Behavior Research Methods (2023) 55:2367–2386
https://doi.org/10.3758/s13428-022-01892-7

Multilevel multivariate meta-analysis made easy: An introduction to MLMVmeta

Blakeley B. McShane[1] · **Ulf Böckenholt**[1]

Accepted: 12 May 2022 / Published online: 1 August 2022

Abstract

The basic random effects meta-analytic model is overwhelmingly dominant in psychological research. Indeed, it is typically employed even when more complex multilevel multivariate meta-analytic models are warranted. In this paper, we aim to help overcome challenges so that multilevel multivariate meta-analytic models will be more often employed in practice. We do so by introducing MLMVmeta—an easy-to-use web application that implements multilevel multivariate meta-analytic methodology that is both specially tailored to contemporary psychological research and easily estimable, interpretable, and parsimonious—and illustrating it across three case studies. The three case studies demonstrate the more accurate and extensive results that can be obtained via multilevel multivariate meta-analytic models. Further, they sequentially build in complexity featuring increasing numbers of experimental factors and conditions, dependent variables, and levels; this in turn necessitates increasingly complex model specifications that also sequentially build upon one another.

Keywords Multilevel · Multivariate · Meta-analysis

Introduction

The basic random effects meta-analytic model is overwhelmingly dominant in psychological research. However, this univariate, two-level model is suitable only when studies are independent and adequately summarized by a single statistic. This is seldom the case in contemporary psychological research, and when it is not, the model can be problematic and more accurate and extensive results can be obtained via multilevel multivariate meta-analytic models.

Nonetheless, the basic random effects meta-analytic model is still typically employed even when these more complex models are warranted (Tipton, Pustejovsky, & Ahmadi, 2019; McShane & Böckenholt, 2020). This is perhaps curious as multilevel multivariate meta-analytic models are by no means new. Indeed, they were introduced and applied in noted research articles (Kalaian & Raudenbush, 1996; Berkey, Hoaglin, Antczak-Bouckoms, Mosteller, & Colditz, 1998) and have been covered in classic textbooks (Raudenbush & Bryk, 2002; Cheung, 2015) and handbooks (Becker, 2000; Cooper, Hedges, & Valentine J.C., 2019; Schmid, Stijnen, & White, 2020).

One reason multilevel multivariate meta-analytic models may not often be employed in practice is that they can be considerably more difficult to implement as compared to the basic random effects meta-analytic model.

A second reason may be that standard multilevel multivariate meta-analytic models are not fully suited to contemporary psychological research in which studies in a given domain can vary considerably in terms of their experimental factors and dependent variables; examine multiple conditions that result from the manipulation of those experimental factors and give rise to multiple dependent effects of interest (e.g., simple effects, main effects, and interaction effects); employ a mix of study designs (e.g., unmoderated versus moderated, between-subjects versus within-subjects, univariate versus multivariate); and vary with respect to their operationalizations of the experimental factors and dependent variables, social contexts, and other method factors. Further, papers can feature multiple studies that, while different, are quite similar—particularly in comparison to studies featured in other papers. Such studies require careful treatment in meta-analysis so that the variation and covariation induced

✉ Blakeley B. McShane
b-mcshane@kellogg.northwestern.edu

Ulf Böckenholt
u-bockenholt@kellogg.northwestern.edu

1 Kellogg School of Management, Northwestern University, Evanston, IL, USA

by the facts that observations differ in their experimental factors and dependent variables and are nested (e.g., individual-level data within experimental conditions within groups of subjects within studies within papers) are accounted for.

A third reason may be that the number of observations is typically not large in contemporary psychological research, especially relative to the potential complexity of the variation and covariation. Consequently, it is desirable to consider multilevel multivariate meta-analytic model specifications that are not only easily estimable and interpretable (as more generally) but also parsimonious.

In this paper, we aim to help overcome these challenges so that multilevel multivariate meta-analytic models will be more often employed in practice. We do so by introducing MLMVmeta—an easy-to-use web application that implements multilevel multivariate meta-analytic methodology that is both specially tailored to contemporary psychological research and easily estimable, interpretable, and parsimonious—and illustrating it across three case studies.

