Received: 23 August 2021 | Revised: 4 March 2022 | Accepted: 7 April 2022

DOI: 10.1002/jrsm.1562

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Location-scale models for meta-analysis

Wolfgang Viechtbauer[1] | **José Antonio López-López**[2]

[1]Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, The Netherlands

[2]Department of Basic Psychology and Methodology, University of Murcia, Murcia, Spain

Correspondence
Wolfgang Viechtbauer, Department of Psychiatry and Neuropsychology, Vijverdalseweg 1, 6226 NB, Maastricht, The Netherlands.
Email: wolfgang.viechtbauer@maastrichtuniversity.nl

Abstract

Heterogeneity is commonplace in meta-analysis. When heterogeneity is found, researchers often aim to identify predictors that account for at least part of such heterogeneity by using mixed-effects meta-regression models. Another potentially relevant goal is to focus on the amount of heterogeneity as a function of one or more predictors, but this cannot be examined with standard random- and mixed-effects models, which assume a constant (i.e., homoscedastic) value for the heterogeneity variance component across studies. In this paper, we describe a location-scale model for meta-analysis as an extension of the standard random- and mixed-effects models that not only allows an examination of whether predictors are related to the size of the outcomes (i.e., their location), but also the amount of heterogeneity (i.e., their scale). We present estimation methods for such a location-scale model through maximum and restricted maximum likelihood approaches, as well as methods for inference and suggestions for visualization. We also provide an implementation via the *metafor* package for R that makes this model readily available to researchers. Location-scale models can provide a useful tool to researchers interested in heterogeneity in meta-analysis, with the potential to enhance the scope of research questions in the field of evidence synthesis.

KEYWORDS

heterogeneity, location-scale model, meta-regression, mixed-effects models

Highlights

- The observed effects or outcomes to be combined in a meta-analysis are often more variable than would be expected based on their sampling variability alone. This suggests that the underlying true effects or outcomes are heterogeneous.
- Via appropriate meta-regression models, one can examine whether the size of the effects or outcomes tends to be larger under certain circumstances.
- Standard meta-analytic models assume that the amount of heterogeneity is constant across circumstances. In the present paper, we describe a location-scale model for meta-analysis that allows researchers to examine not only

whether the size of the effects or outcomes varies across circumstances, but also the amount of heterogeneity.
- The model allows applied researchers to address entirely new research questions (e.g., for what types of studies are treatment effects more consistent?).

1 | INTRODUCTION

When a phenomenon of interest (e.g., the effectiveness of a treatment, the size of a group difference, or the association between two variables) has been examined across multiple studies, a meta-analysis can be conducted to synthesize the various findings. For this, we quantify the relevant results of each study in terms of an outcome or effect size measure (e.g., as raw/standardized mean differences, log risk/odds ratios, or raw or Fisher's r-to-z transformed correlation coefficients), so that the resulting observed outcomes or effects provide commensurable evidence about the phenomenon of interest and then apply appropriate statistical techniques to analyze these values.[1]

Due to sampling variability, the observed outcomes will differ across studies even if they estimate a common underlying parameter. However, in many cases, the observed outcomes are more variable than would be expected based on their sampling variability alone. This is typically interpreted as evidence for the presence of variability in the underlying true outcomes, a phenomenon commonly referred to as "heterogeneity".[2] Random-effects models are then often used to estimate the amount of heterogeneity in the true outcomes, which is then incorporated into the analysis when estimating the average true outcome.[3,4]

When heterogeneity is found, one can also try to examine if some predictor variables (also known as moderators or effect modifiers) are able to account for at least part of the heterogeneity in the outcomes. A subgroup meta-analysis can be used for this purpose by stratifying the observed outcomes according to a factor of interest.[5] However, mixed-effects meta-regression models provide a more flexible approach, as they allow researchers to examine multiple predictors, both continuous and categorical, within a single modeling framework.[6,7]

In a standard mixed-effects meta-regression model, one or more predictors are included in the model and their association with the size of the outcomes is examined. The "residual heterogeneity" (i.e., the heterogeneity not accounted for by the predictors) is assumed to be homoscedastic. However, this assumption may be violated in practice. Moreover, the amount of heterogeneity might actually vary systematically as a function of one or more predictors (which may be a different set of predictors than those related to the size of the outcomes) and this is something that cannot be examined by means of standard meta-regression models.

Regression models that allow for the error variance to depend on predictor variables have been studied extensively in the past.[8,9] These so-called "location-scale models" are also increasingly popular in the field of multilevel modeling,[10,11,12] and a tutorial for their implementation in R and SAS in this context is available.[13] In the meta-analytic context, location-scale models are seldom used, although their merits have been discussed and illustrated before using the sample sizes of the studies as a predictor for the amount of heterogeneity,[14] and a Bayesian variant for categorical scale moderators with regularized parameters has also been proposed.[15]

As an illustration, consider a meta-analysis on the effectiveness of a psychological intervention that can be delivered either in groups or in an individual format. A common research question is whether both delivery formats yield similarly effective results. This question is related to the "location" part of the model (i.e., the outcome magnitude or size of the average effect) and is routinely examined using meta-regression models. However, we could also raise the question whether both delivery formats lead to equally consistent results. For instance, individual therapy might achieve relatively similar (i.e., homogeneous) results across studies, regardless of the types of patients included or other contextual factors that might vary across studies. On the other hand, group therapy might be very effective for certain types of patients and circumstances, but less so for others, which would imply more heterogeneous findings for studies examining this delivery format. The latter question is related to the "scale" part of the model (i.e., outcome variability), which cannot formally be examined with standard meta-regression models. Both outcome magnitude and outcome variability constitute relevant information for decision making, and this warrants the implementation of location-scale models in meta-analysis.

The purpose of the present paper is to describe the extension of the standard mixed-effects meta-regression model to a location-scale model and to illustrate the use of such a model with several examples with different types of predictor variables. The methods described are also

implemented in the R package *metafor*[16] and we provide the data and code to replicate the illustrative analyses.

The structure of the paper is as follows. In the next section, we present the model. We then provide some technical details about the estimation procedures and methods for making inferences about the parameters in the context of such models. We then present an illustrative example and conclude the paper with a general discussion.

2 | META-ANALYTIC MODELS

Below, we briefly review the standard random- and mixed-effects models and then describe their extension in terms of the location-scale model.

2.1 | Standard Random- and Mixed-Effects Models

For a set of $i = 1,...,k$ independent studies, let y_i denote the observed value of the outcome measure of interest in the ith study. The standard random-effects model in meta-analysis is given by

$$y_i = \mu + u_i + e_i, \tag{1}$$

where μ denotes the average true outcome in the population of studies, $u_i \sim N(0,\tau^2)$ is a normally distributed random effect that allows for heterogeneity in the underlying true outcomes (with τ^2 denoting the between-study variance), and $e_i \sim N(0,v_i)$ is the normally distributed sampling error of the ith estimate. The sampling (or within-study) variances (i.e., v_i values) are assumed to be known constants.*

The random-effects model is actually a special case of the more general mixed-effects meta-regression model given by

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + u_i + e_i, \tag{2}$$

where $x_{i1},...,x_{ip}$ are the values of p moderator variables that may be related to the size of the average true outcome as specified by the model, $\beta_1,...,\beta_p$ are the model coefficients that indicate how the average true outcome changes for a one-unit increase in the corresponding moderator variable, and β_0 is the model intercept. Assumptions about u_i and e_i are the same as before, except that τ^2 now denotes the amount of residual heterogeneity, that is, variability in the true outcomes not accounted for by the moderator(s) included in the model.

2.2 | Location-Scale Model

Equation (2) defines a model that describes the relationship between one or multiple moderators and the size of the outcomes. Hence, in this model, moderators are assumed to be related to the "location" of the outcomes. Accordingly, we will refer to $x_{i1},...,x_{ip}$ as "location variables" and to $\beta_0,...,\beta_p$ as the corresponding "location coefficients". However, there may also be a relationship between the moderators and the amount of heterogeneity in the outcomes. Hence, the "scale" (i.e., variance) of the outcomes may also be function of one or multiple moderator variables. Accordingly, we will refer to the latter as "scale variables".

The standard random- and mixed-effects models do not allow for this possibility, since they assume that the amount of (residual) heterogeneity is constant (i.e., homoscedastic) across studies. This assumption can be relaxed by letting τ^2 be a function of one or more scale variables. In particular, let

$$\tau_i^2 = \alpha_0 + \alpha_1 z_{i1} + ... + \alpha_q z_{iq}, \tag{3}$$

where $z_{i1},...,z_{iq}$ are the values of q scale variables that may be related to the amount of heterogeneity and $\alpha_1,...,\alpha_q$ are the corresponding "scale coefficients", with α_0 again denoting the intercept.

A problem with (3) is the possibility that τ_i^2 can be negative for certain combinations of values for the scale variables and scale coefficients. To enforce that the variance cannot become negative for any of the studies, we can use a model with a log link, so that

$$\ln(\tau_i^2) = \alpha_0 + \alpha_1 z_{i1} + ... + \alpha_q z_{iq}. \tag{4}$$

Then τ_i^2 is given by $\exp(\alpha_0 + \alpha_1 z_{i1} + \cdots + \alpha_q z_{iq})$, which is guaranteed to be positive (or possibly indistinguishable from zero if the values of the scale variables and scale coefficients lead to a very negative value of $\ln(\tau_i^2)$ for a particular study). Note that the standard random- and mixed-effects models are just special cases of the location-scale model where the scale part of the model only includes an intercept term, so that $\tau^2 = \alpha_0$ or $\tau^2 = \exp(\alpha_0)$, depending on the link function used.

2.3 | Maximum Likelihood Estimation

The log-likelihood for the location-scale model with a log link is given by

$$ll(\boldsymbol{\beta},\boldsymbol{\alpha}) = -\frac{k}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{M}| - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X\beta}), \tag{5}$$

where $\boldsymbol{M} = \boldsymbol{V} + \text{diag}(\exp(\boldsymbol{Z\alpha}))$, $\boldsymbol{W} = \boldsymbol{M}^{-1}$, $\boldsymbol{y}$ is the $k \times 1$ vector with the observed outcomes, $\boldsymbol{V}$ is the $k \times k$ diagonal matrix with the sampling variances, $\boldsymbol{X}$ is the $k \times (p+1)$ model matrix containing the location variables (with the first column equal to a vector of 1's for the intercept), $\boldsymbol{\beta}$ is the corresponding $(p+1) \times 1$ vector with the location coefficients, $\boldsymbol{Z}$ is the $k \times (q+1)$ model matrix with the scale variables (again with the first column equal to a vector of 1's), $\boldsymbol{\alpha}$ is the $(q+1) \times 1$ vector with the scale coefficients, and diag() is a function that turns a vector into a diagonal matrix. By setting $\boldsymbol{X}$ to only a column of 1's, we obtain a random-effects model, but with heteroscedastic between-study variances. By setting $\boldsymbol{Z}$ to only a column of 1's, we obtain the standard random/mixed-effects model (with homoscedastic between-study variance) as a special case. Note that (5) is the straightforward generalization of the log-likelihood function for the standard random- and mixed-effects meta-regression models,[17,18] where the only change is that $\boldsymbol{M}$ is no longer diagonal with elements $v_i + \tau^2$, but with elements $v_i + \tau_i^2$.

