27(7):954–970.

59. Noortgate W, López-López José A, Marín-Martínez F, Sánchez-Meca J. Meta-analysis of multiple outcomes: a multilevel approach. *Behav Res Methods.* 2015;47(4):1274–1294.

60. Molenberghs G, Kenward M. *Missing data in clinical studies*. Hoboken, NJ: John Wiley & Sons; 2007.

61. Heisterkamp SH, Willigen E, Diderichsen PM, Maringwa J. Update of the nlme package to allow a fixed standard deviation of the residual error. *R Journal.* 2017;9(1):239–251.

62. Orsini N. *DRMETA: Stata module for dose-response meta-analysis.* : ; 2018.

63. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2010;36(3):1–48.

64. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics.* 1997;:983–997.

65. Kalaian HA, Raudenbush SW. A multivariate mixed linear model for meta-analysis. *Psychol Methods.* 1996;1(3):227.

66. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med.* 2007;26(1):78–97.

67. Follmann DA, Proschan M A. Valid inference in random effects meta-analysis. *Biometrics.* 1999;55(3):732–737.

68. Morris TP, Fisher DJ, Kenward MG, Carpenter JR. Meta-analysis of Gaussian individual patient data: two-stage or not two-stage?. *Stat Med.* 2018;.

69. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res.* 2001;10(4):277–303.

70. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Stat Med.* 2003;22(14):2309–2333.

71. Wei Y, Higgins J. Bayesian multivariate meta-analysis with multiple outcomes. *Stat Med.* 2013;32(17):2911–2934.

72. Sutton AJ, Higgins J. Recent developments in meta-analysis. *Stat Med.* 2008;27(5):625–650.

73. Rhodes KM, Turner RM, Payne RA, White IR. Computationally efficient methods for fitting mixed models to electronic health records data. *Stat Med.* 2018;37(29):4557–4570.

ORCID

Francesco Sera: 0000-0002-8890-6848
Ben Armstrong: 0000-0003-4407-0409
Marta Blangiardo: 0000-0002-1621-704X
Antonio Gasparrini: 0000-0002-2271-3568

How to cite this article: Sera F., Armstrong B. and Gasparrini A. (2019), An extended mixed-effects framework for meta-analysis, *Statistics in Medicine*, *2019;00:1–6*.