MLMVmeta is available at https://blakemcshane.shinyapps.io/mlmvmeta/. It implements the novel multilevel multivariate meta-analytic methodology introduced by McShane and Böckenholt (2018) in an easy-to-use manner. This methodology generalizes prior work in three important respects, namely to simultaneously accommodate (i) an arbitrary number of experimental conditions that result from the manipulation of experimental factors and give rise to multiple dependent effects of interest; (ii) an arbitrary number of dependent variables; and (iii) an arbitrary number of levels of nesting. Therefore, it is well suited to contemporary psychological research because it can account for the variation and covariation induced by the facts that observations differ in their experimental factors and dependent variables and are nested. Nonetheless, the particular multilevel multivariate meta-analytic model specifications implemented by MLMVmeta—which include those of McShane and Böckenholt (2018) as well as several additional novel but related ones—remain easily estimable, interpretable, and parsimonious.

The three case studies feature hypothetical data typical of contemporary psychological research and demonstrate the more accurate and extensive results that can be obtained via multilevel multivariate meta-analytic models. Further, they sequentially build in complexity featuring increasing numbers of experimental factors and conditions, dependent variables, and levels; this in turn necessitates increasingly complex model specifications that also sequentially build upon one another.

Specifically, the first case study features the data on twin correlations given in Table 1 and used in studies of human trait heritability (see, for example, Polderman et al., (2015)). In this data, there are two dependent variables (i.e., dizygotic and monozygotic correlations) and two levels (i.e., individual-level data nested within studies).

The second case study features the data on SAT coaching given in Table 5 and used in studies of educational evaluation (see, for example, Kalaian and Raudenbush (1996)). In this data, there are two experimental conditions (i.e., control and treatment) arising from the experimental manipulation of a single experimental factor (i.e., provision of coaching), two dependent variables (i.e., math and verbal scores), and three levels (i.e., individual-level data nested within groups of subjects nested within schools). In addition, there is a continuous covariate for the treatment condition (i.e., hours of coaching).

The third case study features the data on choice overload given in Table 18 and used in studies of consumer psychology (see, for example, Chernev, Böckenholt, and Goodman (2015)). In this data, there are ten experimental conditions arising from the manipulation of three experimental factors (assortment size, decision goal, and decision task difficulty), three dependent variables (i.e., confidence, regret, and satisfaction), and five levels (i.e., individual-level data nested within experimental conditions nested within groups of subjects nested within studies nested within papers).

Although the case studies illustrate experimental research, the multilevel multivariate meta-analytic methodology implemented by MLMVmeta is fully general and thus also accommodates observational research. Specifically, one can simply substitute the notion of an experimental condition that results from the manipulation of one or more experimental factors with the notion of a group of subjects that results from the variation of individual-level covariates (or omit it entirely if only a single group of subjects is of interest).

The remainder of this paper is organized as follows. We next review the basic random effects meta-analytic model specification. We then illustrate multilevel multivariate meta-analytic models as compared to the basic random effects meta-analytic model via the bivariate, two-level case study on twin correlations and the bivariate, three-level case study on SAT coaching. With this background, we introduce the multilevel multivariate meta-analytic model specification in full generality and then illustrate it as compared to the basic random effects meta-analytic model via the trivariate, five-level case study on choice overload. We conclude with a brief discussion and provide an appendix that details how to reproduce the results reported in the case studies using MLMVmeta, how to analyze other data using MLMVmeta, and other details about MLMVmeta.

Basic random effects model specification

The basic random effects model specification is given by

$$y_i = \alpha + \beta_i + \varepsilon_i$$

where i indexes the observations (e.g., studies); the y_i are single statistics that summarize the individual-level data associated with each observation; α is treated as a fixed effect that models the overall average of the observations; the β_i are treated as random effects for each observation; and the ε_i are random (i.e., sampling) errors for each observation.

The model further assumes that the β_i are independently and identically distributed normal with mean zero and variance τ^2; the ε_i are independently distributed normal with mean zero and variance v_i; and the β_i and ε_j are independent. Thus, the v_i give the level one variances that model sampling variation and τ^2 gives the level two variance that models variation among the β_i. As is standard in meta-analysis, the v_i are assumed known, and the goal is to estimate the fixed effect parameter α and the variance component parameter τ^2.