Maximum likelihood estimates (MLEs) of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can be obtained by maximizing (5) simultaneously over the $p+q+2$ location and scale coefficients. The optimization problem can be simplified by noting that

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y} \tag{6}$$

is the MLE of $\boldsymbol{\beta}$ for a given vector of $\boldsymbol{\alpha}$. Hence, after substituting $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ in (5) and some algebraic simplification, we can construct the profile log-likelihood

$$ll_P(\boldsymbol{\alpha}) = -\frac{k}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{M}| - \frac{1}{2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}, \tag{7}$$

where

$$\boldsymbol{P} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}. \tag{8}$$

Now (7) only depends on $\boldsymbol{\alpha}$ (through $\boldsymbol{M}$, $\boldsymbol{W}$, and hence $\boldsymbol{P}$), which reduces the optimization problem to one involving only the $q+1$ scale coefficients. Quasi-Newton or Nelder–Mead type algorithms can be used for this purpose,[19] which avoids the need to compute the Hessian or information matrix, as would be needed for the Newton–Raphson or Fisher scoring algorithms.

The approach described above yields the MLE of $\boldsymbol{\alpha}$ for model (4), which we denote as $\widehat{\boldsymbol{\alpha}}$. Once $\widehat{\boldsymbol{\alpha}}$ has been obtained, we can compute the MLE of $\boldsymbol{\beta}$ with (6), with $\boldsymbol{W} = \boldsymbol{M}^{-1}$ as before where $\boldsymbol{M} = \boldsymbol{V} + \text{diag}(\exp(\boldsymbol{Z}\widehat{\boldsymbol{\alpha}}))$.

The same approach can also be used to obtain the MLE of $\boldsymbol{\alpha}$ for model (3) that uses an identity link by letting $\boldsymbol{M} = \boldsymbol{V} + \text{diag}(\boldsymbol{Z\alpha})$. However, then extra steps must be taken when optimizing (7) over $\boldsymbol{\alpha}$ to ensure non-negativity for all of the $\boldsymbol{Z\alpha}$ values.† Linearly constrained optimization algorithms can be used for this purpose.[19] Here, the feasible region for $\boldsymbol{\alpha}$ is that set of values for which $\boldsymbol{Z\alpha} \geq \boldsymbol{0}$. Again, once $\widehat{\boldsymbol{\alpha}}$ has been obtained, we can compute the MLE of $\boldsymbol{\beta}$ with (6), where $\boldsymbol{W} = \boldsymbol{M}^{-1}$ and now $\boldsymbol{M} = \boldsymbol{V} + \text{diag}(\boldsymbol{Z}\widehat{\boldsymbol{\alpha}})$.

2.4 | Restricted Maximum Likelihood Estimation

MLEs of variance components are known to be negatively biased, while restricted maximum likelihood (REML) estimation yields approximately unbiased estimates.[20,21,22,23] The same has been found for the ML and REML estimators of τ^2 in the standard random-effects model.[17] Accordingly, it may also be preferable to use REML estimation to estimate the scale coefficients for the location-scale model. The restricted log-likelihood is given by

$$ll_R(\boldsymbol{\alpha}) = -\frac{k-p-1}{2}\ln(2\pi) + \frac{1}{2}\ln|\boldsymbol{X}'\boldsymbol{X}| - \frac{1}{2}\ln|\boldsymbol{M}| - \frac{1}{2}\ln|\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}| - \frac{1}{2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}, \tag{9}$$

with all elements as defined previously. Note that (9) depends on $\boldsymbol{\alpha}$ through $\boldsymbol{M}$, $\boldsymbol{W}$, and $\boldsymbol{P}$, but no longer involves the location coefficients. However, once ll_R has been maximized over the the $q+1$ scale coefficients (either for the log or identity link model), we can again obtain estimates of the elements in $\boldsymbol{\beta}$ with (6).

2.5 | Inference

Once the ML or REML estimates $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ have been obtained, making statistical inferences about the location and scale parts of the model is typically the next step in the analysis. Wald-type methods and methods based on the likelihood ratio of nested models can be used for this purpose and are described below.

2.5.1 | Inference about the Location Part of the Model

The variance–covariance matrix of the elements in $\widehat{\boldsymbol{\beta}}$ can be estimated with

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}. \tag{10}$$

Hence, we can conduct a Wald-type test of the null hypothesis $H_0: \beta_j = 0$ (with $j = 0, ..., p$) by computing

$$z = \frac{\widehat{\beta}_j}{\mathrm{SE}\left[\widehat{\beta}_j\right]}, \tag{11}$$

where $\mathrm{SE}\left[\widehat{\beta}_j\right] = \sqrt{\mathrm{Var}\left[\widehat{\beta}_j\right]}$ and $\mathrm{Var}\left[\widehat{\beta}_j\right]$ is the corresponding (i.e., jth $+ 1$) diagonal element of (10). Under H_0, (11) follows asymptotically a standard normal distribution based on which we can compute the p-value for the test. An approximate 95% confidence interval (CI) for β_j can also be obtained with $\widehat{\beta}_j \pm 1.96 \times \mathrm{SE}\left[\widehat{\beta}_j\right]$.

Multiple coefficients in $\widehat{\boldsymbol{\beta}}$ can be tested simultaneously by computing

$$Q_\beta = \widehat{\boldsymbol{\beta}}_{[2]}' \left(\mathrm{Var}\left[\widehat{\boldsymbol{\beta}}\right]_{[2]}\right)^{-1} \widehat{\boldsymbol{\beta}}_{[2]}, \tag{12}$$

where $\widehat{\boldsymbol{\beta}}_{[2]}$ includes the set of location coefficients to be tested and $\mathrm{Var}\left[\widehat{\boldsymbol{\beta}}\right]_{[2]}$ contains the corresponding rows and columns from (10). Under $H_0: \boldsymbol{\beta}_{[2]} = \mathbf{0}$, (12) follows asymptotically a chi-square distribution with degrees of freedom equal to the number of coefficients tested. A common application of (12) is to test all location coefficients except for the model intercept $\left(\text{i.e., } \widehat{\boldsymbol{\beta}}_{[2]} = \left[\widehat{\beta}_1, ..., \widehat{\beta}_p\right]'\right)$, yielding an omnibus test of the location part of the model.

The predicted average outcome for a particular combination of values for the location variables can be computed with $\widehat{y}_h = \boldsymbol{x}_h \widehat{\boldsymbol{\beta}}$, where $\boldsymbol{x}_h$ is either a particular row from $\boldsymbol{X}$ (yielding the fitted value, $\widehat{y}_i$, for the corresponding study) or contains some other combination of values for the location variables. An approximate 95% CI for a predicted/fitted value is then given by $\widehat{y}_h \pm 1.96\sqrt{\boldsymbol{x}_h \mathrm{Var}\left[\widehat{\boldsymbol{\beta}}\right] \boldsymbol{x}_h'}$.

2.5.2 | Inference about the Scale Part of the Model

Estimating the variance–covariance matrix of the elements in $\widehat{\boldsymbol{\alpha}}$ is not as straightforward. Although computationally more demanding, we can make use of numerical differentiation[24] to approximate the matrix of second derivatives (i.e., the Hessian) of $ll_P(\boldsymbol{\alpha})$ or $ll_R(\boldsymbol{\alpha})$ with respect to the elements in $\boldsymbol{\alpha}$. Once the Hessian is obtained, the inverse of the negative Hessian matrix yields the estimated variance–covariance matrix of the scale coefficients, which we denote by $\mathrm{Var}[\widehat{\boldsymbol{\alpha}}]$. Hence, a Wald-type test of the null hypothesis $H_0: \alpha_j = 0$ (with $j = 0, ..., q$) can be conducted by computing

$$z = \frac{\widehat{\alpha}_j}{\mathrm{SE}\left[\widehat{\alpha}_j\right]}, \tag{13}$$

where $\mathrm{SE}\left[\widehat{\alpha}_j\right]$ is the square-root of the corresponding (i.e., jth $+ 1$) diagonal element of $\mathrm{Var}[\widehat{\boldsymbol{\alpha}}]$. Under H_0, (13) again follows asymptotically a standard normal distribution. As before, we can also construct an approximate 95% CI for α_j with $\widehat{\alpha}_j \pm 1.96 \times \mathrm{SE}\left[\widehat{\alpha}_j\right]$.

Similarly, multiple scale coefficients in $\widehat{\boldsymbol{\alpha}}$ can be tested simultaneously by computing

$$Q_\alpha = \widehat{\boldsymbol{\alpha}}_{[2]}' \left(\mathrm{Var}[\widehat{\boldsymbol{\alpha}}]_{[2]}\right)^{-1} \widehat{\boldsymbol{\alpha}}_{[2]}, \tag{14}$$

where $\widehat{\boldsymbol{\alpha}}_{[2]}$ and $\mathrm{Var}[\widehat{\boldsymbol{\alpha}}]_{[2]}$ again include the rows (and columns) corresponding to the coefficients to be tested. Under $H_0: \boldsymbol{\alpha}_{[2]} = \mathbf{0}$, (14) follows asymptotically a chi-square distribution with degrees of freedom equal to the number of coefficients tested. By including all scale coefficients except for the intercept in this test $\left(\text{i.e., } \widehat{\boldsymbol{\alpha}}_{[2]} = \left[\widehat{\alpha}_1, ..., \widehat{\alpha}_q\right]'\right)$, we can conduct an omnibus test of the scale part of the model.

The predicted amount of (residual) heterogeneity for a particular combination of values for the scale variables can be computed with $\widehat{\tau}_h^2 = \exp(\boldsymbol{z}_h \widehat{\boldsymbol{\alpha}})$ or $\widehat{\tau}_h^2 = \boldsymbol{z}_h \widehat{\boldsymbol{\alpha}}$ when using a log or identity link for the scale part of the model, respectively. Here, $\boldsymbol{z}_h$ denotes either a particular row from $\boldsymbol{Z}$ (yielding $\widehat{\tau}_i^2$) or some other combination of values for the scale variables. A corresponding approximate 95% CI for τ_h^2 is then given by either $\exp\left(\boldsymbol{z}_h \widehat{\boldsymbol{\alpha}} \pm 1.96\sqrt{\boldsymbol{z}_h \mathrm{Var}[\widehat{\boldsymbol{\alpha}}] \boldsymbol{z}_h'}\right)$ or $\boldsymbol{z}_h \widehat{\boldsymbol{\alpha}} \pm 1.96\sqrt{\boldsymbol{z}_h \mathrm{Var}[\widehat{\boldsymbol{\alpha}}] \boldsymbol{z}_h'}$. Note that when using an identity link, $\widehat{\tau}_h^2$ is only guaranteed to be non-negative when $\boldsymbol{z}_h$ is a row from $\boldsymbol{Z}$. Moreover, even then, the lower bound of the CI may be negative. While one could set negative $\widehat{\tau}_h^2$ values or CI bounds to 0, we can avoid these issues altogether by using the log link, since exponentiation guarantees non-negative predicted values and CI bounds in all cases.

2.5.3 | Likelihood ratio tests and confidence intervals

Likelihood ratio tests (LRTs) can also be used to compare models. For the LRT of one or multiple location/scale

coefficients, let $ll\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}}\right)$ denote the maximized log-likelihood under the full model and $ll\left(\widehat{\boldsymbol{\beta}}_0,\widehat{\boldsymbol{\alpha}}_0\right)$ the maximized log-likelihood under the reduced model where the location/scale coefficients to be tested are constrained to zero (which is equivalent to fitting a model where we remove from $\boldsymbol{X}$ and/or $\boldsymbol{Z}$ the columns that correspond to the location and/or scale coefficients tested). Under the null hypothesis that the corresponding true values of the location/scale coefficients are equal to zero, the LRT statistic

$$X^2 = -2 \times \left(ll\left(\widehat{\boldsymbol{\beta}}_0,\widehat{\boldsymbol{\alpha}}_0\right) - ll\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}}\right)\right) \tag{15}$$

then follows asymptotically a chi-square distribution with degrees of freedom equal to the number of coefficients tested.

It is also possible to test scale coefficients in this manner when using REML estimation. Here, we let $ll_R(\widehat{\boldsymbol{\alpha}})$ and $ll_R(\widehat{\boldsymbol{\alpha}}_0)$ denote the maximized restricted log-likelihood under the full and reduced model, respectively. Then

$$X^2_R = -2 \times \left(ll_R(\widehat{\boldsymbol{\alpha}}_0) - ll_R(\widehat{\boldsymbol{\alpha}})\right) \tag{16}$$

follows asymptotically a chi-square distribution with degrees of freedom equal to the number of scale coefficients tested.

For the standard random-effects model, Hardy and Thompson[25] describe how to "invert" the LRT to construct profile likelihood CIs for μ and τ^2 (or a confidence region for both parameters jointly). The same idea can be generalized to the present model, yielding profile likelihood CIs for particular model coefficients (or a confidence region for multiple coefficients). This approach is especially advantageous for the scale coefficients, since their sampling distribution may not be normal (which is implicitly assumed by (13) and the corresponding Wald-type CI) and we will therefore focus on the construction of profile likelihood CIs for this purpose.

Let $ll_P(\widetilde{\boldsymbol{\alpha}})$ denote the maximized profile log-likelihood when one or multiple scale coefficients are constrained not to zero, but to arbitrary values and $\chi^2_{r,.95}$ the 95th quantile of a chi-square distribution with r degrees of freedom, where r denotes the total number of scale parameters that were constrained. Then the set of all $\widetilde{\boldsymbol{\alpha}}$ values that satisfy

$$ll_P(\widetilde{\boldsymbol{\alpha}}) \geq ll_P(\widetilde{\boldsymbol{\alpha}}) - \chi^2_{r,.95}/2 \tag{17}$$

denotes a 95% CI (or confidence region) for the coefficients that were constrained. Similarly, letting $ll_R(\widetilde{\boldsymbol{\alpha}})$ denote the maximized restricted log-likelihood when one or multiple scale coefficients are constrained to arbitrary values, then the set of all $\widetilde{\boldsymbol{\alpha}}$ values that satisfy

$$ll_R(\widetilde{\boldsymbol{\alpha}}) \geq ll_R(\widetilde{\boldsymbol{\alpha}}) - \chi^2_{r,.95}/2 \tag{18}$$

denotes a 95% CI (or confidence region) for the constrained scale coefficients under REML estimation.

2.5.4 | Small-Sample Performance

As noted above, the distributional assumptions underlying the inferential methods presented here are based on asymptotics, that is, they rely on large-sample approximations. To be precise, "large-sample" in the present context primarily refers to the number of studies included in the analysis (although as noted in the footnote in section 2.1, the within-study sample sizes also need to be sufficiently large so that the sampling variances can be treated as approximately known). Moreover, when the model includes categorical predictors, then the number of studies within each category needs to be sufficiently large for the approximations to hold.

The methods described in section 2.5.1 for making inferences about the location part of the model are identical to those used in standard mixed-effects meta-regression models.[26,27] However, based on simulation studies in this context,[18] we know that the tests and CIs may not have nominal properties (i.e., their actual Type I error and coverage rates can deviate from the chosen level), especially when k is small. The Knapp-Hartung method[28] is a well known improvement over the standard Wald-type methods, leading to tests and CIs with close to nominal performance.[18] A generalization of the method is also possible for location-scale models.

Let $s^2 = \sum_{i=1}^{k} w_i (y_i - \widehat{y}_i)^2/(k-p-1)$ where $w_i = 1/\left(v_i + \widehat{\tau}^2_i\right)$. Now using $\mathrm{Var}\left[\widehat{\boldsymbol{\beta}}\right] = s^2(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}$ as the variance–covariance of the elements in $\widehat{\boldsymbol{\beta}}$, the test statistic (11) then follows an approximate t-distribution with $k-p-1$ degrees of freedom under H_0, while the 95% CI for β_j is obtained with $\widehat{\beta}_j \pm t_{.975;k-p-1} \times \mathrm{SE}\left[\widehat{\beta}_j\right]$, where $t_{.975;k-p-1}$ denotes the 97.5th quantile of a t-distribution with the same degrees of freedom. The test of multiple coefficients is then conducted with $F_\beta = Q_\beta/m$, which follows an approximate F-distribution with m and $k-p-1$ degrees of freedom under H_0, where m denotes the number of coefficients tested. Finally, to construct the 95% CI for a predicted average outcome, we then use $\widehat{y}_h \pm t_{.975;k-p-1}\sqrt{\boldsymbol{x}_h \mathrm{Var}\left[\widehat{\boldsymbol{\beta}}\right]\boldsymbol{x}'_h}$. These results follow directly from those given by Knapp and Hartung.[28]

An analogous generalization of the methods given in section 2.5.2 for making inferences about the scale part of the model is not currently available. We can, however, heuristically still compare the test statistic (13) to a t-distribution, now with $k-q-1$ degrees of freedom and construct the CI for α_j accordingly. Similarly, we can use $F_\alpha = Q_\alpha/m$ to test multiple scale coefficients, now letting m denote the number of scale coefficients tested, and use an F-distribution with m and $k-q-1$ degrees of freedom as a reference. Finally, in the 95% CI for τ^2_h, we simply replace 1.96 with $t_{.975;k-q-1}$. In essence, these are analogous heuristic adjustments that have previously been considered in the context of standard random-effects and meta-regression models.[29,30]

It is currently unknown how well the standard or adjusted methods perform in small samples and how large the number of studies needs to be for the methods to have nominal properties. However, given that accurate estimation and inferences about the amount of heterogeneity in standard random-effects models is already a difficult endeavor to begin with,[17,31,32] we suspect that k will need to be fairly large for the methods to have nominal properties. Therefore, at the moment, we would caution against the application of location-scale models in small meta-analyses.

2.6 | Profile Likelihood Plots

Fitting location-scale models is a non-trivial optimization problem, especially when the model includes a large number of scale variables. The $q+1$ dimensional surface of the profiled log-likelihood (7) or the restricted log-likelihood (9) may involve ridges, local optima, and saddle points, which can lead to convergence to a non-optimal solution. To obtain some reassurance that $ll_P(\widehat{\boldsymbol{\alpha}})$ or $ll_R(\widehat{\boldsymbol{\alpha}})$ really does correspond to its respective global maximum, we can make use of profile likelihood plots for each of the scale coefficients in the model.

To construct such a plot for a particular scale coefficient α_j, we fix the coefficient to some value near $\widehat{\alpha}_j$ and maximize (7) or (9) over the remaining scale coefficients. By repeating this process for a range of values around $\widehat{\alpha}_j$, we can examine how $ll_P(\boldsymbol{\alpha})$ or $ll_R(\boldsymbol{\alpha})$ changes as a function of α_j (i.e., we construct a profile of (7) or (9) along the dimension corresponding to α_j). Note that this is in essence the same process that is involved in finding a profile likelihood CI for α_j as described in the previous section.

The profile likelihood function constructed in this manner should have a peak at $\widehat{\alpha}_j$, indicating that $ll_P(\boldsymbol{\alpha})$ or $ll_R(\boldsymbol{\alpha})$ is really maximized along the dimension corresponding to α_j within the range of α_j values examined. By constructing such profiles for each scale parameter, we can check that the respective likelihoods are maximized along each dimension, at least within the vicinity of $\widehat{\boldsymbol{\alpha}}$.‡ See Raue et al.[33] for further details on the use of profile likelihoods for checking on the identifiability of parameters in complex models.

2.7 | Prediction Intervals

For the standard random-effects model (1), Raudenbush[26] suggested to compute $\widehat{\mu} \pm 1.96\widehat{\tau}$ as a "plausible value interval" that should contain approximately 95% of the true outcomes. A similar type of interval, referred to as a "credibility interval", was proposed by Hunter and Schmidt[34] to quantify the degree to which the underlying true outcomes may vary over studies. However, these intervals ignore the uncertainty in $\widehat{\mu}$ and hence an improved interval could be computed with $\widehat{\mu} \pm 1.96\sqrt{\widehat{\tau}^2 + \text{SE}[\widehat{\mu}]^2}$.§ Intervals of this type have also been referred to as "prediction intervals" (PI),[35] as they can also be interpreted as the range for the predicted true outcome in a new study. Given that this term has also found its way into popular textbooks on meta-analysis[36] and is based on similar concepts in regression modeling,[37] we will adopt the same terminology below.

To compute PIs in the context of a location-scale model, we need to specify the values of the location and scale variables. In particular, recall that the predicted average outcome for a particular combination of values for the location variables is given by $\widehat{y}_h = \boldsymbol{x}_h\widehat{\boldsymbol{\beta}}$, while $\widehat{\tau}^2_h = \exp(\boldsymbol{z}_h\widehat{\boldsymbol{\alpha}})$ (or $\widehat{\tau}^2_h = \boldsymbol{z}_h\widehat{\boldsymbol{\alpha}}$ when using an identity link) yields the predicted amount of (residual) heterogeneity for a particular combination of values for the scale variables. Hence, an approximate 95% PI for the true outcomes of studies (or a future study) at the chosen values of $\boldsymbol{x}_h$ and $\boldsymbol{z}_h$ is given by $\widehat{y}_h \pm 1.96\sqrt{\widehat{\tau}^2_h + \boldsymbol{x}_h \text{Var}\left[\widehat{\boldsymbol{\beta}}\right]\boldsymbol{x}'_h}$, with $\widehat{\tau}^2_h$ as defined above. When using the generalization of the Knapp-Hartung method described earlier, we replace 1.96 with $t_{.975;k-p-1}$.