This model is quite limited in that it assumes that the observations are independent and have both a common average and a common degree of (non-sampling) variation. It may be suitable when studies are independent and adequately summarized by a single statistic. However, this is seldom the case in contemporary psychological research (e.g., it may be the case when papers feature only a single study and studies feature only a single experimental condition and a single dependent variable and interest centers on only a single statistic summarizing the individual-level data). Further, standard generalizations to the model (e.g., accommodating covariates by replacing α with observation-specific α_i, which are specified as $\alpha_i = \alpha_0 + \sum_{p=1}^{P} \alpha_p x_{i,p}$ where $x_{i,p}$ is the p^{th} covariate for observation i and P is the number of covariates) do not overcome these limitations. Therefore, the model can be problematic and more accurate and extensive results can be obtained via multilevel multivariate meta-analytic models as the case studies demonstrate.

Table 1 Twin correlation data

Dependent variable	Study ID	y	v
Dizygotic	1	0.3031	0.0042
Monozygotic	1	0.5428	0.0047
Dizygotic	2	0.0972	0.0045
Monozygotic	2	0.6115	0.0040
Dizygotic	3	0.3405	0.0027
Monozygotic	3	0.6912	0.0027
Dizygotic	4	0.2852	0.0072
Monozygotic	4	0.5366	0.0065
Dizygotic	5	0.2331	0.0031
Monozygotic	5	0.7522	0.0029
Dizygotic	6	0.3118	0.0026
Monozygotic	6	0.6045	0.0029
Dizygotic	7	0.5644	0.0028
Monozygotic	7	0.9238	0.0028
Dizygotic	8	0.4157	0.0076
Monozygotic	8	0.8519	0.0069
Dizygotic	9	0.5748	0.0038
Monozygotic	9	0.5559	0.0033
Dizygotic	10	0.3201	0.0072
Monozygotic	10	0.5560	0.0083
Dizygotic	11	0.4638	0.0103
Monozygotic	11	0.8653	0.0106
Dizygotic	12	0.2178	0.0031
Monozygotic	12	0.6034	0.0036
Dizygotic	13	0.3370	0.0031
Monozygotic	13	0.5777	0.0032
Dizygotic	14	0.4939	0.0035
Monozygotic	14	0.7987	0.0033
Dizygotic	15	0.1510	0.0028
Monozygotic	15	0.5111	0.0031

Case study 1: Twin correlations

Studies of the correlations of pairs of dizygotic and monozygotic twins on a given trait are an important source of data for investigating the heritability of the trait in question. In Table 1, we present hypothetical data typical of such studies. Each study reports an estimate of the correlation (labeled y in the table) of pairs of dizygotic and monozygotic twins on a given trait and an estimate of the sampling variance of the estimate of the correlation (labeled v in the table).

The multilevel multivariate meta-analytic model specification for this data is given by

$$y_i = \alpha_{d_i} + \beta_i + \varepsilon_i$$

where i indexes the observations in Table 1; the α_d are treated as fixed effects that model the overall average for each dependent variable; the d_i denote the dependent variable $d \in \{1,2\}$ (here denoting dizygotic and monozygotic, respectively) which each observation measures; the β_i are treated as random effects for each observation; and the ε_i are random errors for each observation.

The model further assumes that the pairs of β_i for each study are independently and identically distributed bivariate normal with mean zero and variance-covariance matrix specified according to the multilevel multivariate compound symmetry (MMCS) model specification with variance component parameters $\tau^2_{2,d}$ and $\rho_{2,d,d'}$ to be discussed below. It also assumes that the ε_i are independently distributed normal with mean zero and variance v_i and that the β_i and ε_j are independent. As is standard in meta-analysis, the v_i are assumed known, and the goal is to estimate the fixed effect parameters α_d and the variance component parameters $\tau^2_{2,d}$ and $\rho_{2,d,d'}$. Here and hereafter, $\tau^2_{k,d}$ denote variance parameters and $\rho_{k,d,d'}$ denote correlation parameters.