2.8 | Visualization

The results of a meta-regression model involving a numerical/quantitative predictor can be visualized by plotting the observed outcomes on the y-axis against the values of the predictor on the x-axis and adding the regression line based on the model (with or without corresponding CI and/or PI bands) to such a plot.[38,39] Typically, the outcomes are drawn proportional in size to

some measure of their precision (e.g., $1/v_i$) or the model weights (i.e., the diagonal elements of $\boldsymbol{W}$, which are equal to $w_i = 1/(v_i + \widehat{\tau}_i^2)$ for location-scale models). Given their appearance, such scatter plots are also at times referred to as "bubble plots". This type of graph is equally applicable to visualize the results from the location part of a location-scale model involving a quantitative location variable. On the other hand, for a categorical predictor, one can simply add the estimated average outcomes for the various levels of the predictor (i.e., the subgroups defined by the predictor) to a standard forest plot.[39]

For illustrating the results from the scale part of a model, we suggest the following visualization. First, let $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}$ denote the hat matrix and h_i the ith diagonal element thereof. Furthermore, let $e_i = y_i - \widehat{y}_i$ denote the observed residual of the ith study which, under the assumptions of the model, can be shown to have expectation 0 and variance $\mathrm{Var}[e_i] = (1-h_i)(v_i + \tau_i^2)$. Hence, we can use $\widehat{\tau}_i^2 = e_i^2/(1-h_i) - v_i$ as an estimate of τ_i^2 (setting negative $\widehat{\tau}_i^2$ values to 0). We can therefore plot these estimates against the values of a quantitative scale variable as an analogue to the bubble plot described above. As above, the regression line, now for the predicted amount of heterogeneity as a function of the predictor, can be added to such a figure (with or without a corresponding CI band). For a categorical scale variable, differences in the amount of heterogeneity across subgroups can again be visualized as part of a forest plot, for example by showing the different PIs for the various subgroups defined by the predictor.

2.9 | Model Selection

In practice, one is often faced with a large number of potentially relevant location and/or scale variables one could include in a model. The problem of finding those predictors that are truly related (if any) to the location and/or scale of the outcomes can therefore be framed as a model selection problem.[40] While it is still common practice to examine one predictor at a time in a series of univariate meta-regression models,[41] this approach increases the risk of finding spurious relationships due to the fact that predictors are often correlated. Fitting models which include multiple predictors of interest can mitigate this problem at least to some extent.[42] The use of information-theoretic methods[43] for model selection in the context of meta-regression analyses was also recently explored and might constitute a promising alternative to the use of null-hypothesis significance testing.[40]

For this, we compute, for a set of potentially plausible models, one of several different information criteria such as the Akaike Information Criterion (AIC),[44] which is given by

$$AIC = -2ll + 2(p+q+2), \tag{19}$$

where ll is either ll_P or ll_R for ML or REML estimation, respectively, and $p+q+2$ corresponds to the total number of parameters of the location-scale model. Heuristically, we can regard the AIC as a measure that penalizes the fit of a model (as indicated by its [restricted] log-likelihood) for its complexity (as indicated by the number of included parameters). As expressed in (19), models with lower AIC values strike a better balance between fit and complexity and are therefore to be preferred.

An alternative is the Bayesian Information Criterion (BIC),[45] which is given by

$$BIC = -2ll + 2(p+q+2)\ln(k^*), \tag{20}$$

where $k^* = k$ for ML and $k^* = k-p-1$ for REML estimation. When $k^* \geq 8$, the BIC imposes a greater penalty for the model complexity compared to the AIC and therefore tends to favor simpler models. Similarly, the corrected AIC is given by

$$AICc = -2ll + 2(p+q+2)\left(\frac{k^*}{k^* - (p+q+2) - 1}\right), \tag{21}$$

where $k^* = \max(k, p+q+4)$ for ML and $k^* = \max(k-p-1, p+q+4)$ for REML estimation, which ensures that the additional multiplicative term is ≥ 1 and hence again implies a greater penalty for the number of parameters compared to the AIC.[46]

Strictly speaking, models that differ in terms of their fixed effects (i.e., location variables) should not be compared with respect to their restricted log-likelihoods,[47,48] which would imply that information criteria computed based on REML estimation would only be valid for model selection when comparing models including different scale but the same set of location variables. However, recent evidence suggests that information criteria computed based on REML estimation may even serve as a model selection tool when their fixed effects differ.[40,49] Regardless of this issue, further research is needed to examine the performance of information-theoretic methods for model selection in the present context.

2.10 | Implementation Details

The option to fit location-scale models was recently added in an update to the *metafor* package[16] for R[50] as part of the rma() function. Maximization of the profiled log-likelihood (7) or the restricted log-likelihood (9) (for ML and REML estimation, respectively) is accomplished by default using the quasi-Newton algorithm

implemented in the nlminb() function,[51] but the user also has the option to choose from a wide variety of alternative optimization routines should convergence issues arise. When using an identity link, constrained optimization using the Nelder–Mead (downhill simplex) method[52] is used in combination with an adaptive barrier algorithm[53] as implemented in the constrOptim() function to ensure non-negativity of $\boldsymbol{Z\alpha}$. The *numDeriv* package,[54] which provides accurate methods for numerical differentiation using Richardson extrapolation, is used to obtain the Hessian matrix of the scale coefficients, from which their variance–covariance matrix is estimated. The output for a fitted model includes Wald-type tests and CIs for the location and scale coefficients,¶ while LRTs and profile likelihood CIs (the latter only for the scale coefficients) can be obtained using the anova() and confint() functions, respectively. Model identifiability can be checked by drawing profile likelihood plots with the profile() function and fit statistics (including the AIC, BIC, and AICc) can be obtained with the fitstats() function. Finally, the predict() function can compute the predicted average outcome (with CI and PI) for a particular combination of values for the location variables and the predicted amount of (residual) heterogeneity for a particular combination of values for the scale variables.

3 | ILLUSTRATIVE EXAMPLE

In this section, we demonstrate the application of location-scale models using a dataset readily available in the *metafor* package. In this illustration, we model the scale part using a log link, use REML estimation (except when otherwise noted), and report Wald-type CIs for both the location and scale parts of the models (using the Knapp-Hartung generalization we described in section 2.5.4 for drawing inferences about location coefficients and the approximate t- and F-distributions for inferences about scale coefficients). We also illustrate the use profile likelihood plots and CIs for the scale coefficients. The analysis code can be found at https://osf.io/53mtg/.

Bangert-Drowns et al.[55] integrated the results from 48 studies examining the effectiveness of school-based interventions to improve educational achievement. Each study compared an experimental group of students who received an intervention focused on writing tasks (experimental group) against another group receiving conventional instruction (control group) with respect to some measure of academic achievement (e.g., final grade, an exam/quiz/test score). The outcome measure was the standardized mean difference (with positive scores favoring the intervention group),[56] which we corrected for its small-sample bias.**

The standard random-effects model yields an estimated overall effect of $\widehat{\mu} = 0.22$ (95% CI: 0.12 to 0.32, $p < 0.001$), suggesting that intervention groups obtained on average higher academic achievement scores than control groups. However, the between-study variance estimate of $\widehat{\tau}^2 = 0.050$ leads to a 95% PI around $\widehat{\mu}$ from -0.24 to 0.68, which reveals substantial heterogeneity in the effectiveness of such interventions across studies. Figure 1 shows a forest plot of the individual effect size estimates with the results from the random-effects model at the bottom (the dotted interval around the summary polygon indicates the PI bounds). The results from the location-scale model will be discussed further below.

As a quick check of the routines, we can also fit the standard random-effects model as a location-scale model by setting $\boldsymbol{X}$ and $\boldsymbol{Z}$ both to column vectors of 1's. Doing so yields the same estimate and CI for μ and an estimate of $\widehat{\alpha}_0 = -2.997$ (with $\mathrm{SE}[\widehat{\alpha}_0] = 0.4603$) and therefore the model implies $\widehat{\tau}^2 = \exp(-2.997) = 0.050$ as above. An interesting feature of this approach is that a 95% CI for τ^2 can be readily constructed with $\exp(-2.997 \pm 2.01 \times 0.4603) = (0.020, 0.126)$ (where $t_{.975;47} = 2.01$), although it remains to be examined how this CI compares (in terms of coverage and width) to other methods for constructing CIs for τ^2 in the context of the random-effects model.[32,57]

Next, we explored the association of two predictors with the size (i.e., location) and amount of heterogeneity (i.e., scale) of the outcomes. As an example of a quantitative predictor, we included the total sample size of each study (range 16–542, with a mean and median of 116 and 68 participants, respectively) in the model, which we rescaled for interpretation purposes for the analyses (keeping the original scale for graphical display), so that the location and scale parts of the model can be written as $y_i = \beta_0 + \beta_1(n_i/100)$ and $\ln(\tau_i^2) = \alpha_0 + \alpha_1(n_i/100)$, respectively. We also examined a categorical predictor, namely the subject matter that was taught in each study (mathematics: 28 studies; science: 9 studies; social science: 11 studies). This predictor was incorporated into the model as two dummy variables, one for science and the other for social science subjects (and hence using math as the reference category), resulting in the model $y_i = \beta_0 + \beta_1 sci_i + \beta_2 soc_i$ for the location part and $\ln(\tau_i^2) = \alpha_0 + \alpha_1 sci_i + \alpha_2 soc_i$ for the scale part. We first fitted two separate models testing the association of each predictor with both the location and scale of the outcomes, and then ran an additional analysis incorporating both into a model with multiple predictors, that is, $y_i = \beta_0 + \beta_1(n_i/100) + \beta_2 sci_i + \beta_3 soc_i$ for the location part and $\ln(\tau_i^2) = \alpha_0 + \alpha_1(n_i/100) + \alpha_2 sci_i + \alpha_3 soc_i$ for the scale part.

The model with sample size as predictor provided evidence of a negative association with the size of the

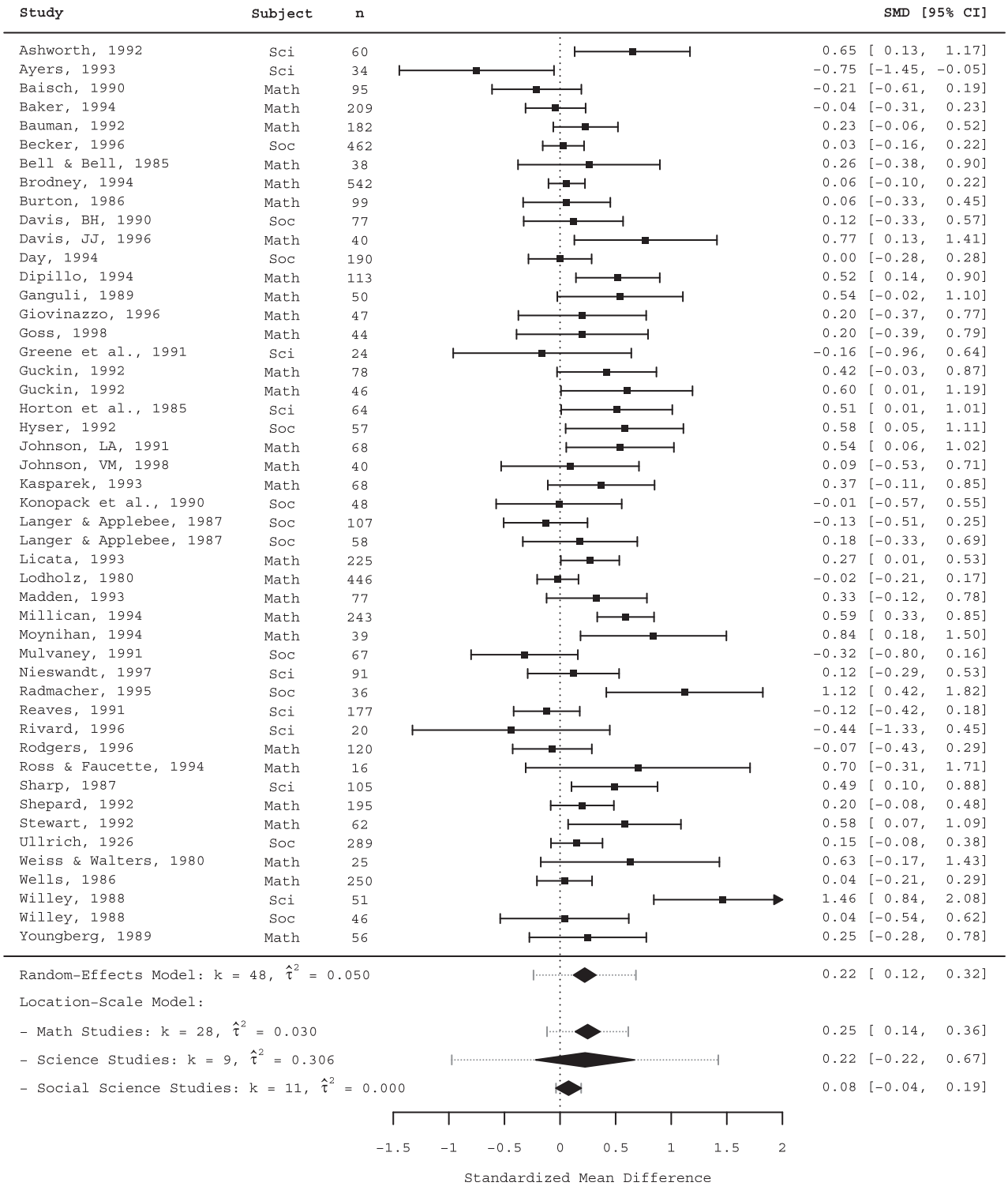

FIGURE 1 Forest plot of the studies from Bangert-Drowns et al.[55] with the results from the random-effects model and the location-scale model (including the study subject as predictor) shown.

outcomes ($\widehat{\beta}_1 = -0.055$, 95% CI: −0.095 to −0.015, $p = .008$) and weaker evidence of a negative association with the amount of heterogeneity ($\widehat{\alpha}_1 = -0.917$, 95% CI: −1.952 to 0.117, $p = 0.081$). Therefore, studies with larger sample sizes tended to yield smaller (and maybe more homogeneous) outcomes. These associations are shown in the bubble plots in Figure 2. In Figure 2a, we included both the CI and PI bands around the regression line, the latter illustrating the shrinking of the amount of heterogeneity for larger studies. This is also what Figure 2b shows, in terms

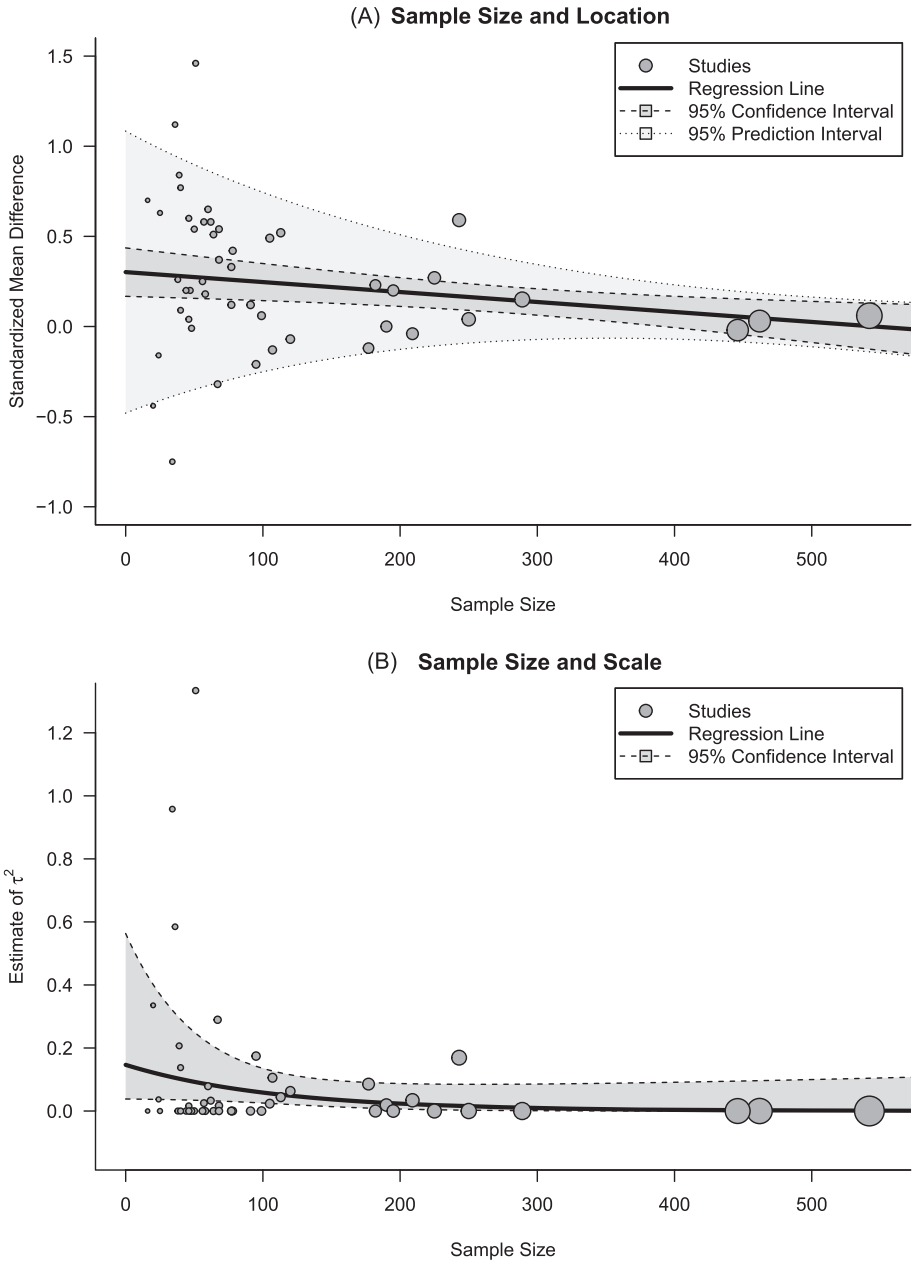

FIGURE 2 Bubble plots showing the association between the sample size and the location and scale parts of the model in the example of Bangert-Drowns et al.[55]

of the predicted value of τ^2 as a function of the sample size (with corresponding CI for the predicted values).

With regards to subject type, the estimated average effects and between-study variances for the three subject categories are presented at the bottom of Figure 1. The estimated average effect for studies focused on social science subjects was $\widehat{\beta}_0 + \widehat{\beta}_2 = \widehat{\mu}_{soc} = 0.08$ (95% CI: −0.04 to 0.19, $p = 0.18$), as opposed to $\widehat{\beta}_0 + \widehat{\beta}_1 = \widehat{\mu}_{sci} = 0.22$ (95% CI: −0.22 to 0.67, $p = 0.31$) for science studies and $\widehat{\beta}_0 = \widehat{\mu}_{mat} = 0.25$ (95% CI: 0.14 to 0.36, $p < .001$) for math studies. The omnibus test for the location part of the model (i.e., $H_0 : \beta_1 = \beta_2 = 0$) yielded weak evidence of an association ($F_\beta(2,45) = 2.43$, $p = 0.099$). However, hypothesis tests for specific location coefficients showed evidence that social science studies reported on average effect size estimates of smaller magnitude than math studies ($\widehat{\beta}_2 = -0.17$, 95% CI: −0.33 to −0.01, $p = 0.034$). Similarly, the omnibus test for the scale part (i.e., $H_0 : \alpha_1 = \alpha_2 = 0$) provided some evidence of an association between subject type and the amount of heterogeneity in the outcomes ($F_\alpha(2,45) = 3.32$, $p = 0.045$). In particular, studies focused on science subjects yielded more heterogeneous outcomes ($\exp(\widehat{\alpha}_0 + \widehat{\alpha}_1) = \widehat{\tau}^2_{sci} = 0.306$, 95% PI around $\widehat{\mu}_{sci}$ from −0.97 to 1.42) than math studies ($\exp(\widehat{\alpha}_0) = \widehat{\tau}^2_{mat} = 0.030$, 95% PI around $\widehat{\mu}_{mat}$ from −0.12 to 0.61) and social science studies ($\exp(\widehat{\alpha}_0 + \widehat{\alpha}_2) = \widehat{\tau}^2_{soc} = 0.000$, 95% PI around $\widehat{\mu}_{soc}$ from −0.04 to 0.19).

TABLE 1 Multiple meta-regression results using the data from Bangert-Drowns et al.[55]

Coefficient	$\widehat{\beta}$	95% CI for $\widehat{\beta}$	$\widehat{\alpha}$	95% CI for $\widehat{\alpha}$	95% PLCI[a] for $\widehat{\alpha}$
Intercept	0.344	0.210 to 0.478	−3.102	−5.100 to −1.105	< −8 to −1.276
Sample Size	−0.058	−0.099 to −0.018	−0.539	−1.682 to 0.604	−7.159 to 0.551
Sciences vs. Math	−0.080	−0.487 to 0.327	2.233	0.122 to 4.344	0.332 to > 10
Social Sciences vs. Math	−0.109	−0.274 to 0.057	0.401	−2.425 to 3.227	< −10 to > 10

[a]PLCI = profile likelihood confidence interval.