Table 2 Twin correlation MMCS model performance

Model specification	Variance component parameters	REML LL	AIC
Fixed effects	0	-19.84	39.68
Equal variance, zero correlation	1	15.43	-28.85
Unequal variance, zero correlation	2	15.46	-26.91
Unconstrained	3	17.74	-29.48

The MMCS model specification implies that

$$\mathrm{var}(y_i) = \mathrm{var}(\beta_i) + \mathrm{var}(\varepsilon_i) = \tau^2_{2,d_i} + v_i$$

and

$$\mathrm{cov}(y_i, y_j) = \mathrm{cov}(\beta_i, \beta_j) + \mathrm{cov}(\varepsilon_i, \varepsilon_j) = \rho_{2,d_i,d_j}\tau_{2,d_i}\tau_{2,d_j}m_{2,i,j} + 0$$

where the $\tau^2_{2,d}$ and $\rho_{2,d,d'}$ respectively give the variances and correlations that model the variation and covariation among the β_i induced by level 2 of the nesting structure (here level 2 denotes the study); $m_{2,i,j}$ is one if observations i and j are nested in the same group at level 2 (i.e., are from the same study) and zero otherwise; the v_i are the assumed known sampling variances of observations; and cov(ε_i,ε_j) is zero because the ε_i and ε_j are independent for $i \neq j$ (i.e., because the observations were of different pairs of twins).

In addition to this unconstrained model specification, we consider a sequence of nested simplifications: a fixed effects specification that sets the $\tau^2_{2,d} = 0$ for both d such that $\rho_{2,d,d'}$ is irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_{2,d} = \tau^2_2$ for both d and $\rho_{2,d,d'} = 0$ for $d \neq d'$; and an unequal variance, zero correlation specification that sets $\rho_{2,d,d'} = 0$ for $d \neq d'$. We consider these various simplifications for reasons of estimability, interpretability, and parsimony; while these issues are not critical in this case study because the unconstrained specification has only three variance component parameters (i.e., $\tau^2_{2,1}$, $\tau^2_{2,2}$, and $\rho_{2,1,2}$), in other case studies and applications, these simplifications can be highly important.

We present the number of variance component parameters estimated by, the restricted maximum likelihood log likelihood (REML LL) of, and the Akaike information criterion (AIC) of each of these model specifications in Table 2. As can be seen, the unconstrained model specification performs best in terms of AIC. Consequently, we present estimates from it in Table 3. The estimates indicate that monozygotic twins have a larger trait correlation than dizygotic twins (i.e., 0.664 versus 0.339), a comparison to which we return at the end of this case study. In addition, they indicate that the variation among the β_i is comparable for both types of twins (i.e., 0.017 ≈ 0.014). Finally, they indicate that there is substantial correlation (i.e., 0.664) between the β_i for the two types of twins at the study level (i.e., studies with low (high) trait correlations for dizygotic twins tend to have low (high) trait correlations for monozygotic twins).

Table 3 Twin correlation unconstrained MMCS model estimates

Dependent variable	α		τ^2	ρ
	Estimate	Std. error	Estimate	Estimate
Dizygotic	0.339	0.037	0.017	0.664
Monozygotic	0.664	0.035	0.014	

Table 4 Twin correlation basic random effects model estimates

Model	Dependent variable	α		τ^2
		Estimate	Std. error	Estimate
Model 1	Not applicable	0.502	0.039	0.042
Model 2	Dizygotic	0.339	0.036	0.015
	Monozygotic	0.664	0.036	

In addition to considering multilevel multivariate meta-analytic model specifications, we also present estimates from the basic random effects model in Table 4. In particular, we present estimates from two variants of this model: Model 1 is the basic random effects model with a single fixed effect parameter α and a single variance component parameter τ^2 while Model 2 generalizes Model 1 to allow for fixed effect parameters that accommodate differences between dizygotic and monozygotic twins. We note that REML LL and AIC are comparable only across models that include the same fixed effect parameters; consequently, only Model 2 can be compared to the model specifications presented in Table 2. However, because in this case study there are only two levels in the nesting structure and the random errors for each observation are independent (i.e., because the observations were of different pairs of twins), Model 2 is equivalent to the equal variance, zero correlation multilevel multivariate meta-analytic model specification. Therefore, it performs worse in terms of AIC as compared to the multilevel multivariate meta-analytic model specification discussed above.

From the perspective of theory, these basic random effects model variants are inappropriate because they ignore differences between dependent variables. Further, from the perspective of the data structure, they are inappropriate because they ignore the fact that observations are nested within studies and treat all observations as independent. Both of these have important implications for the estimates obtained from these models.