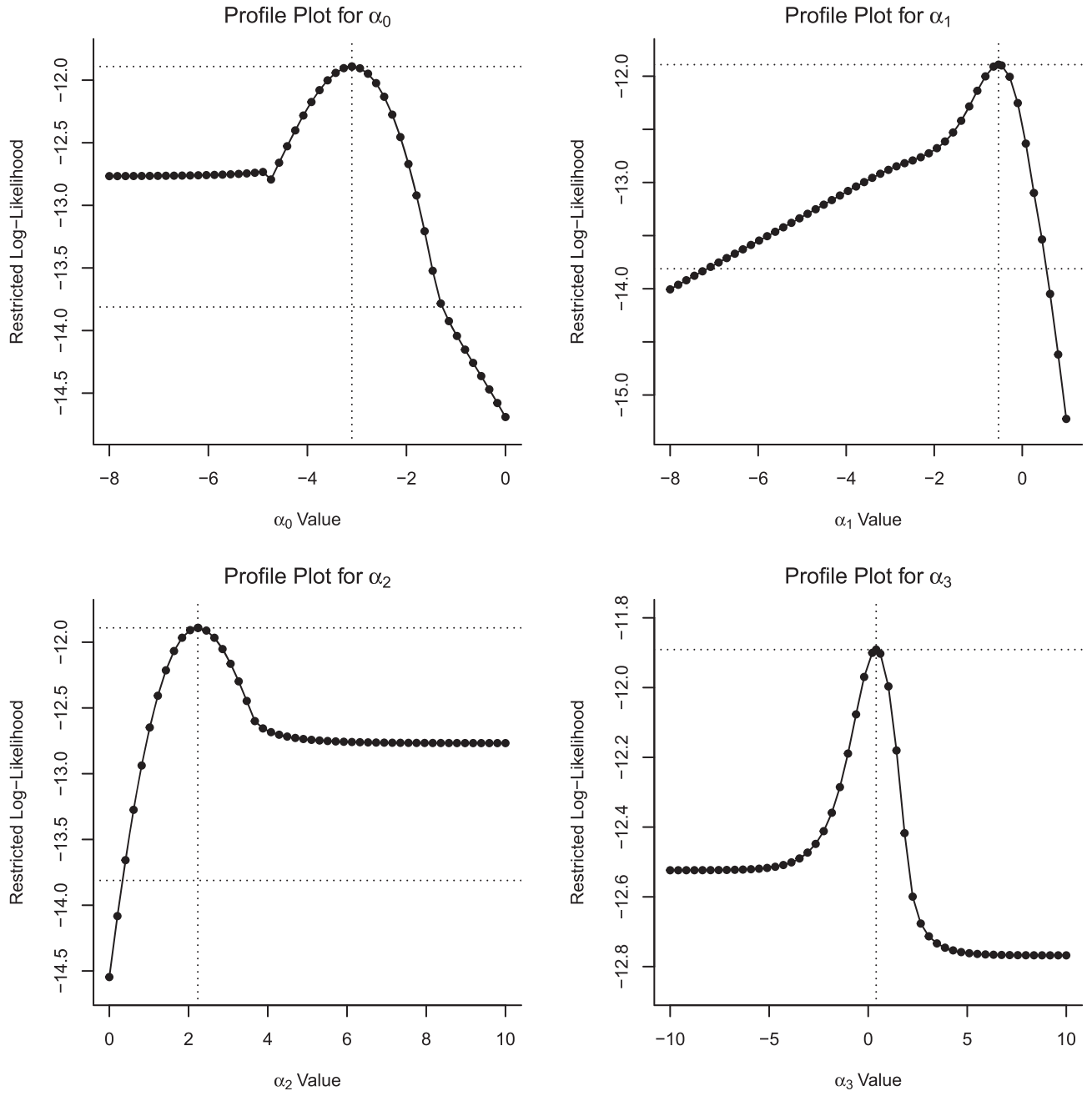

FIGURE 3 Profile likelihood plots for the scale coefficients in the model with multiple location and scale predictors in the example of Bangert-Drowns et al.[55]

Of note, the estimates of the average effect and between-study variance for each subject type from the previous location-scale model can also be obtained by fitting separate random-effects models for each category of the moderator. Doing so leads to estimates of $\widehat{\mu}_{mat} = 0.25$ and $\widehat{\tau}^2_{mat} = 0.030$ for math studies, $\widehat{\mu}_{sci} = 0.22$ and $\widehat{\tau}^2_{sci} = 0.306$ for science studies, and $\widehat{\mu}_{soc} = 0.08$ and $\widehat{\tau}^2_{soc} = 0.000$ for social science studies, respectively. This illustrates the equivalence between separate random-effects and location-scale models when a single categorical moderator is considered (but only in this special case).††

Table 1 presents the results of the model including sample size and subject type as predictors for both the location and scale of the outcomes. For illustration purposes and due to the increasing complexity of the model, Figure 3 presents profile likelihood plots for the scale coefficients in the model, which do not suggest estimation problems.‡‡ Furthermore, in addition to the Wald-type CIs reported throughout the paper, we also report in Table 1 profile likelihood CIs for the scale parameters (calculated with Equation 18). However, due to the flatness of some of the likelihood profiles for values of α_j further away from $\widehat{\alpha}_j$ (see Figure 3), some of the bounds cannot be obtained exactly.

Irrespective, both interval types lead to the same statistical conclusions, and hence for simplicity we focus on the Wald-type CIs for the result interpretation. The omnibus tests of the location and scale coefficients (i.e., $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ and $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$) showed some evidence of associations for both parts ($F_\beta(3,44) = 3.44$, $p = 0.025$ and $F_\alpha(3,44) = 2.70$, $p = 0.057$, respectively). After controlling for subject type, there was still evidence of an association between the sample size of the studies and the size of the outcomes ($\widehat{\beta}_1 = -0.058$, 95% CI: −0.099 to −0.018, $p = 0.006$) but not with the amount of heterogeneity ($\widehat{\alpha}_1 = -0.539$, 95% CI: −1.682 to 0.604, $p = 0.35$). Tests of $H_0: \beta_2 = \beta_3 = 0$ and $H_0: \alpha_2 = \alpha_3 = 0$ can be used to test for differences between the subject types after controlling for sample size. For the location part, the test indicated no evidence of an association between subject type and the size of the outcomes ($F_\beta(2,44) = 0.91$, $p = 0.41$), whereas the test result for the scale part of the model suggested some weak evidence of an association with the amount of heterogeneity ($F_\alpha(2,44) = 2.39$, $p = 0.10$). When examining the individual scale coefficients, there was evidence that interventions focused on science subjects yielded more heterogeneous outcomes than those focused on math subjects ($\widehat{\alpha}_2 = 2.233$, 95% CI: 0.122 to 4.344, $p = 0.039$).

The previous model can be used to make predictions for μ and τ^2 in future studies. For the location part, the multiple meta-regression model predicts a value of $\widehat{\mu} = 0.32$ (95% PI: −0.08 to 0.71) for a study focused on math subjects with 50 participants, but the predicted effect decreases to $\widehat{\mu} = 0.29$ (95% PI: −0.06 to 0.63) and $\widehat{\mu} = 0.26$ (95% PI: −0.04 to 0.56) as the sample size increases to 100 and 150 participants, respectively. With regards to the scale part, keeping the sample size fixed at 100 participants, the model predicts an amount of heterogeneity of $\widehat{\tau}^2 = 0.026$ (95% CI: 0.006 to 0.121) for math-focused interventions, which increases to $\widehat{\tau}^2 = 0.039$ (95% CI: 0.004 to 0.437) for social science and to $\widehat{\tau}^2 = 0.245$ (95% CI: 0.060 to 1.001) for science subjects.

In the models above, the same predictors were used for the location and scale parts. To illustrate that this is not a requirement, we fitted a model with sample size as a location moderator and subject type as a scale moderator, that is, $y_i = \beta_0 + \beta_1(n_i/100)$ for the location part and $\ln(\tau^2_i) = \alpha_0 + \alpha_1 sci_i + \alpha_2 soc_i$ for the scale part. Results in Table 2 show evidence of a negative association between sample size and the magnitude of the outcomes ($\widehat{\beta}_1 = -0.062$, 95% CI: −0.116 to −0.008, $p = 0.026$). Furthermore, there was evidence that science studies yielded more heterogeneous outcomes than math studies ($\widehat{\alpha}_1 = 2.597$, 95% CI: 0.529 to 4.666, $p = 0.015$), whereas no significant difference was found between social science and math studies ($\widehat{\alpha}_2 = 0.520$, 95% CI: −2.739 to 3.779, $p = 0.75$).

Since this model is nested within the model that includes both sample size and subject type in the location and scale parts of the model, we can therefore also conduct a LRT, examining if the full model (including both predictors in both parts) provides a significantly better fit than the model that only includes sample size as a

TABLE 2 Multiple meta-regression results with different predictors for the location and scale parts, using the data from Bangert-Drowns et al.[55]

Coefficient	$\widehat{\beta}$	95% CI for $\widehat{\beta}$	$\widehat{\alpha}$	95% CI for $\widehat{\alpha}$	95% PLCI[a] for $\widehat{\alpha}$
Intercept	0.319	0.190 to 0.448	−3.957	−5.511 to −2.402	−11.217 to −2.718
Sample Size	−0.062	−0.116 to −0.008			
Sciences vs. Math			2.597	0.529 to 4.666	0.654 to 9.856
Social Sciences vs. Math			0.520	−2.739 to 3.779	< −10 to 7.700

[a]PLCI = profile likelihood confidence interval.

TABLE 3 Log-likelihoods and fit criteria values for the various models fitted using the data from Bangert-Drowns et al.[55]

	Location		Scale					
Method	Sample Size	Subject Type	Sample Size	Subject Type	logLik	AIC	BIC	AICc
ML					−18.26	40.52	44.27	40.79
	✓		✓		−13.24	34.48	41.96	35.41
		✓		✓	−13.20	38.40	49.62	40.45
	✓	✓	✓	✓	−10.08	36.16	51.13	39.86
	✓			✓	−12.50	35.00	44.35	36.43
REML					−18.49	40.99	44.69	41.26
	✓		✓		−14.65	37.30	44.62	38.28
		✓		✓	−13.99	39.97	50.81	42.18
	✓	✓	✓	✓	−11.89	39.78	54.06	43.90
	✓			✓	−13.55	37.10	46.24	38.60

location moderator and subject type as a scale moderator. Since the two models differ in terms of their location coefficients, we must refit the two models using ML estimation for the LRT to be meaningful. After doing so, (15) yields $X^2 = 4.83$ based on 3 degrees of freedom, resulting in $p = 0.18$ and hence no statistically significant evidence that the full model provides a better fit.

To illustrate the use of information criteria for model selection, Table 3 shows the log-likelihood, AIC, BIC, and AICc values computed based on ML and REML estimation for all models considered above. All criteria favor the model including sample size in the location and scale parts of the model regardless of the estimation method used. The only exception to this was the AIC computed based on REML estimation, which was lower for the model including sample size in the location and subject type in the scale part of the model, although only by a thin margin.

4 | DISCUSSION

In this paper, we have described a location-scale model for meta-analysis as an extension of the standard random- and mixed-effects models that not only allows an examination of whether predictors are related to the size of the outcomes (i.e., their location), but also the amount of heterogeneity (i.e., their scale). Together with a description of the methods for fitting and drawing inferences based on this model and an example illustrating its use, we have also provided an implementation via the *metafor* package for R that makes this model readily available to researchers. Of note, the use of this model does not require any additional information beyond what is necessary for fitting standard meta-regression models (except if hypotheses concerning the scale part involve variables that have not already been collected for the purposes of a standard moderator analysis).

At the same time, we want to emphasize that guidelines and caveats related to standard meta-regression analyses[38,41,58] are equally applicable when examining scale variables. In particular, researchers should formulate a priori hypotheses to motivate the examination of the scale variables and why/how they might be related to the amount of heterogeneity in the outcomes. For example, if the specification of clinical guidelines has led to an increased consistency in how a particular treatment has been implemented over time, one could hypothesize that the results from more recent trials might tend to be more consistent (i.e., exhibit lower amounts of heterogeneity) than earlier trials. On the other hand, more recent studies might explore the generalizability of a treatment effect by examining its effectiveness in more diverse populations, which in turn might lead to increased heterogeneity. In either case, such hypotheses should be formulated before embarking on such analyses.

Even when the analyses are pre-specified and hypothesis driven, the potential for making at least one Type I error increases with the number of predictors examined. Although not common practice in meta-regression analyses,[41] researchers should consider the use of corrections for multiple testing to reduce the number of false positive associations. This, however, comes at the cost of decreased power to detect true associations. In fact, we suspect that the number of studies required for location-scale models to have sufficient power to detect associations between scale variables and the amount of heterogeneity is fairly high to begin with. This, however, needs to be examined further via simulation studies.

It is also important to emphasize that such analyses (whether they concern location or scale variables) are purely observational and hence any associations found could be confounded by other variables not controlled for in the analyses.[42] Hence, as is the case for standard meta-regression analyses,[59] "synthesis-generated evidence" about scale variable associations should be treated with due caution.

Finally, any associations found, either with respect to location or scale variables, reflect relationships that exist at the study level which does not imply that similar relationships exist at the participant level. For example, if the mean age of the participants is found to be positively related to the size of a treatment effect across studies, then this indicates that studies including on average older participants tended to find larger effects, but this does not imply that the treatment tended to be more effective for older participants within studies (this may or may not be true). We illustrate this idea schematically in Figure 4a, showing the results from 50 trials where indeed such a relationship at the study level is present (i.e., the points correspond to the mean age values and

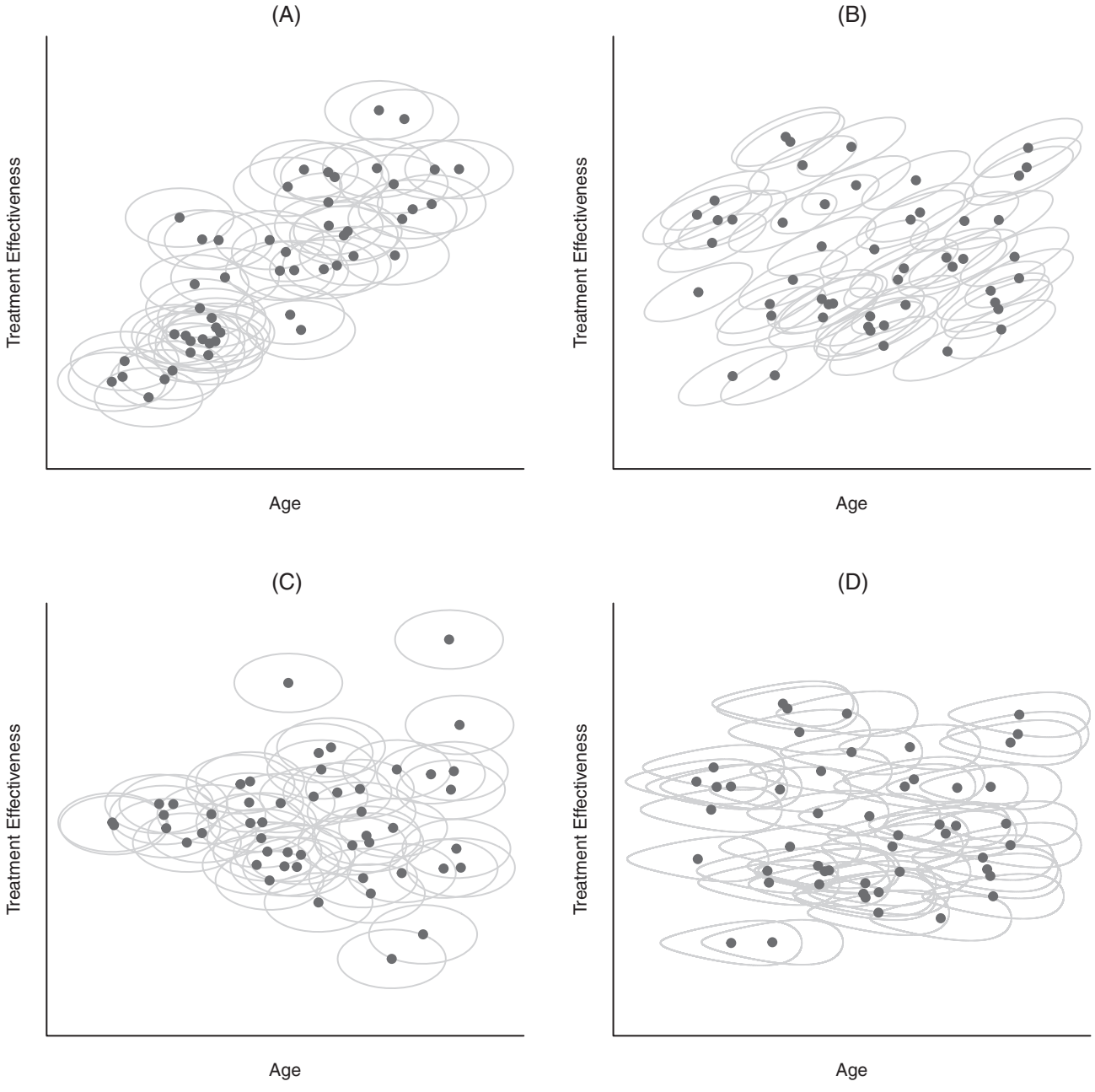

FIGURE 4 Schematic illustrations of the between- and within-study relationship between the (mean) age of study participants and the treatment effects in a set of 50 hypothetical trials (see text for explanations).

the corresponding observed effects), but within studies, there is no relationship between the age of the participants and the treatment effectiveness (i.e., the ovals represent envelopes containing the raw data). In contrast, Figure 4b illustrates the opposite scenario, where no relationship exists at the study level, but within studies the treatment is more effective for older participants.

Similarly, if the mean age of the study participants is found to be positively associated with the amount of heterogeneity in the studies, then this indicates that studies including older participants tended to yield more heterogeneous results. We illustrate this scenario in Figure 4c, where we see increased between-study variance in the effects for studies with older participants. However, such a finding does not imply that the treatment effect varies more strongly for older participants within studies. This case is shown in Figure 4d, where treatment effectiveness indeed varies more strongly for older participants within studies (note the increase in variance in the raw data as a function of age).

To properly examine such within-study relationships, studies would need to quantify these types of associations directly, for example by reporting the correlation between the age of the study participants and the observed treatment effects or their variability. Alternatively, one may be able to quantify within-study differences in treatment effects or their variability if the studies report subgroup results (e.g., for younger versus older participants). Ideally, if the raw data of the individual studies are available, then an individual participant data meta-analysis could also be conducted where between- and within-study relationships can be properly disentangled.[60] This would be equally true for relationships involving scale variables, in which case multilevel location-scale models[10,11,12] would need to be used for the meta-analysis.

On a related note, there has been an increased interest recently in meta-analyzing not only outcome measures that quantify the central tendency (i.e., average) of a quantitative response variable, but also its variability within groups or group differences thereof.[61,62,63] In other words, while a more "traditional" meta-analysis might for example synthesize estimates of the difference in the mean blood pressure of treatment groups receiving antihypertensive therapy versus control groups, a meta-analysis could also examine if the within-group variability (as measured, for example, by the variance or standard deviation of the blood pressure measurements) differs between treatment versus control groups (e.g., by computing the ratio of the two standard deviations for each study and, after applying a suitable normalizing and variance stabilizing transformation, synthesizing these outcomes). It should be noted that meta-analyses of this type are addressing a different question than what can be examined with the location-scale model we have described in the present paper, which is focused on the between-study variability of the outcomes (i.e., are the findings of studies more heterogeneous under certain conditions than others?).§§ Hence, while there is a difference in purpose between these different approaches, they both shift (at least to some extent) attention away from questions about averages to questions about variances, opening up avenues for new insights.

Moreover, as we have demonstrated in the illustrative example, variables used as predictors for the location and scale parts of the model do not have to coincide. Therefore, one could even consider a model containing no location variables at all (except for the intercept term, allowing the average outcome to differ from zero), placing the focus entirely on an examination of scale variables to investigate under what conditions the outcomes of studies are more or less heterogeneous within a particular meta-analysis.

Questions about differences in heterogeneity have been raised previously. In particular, a number of prior studies examined to what extent estimates of heterogeneity (or some derivative measure such as I^2) differ across meta-analyses.[64,65,66,67,68] However, these studies compared measures of heterogeneity across entire meta-analyses (differing for example in terms of the effect size measure used or the types of outcomes or interventions studied), while the location-scale model described in the present paper allows for an examination of differences in heterogeneity across the studies included in a single meta-analysis.

Conceptually closer to this idea was the study by IntHout et al.,[69] who examined differences in estimates of τ^2 for the larger versus smaller studies within individual meta-analyses. Across 235 meta-analyses that had used the standardized mean difference as the outcome measure, they found that smaller studies (with a total sample size below roughly 50 participants) tended to yield an estimate of τ^2 that was on average 3.11 times larger than the estimate of τ^2 of larger studies (with more than 50 participants). This is in line with what we found in our illustrative example (see Figure 2b), showing a decrease in the amount of heterogeneity for larger studies. In fact, the predicted value of τ^2 for a sample size of 36 (the mean sample size of the smaller studies with less than 50 participants) was $\widehat{\tau}^2 = 0.105$, compared to $\widehat{\tau}^2 = 0.035$ for a sample size of 156 (the mean sample size of the larger studies with more than 50 participants), yielding a ratio of 3, which is remarkably close to the ratio found by IntHout et al.[69] By using a location-scale model, we could however avoid the arbitrary dichotomization of the studies into "small" versus "large" ones and directly model the relationship between τ^2 and the sample sizes.

It needs to be emphasized that this finding is unrelated to what we expect to see in a funnel plot, namely a decrease in the variability of the estimates for larger studies (at least when the true outcomes are sufficiently homogeneous). This phenomenon is attributable to the decrease in the sampling (or within-study) variance for larger studies and is a natural consequence of the consistency of the estimators used for calculating the observed estimates. However, the extent to which the amount of heterogeneity in the underlying true effects/outcomes differs across smaller versus larger studies within a particular meta-analysis is an empirical question that needs to be examined on a case-by-case basis, irrespective of the general trend found by IntHout et al.[69]

The location-scale model we have described in the present paper assumes independence between the observed outcomes or effect size estimates included in the same analysis. This assumption is often violated in practice, for example when multiple effect size estimates (e.g., for different response scales) are computed based on the same group of study participants, when multiple effect size estimates are computed by contrasting several different treatment groups against a common control group within at least some of the studies, or when the data have some other hierarchical structure (e.g., when multiple studies included in the meta-analysis were conducted by the same author or research lab).[70,71,72,73] An appropriate analysis of such dependent estimates requires the use of more complex models, possibly including multiple random effects (e.g., for studies and estimates within studies) while accounting for potential covariance in the sampling errors of the estimates. One could extend such multilevel/multivariate meta-analysis models to also include a scale model for each variance component (e.g., for the amount of between- and within-study heterogeneity), although this would increase the complexity considerably and require even more nuanced considerations as to the types of scale variables that may be associated with the various sources of variability. At the moment, the rma.mv() function in the *metafor* package that can be used to fit multilevel and multivariate meta-analysis models has not been extended to allow for this possibility, although this could be considered in a future update. For now, one could circumvent this issue by using subsets of the data that contain only independent estimates and/or data aggregated to a level at which the estimates can be assumed to be independent to fit location-scale models.

Aside from some interesting applications, we hope that the present paper will spark further research into the statistical properties of location-scale models in the present context and further extensions. As we alluded to earlier, we suspect that the increased complexity of such models will require a sufficiently large number of studies to yield accurate estimates especially for the scale part of the model. Under the usual regularity conditions, we can reason that the ML/REML estimates of location-scale models will be asymptotically fully efficient and that the size of tests and the coverage rate of CIs will be nominal when k is large, but the specific conditions under which such behavior holds will require examination.

Still, we believe that location-scale models are a useful tool for researchers interested in exploring whether the amount of heterogeneity may differ as a function of one or multiple predictor variables within a meta-analysis, broadening the research questions that can be addressed in the field of evidence synthesis.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

ENDNOTES

* For many outcome or effect size measures, the sampling variances are a function of one or multiple unknown parameters, which in practice need to be estimated based on the sample characteristics. In this case, the sampling variances are not really known constants, but estimates themselves. However, as long as the sample sizes of the studies are not too small, the sampling variances can be treated as approximately known.

† Technically, we only need to ensure that the diagonal elements of $\boldsymbol{M}$ (i.e., the $v_i + \tau_i^2$ values) are positive, so that $\boldsymbol{M}$ can be inverted. However, since we may be interested in and want to interpret the τ_i^2 values themselves, we prefer to enforce non-negativity of the $\boldsymbol{Z\alpha}$ values directly.

‡ Even if this is indeed the case, it is of course possible that $\widehat{\boldsymbol{\alpha}}$ only corresponds to a local maximum and that the global maximum lies in some region of the likelihood surface even further away from $ll_P(\widehat{\boldsymbol{\alpha}})$ or $ll_R(\widehat{\boldsymbol{\alpha}})$.

§ This interval still ignores the uncertainty in $\widehat{\tau}^2$. As a heuristic suggestion, Higgins et al.[35] propose to improve on this further by using the 97.5th quantile of a t-distribution with $k-2$ degrees of freedom in place of 1.96.

¶ To avoid any potential confusion, we note that the omnibus test statistics Q_β and Q_α are denoted as QM and QS, respectively, in the output, although when using the adjustments described in section 2.5.4, these omnibus tests are denoted as F statistics.

** Bangert-Drowns et al.[55] only report the total sample size of each study (n_i), not the sizes of the experimental and control groups separately (i.e., n_{1i} and n_{2i}, respectively). We therefore assumed $n_{1i} = n_{2i} = n_i/2$ for computing the sampling variances of the (bias-corrected) standardized mean differences.

†† A slight difference will still arise with respect to the inferences about the location coefficients when applying the Knapp-Hartung method. In the location-scale model, the method involves a single scaling factor, s^2, to adjust the variance–covariance matrix of the elements in $\widehat{\boldsymbol{\beta}}$ and the degrees of freedom for the t-distribution are taken to be $k-p-1$. On the other hand, when fitting separate random-effects models within each level of the categorical moderator, separate scaling factors are calculated within each

level and the degrees of freedom are $k_l - 1$, where k_l denotes the number of studies within level l of the moderator.

[‡‡] However, note that the profile for α_0 shows non-monotonic behavior for low values of α_0. Hence, the value of $\alpha_0 \approx -4.887$ corresponds to a local maximum. Fortunately, this did not prevent the optimization algorithm from finding the (presumably) global maximum at $\widehat{\alpha}_0 \approx -3.102$.

[§§] The location-scale model could however also be used to meta-analyze outcome measures that reflect such between-group differences in within-group variability. Location variables would then be used to examine if the size of such between-group differences in variability are on average larger under certain circumstances than others, while the scale part of the model would be used to examine if the amount of heterogeneity in such between-group differences is larger for certain types of studies versus others.

DATA AVAILABILITY STATEMENT

The code and data that support the findings of this study are openly available at the Open Science Framework (https://osf.io/53mtg/).

ORCID

Wolfgang Viechtbauer https://orcid.org/0000-0003-3463-4063
José Antonio López-López https://orcid.org/0000-0002-9655-3616

REFERENCES

1. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97-111.
2. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J*. 1994;309(6965):1351-1355.
3. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.
4. Hedges LV. A random effects model for effect sizes. *Psychol Bull*. 1983;93(2):388-395.
5. Borenstein M, Higgins JPT. Meta-analysis and subgroups. *Prev Sci*. 2013;14(2):134-143.
6. Raudenbush SW, Bryk AS. Empirical Bayes meta-analysis. *Journal of Educational Statistics*. 1985;10(2):75-98.
7. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med*. 1999;18(20):2693-2708.
8. Cook RD, Weisberg S. Diagnostics for heteroscedasticity in regression. *Biometrika*. 1983;70(1):1-10.
9. Carroll RJ, Ruppert D. *Transformation and Weighting in Regression*. Chapman & Hall; 1988.
10. Hedeker D, Mermelstein RJ, Demirtas H. An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*. 2008;64(2):627-634.
11. Hedeker D, Mermelstein RJ, Demirtas H. Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Stat Med*. 2012;31(27):3328-3336.
12. Li X, Hedeker D. A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Stat Med*. 2012;31(26):3192-3210.
13. Hedeker D, Nordgren R. MIXREGLS: a program for mixed-effects location scale analysis. *J Stat Softw*. 2013;52(2):1-38.
14. Bowater RJ, Escarela G. Heterogeneity and study size in random-effects meta-analysis. *Journal of Applied Statistics*. 2013;40(1):2-16.
15. Thompson CG, Becker BJ. A group-specific prior distribution for effect-size heterogeneity in meta-analysis. *Behav Res Methods*. 2020;52(5):2020-2030.
16. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1-48.
17. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*. 2005;30(3):261-293.
18. Viechtbauer W, López-López JA, Sánchez-Meca J, Marín-Martínez F. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychol Methods*. 2015;20(3):360-374.
19. Nocedal J, Wright SJ. *Numerical optimization*. 2nd ed. Springer; 2006.
20. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc*. 1977;72(358):320-338.
21. Patterson HD, Thompson R. Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometrics Conference*; Biometric Society. 1974:197-207.
22. Corbeil RR, Searle SR. A comparison of variance component estimation. *Biometrics*. 1976;32(4):779-791.
23. Corbeil RR, Searle SR. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Dent Technometrics*. 1976;18(1):31-38.
24. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes: the Art of Scientific Computing*. 3rd ed. Cambridge University Press; 2007.
25. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med*. 1996;15(6):619-629.
26. Raudenbush SW. Analyzing effect sizes: random-effects models. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. 2nd ed. Russell Sage Foundation; 2009:295-315.
27. Konstantopoulos S, Hedges LV. Statistically analyzing effect sizes: fixed- and random-effects models. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. 3rd ed. Russell Sage Foundation; 2019:245-279.
28. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med*. 2003;22(17):2693-2710.
29. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. *Biometrics*. 1999;55(3):732-737.
30. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Stat Med*. 1995;14(4):395-411.
31. Viechtbauer W. Hypothesis tests for population heterogeneity in meta-analysis. *British Journal of Mathematical and Statistical Psychology*. 2007;60(1):29-60.

32. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat Med*. 2007;26(1):37-52.
33. Raue A, Kreutz C, Maiwald T, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*. 2009;25(15):1923-1929.
34. Hunter JE, Schmidt FL. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage; 1990.
35. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A*. 2009;172(1):137-159.
36. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Wiley; 2009.
37. Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. 5th ed. McGraw-Hill; 2005.
38. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559-1573.
39. Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: an overview with suggestions for practice. *Res Synth Methods*. 2010;1(1):66-80.
40. Cinar O, Umbanhowar J, Hoeksema JD, Viechtbauer W. Using information-theoretic approaches for model selection in meta-analysis. *Res Synth Methods*. 2021;12(4):537-556.
41. Tipton E, Pustejovsky JE, Ahmadi H. Current practices in meta-regression in psychology, education, and medicine. *Res Synth Methods*. 2019;10(2):180-194.
42. Lipsey MW. Those confounded moderators in meta-analysis: good, bad, and ugly. *Ann Am Acad pol Soc Sci*. 2003;587:69-81.
43. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. Springer; 2002.
44. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974;19(6):716-723.
45. Schwarz G. Estimating the dimension of a model. *Annals of Statistics*. 1978;6(2):461-464.
46. Hurvich CM, Tsai CL. Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*. 1991;78(3):499-509.
47. Pinheiro JC, Bates D. *Mixed-Effects Models in S and S-PLUS*. Springer; 2000.
48. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer; 2000.
49. Gurka MJ. Selecting the best linear mixed model under REML. *The American Statistician*. 2006;60(1):19-26.
50. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021.
51. Gay DM. *Usage Summary for Selected Optimization Routines*. Technical Report. AT&T Bell Laboratories; 1990.
52. Nelder JA, Mead R. A simplex method for function minimization. *The Computer Journal*. 1965;7(4):308-313.
53. Lange K. *Numerical Analysis for Statisticians*. Springer; 1999.
54. Gilbert P, Varadhan R. numDeriv: Accurate numerical derivatives. 2016. R package version 2016.8-1.
55. Bangert-Drowns RL, Hurley MM, Wilkinson B. The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Review of Educational Research*. 2004;74(1):29-58.
56. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Academic Press; 1985.
57. Jackson D. Confidence intervals for the between-study variance in random effects meta-analysis using generalised Cochran heterogeneity statistics. *Res Synth Methods*. 2013;4(3):220-229.
58. Viechtbauer W. Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift für Psychologie / Journal of Psychology*. 2007;215(2):104-121.
59. Cooper HM. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. 5th ed. Sage; 2017.
60. 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.
61. Nakagawa S, Poulin R, Mengersen K, et al. Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*. 2015;6(2):143-152.
62. Prendergast LA, Staudte RG. Meta-analysis of ratios of sample variances. *Stat Med*. 2016;35(11):1780-1799.
63. Senior AM, Viechtbauer W, Nakagawa S. Revisiting and expanding the meta-analysis of variation: the log coefficient of variation ratio. *Res Synth Methods*. 2020;11(4):553-567.
64. Alba AC, Alexander PE, Chang J, MacIsaac J, DeFry S, Guyatt GH. High statistical heterogeneity is more frequent in meta-analysis of continuous than binary outcomes. *J Clin Epidemiol*. 2016;70:129-135.
65. Rhodes KM, Turner RM, Higgins JP. Predictive distributionswere developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *J Clin Epidemiol*. 2015;68(1):52-60.
66. Rhodes KM, Turner RM, Higgins JP. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. *Res Synth Methods*. 2016;7(4):346-370.
67. Senior AM, Grueber CE, Kamiya T, et al. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology*. 2016;97(12):3293-3299.
68. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JP. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Stat Med*. 2015;34(6):984-998.
69. IntHout J, Ioannidis JPA, Borm GF, Goeman JJ. Small studies are more heterogeneous than large ones: a meta-metaanalysis. *J Clin Epidemiol*. 2015;68(8):860-869.
70. Kalaian HA, Raudenbush SW. A multivariate mixed linear model for meta-analysis. *Psychol Methods*. 1996;1(3):227-235.
71. 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.
72. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. 2nd ed. Russell Sage Foundation; 2009:357-376.
73. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. *Res Synth Methods*. 2011;2(1):61-76.

How to cite this article: Viechtbauer W, López-López JA. Location-scale models for meta-analysis. *Res Syn Meth*. 2022;13(6):697-715. doi:10.1002/jrsm.1562