First, the estimate of the fixed effect from Model 1 is rather nonsensical and uninterpretable because it ignores differences between dizygotic and monozygotic twin correlations. As a consequence, the estimate of the variance component is inflated relative to those of the multilevel multivariate meta-analytic model.

Second, while the estimates of the fixed effects and the variance component from Model 2 are reasonable and match those of the multilevel multivariate meta-analytic model, this is direct consequence of the fact that the multilevel multivariate meta-analytic model estimates of the variance components are similar for both types of twins (i.e., $0.017 \approx 0.014$). Had they differed—which they may well have empirically—the estimates of the fixed effects and the variance component from Model 2 would not necessarily have been reasonable. Further, because the basic random effects model and its variants assume independence across observations, the estimates of the fixed effects are necessarily assumed independent; however, the multilevel multivariate meta-analytic model estimates indicate substantial correlation (i.e., 0.664) between the β_i at the study level which results in a substantial and heretofore unmentioned estimate of the correlation (i.e, 0.524) between the estimates of the fixed effects.

Consequently, these basic random effects model variants are unsuitable for conducting statistical inference (e.g., confidence interval (CI) estimation, null hypothesis significance testing) on linear combinations—in particular, in this case study, the difference between monozygotic and dizygotic twin correlations, $\alpha_{\text{Monozygotic}} - \alpha_{\text{Dizygotic}}$—of the fixed effect parameters[1]. Therefore, we conduct statistical inference for this difference for only the multilevel multivariate meta-analytic model. The estimate of this difference is $0.664 - 0.339 = 0.325$ and the estimate of the standard error of this estimate of this difference is

$$\sqrt{0.037^2 + 0.035^2 - 2 \cdot 0.524 \cdot 0.037 \cdot 0.035} = 0.035$$

where these follow from the estimates of the fixed effects and the estimates of the standard errors of the fixed effect estimates in Table 3 as well as the estimate of the correlation of the fixed effect estimates discussed above. Therefore, a 95% CI estimate of this difference is $0.325 \pm 1.960 \cdot 0.035 = [0.255, 0.394]$, and the z-statistic and p-value against the point null hypothesis of zero difference are $0.325/0.035 = 9.214$ and less than 0.001, respectively.

Case study 2: SAT coaching

Studies of the effect of coaching on multiple measures of test performance are an important source of data for evaluating educational interventions. In Table 5, we present hypothetical data typical of such studies. Each study reports an estimate of the mean performance (labeled y in the table) on the math and verbal sections of the SAT of a group of control students provided no coaching and a group of treated students provided coaching in a given school and an estimate of the sampling variance of the estimate of the mean performance (labeled v in the table). The table also reports the number of hours of coaching provided to the groups of students. In Table 6, we present the estimate of the sampling variance-covariance matrix of the estimates of the mean performance (of which v forms the diagonal) which gives the level one variances and covariances that model sampling variation and covariation, respectively; it follows a 2×2 block diagonal structure because sampling covariances are zero when observations are of different groups of students.

In the absence of the continuous covariate (i.e., hours of coaching), the multilevel multivariate meta-analytic model specification for this data is given by

$$y_i = \alpha_{c_i,d_i} + \beta_i + \varepsilon_i$$

where i indexes the observations in Table 5; the $\alpha_{c,d}$ are treated as fixed effects that model the overall average for each experimental condition and dependent variable; the c_i denote the experimental condition $c \in \{1,2\}$ (here denoting control and treatment, respectively) to which each observation is assigned; the d_i denote the dependent variable $d \in \{1,2\}$ (here denoting math and verbal, respectively) which each observation measures; the β_i are treated as random effects for each observation; and the ε_i are random errors for each observation.

To account for the number of hours of coaching provided to the treated group of students, the model replaces the $\alpha_{c,d}$ with observation-specific α_i which are specified as

[1] To conduct statistical inference for linear combinations of fixed effect parameters in general, we let **a** be a column vector containing the estimates of the fixed effects (e.g., those in Table 3), **S** be the estimate of the variance-covariance matrix of **a** (available on MLMVmeta), and **L** be the matrix whose rows are the linear combination vectors (e.g., $(-1\ \ 1)$ for the difference between monozygotic and dizygotic twin correlations), and use Wald standard errors for inference. Specifically, estimates of the linear combinations are given by **La** and estimates of the standard errors of these estimates are given by the square root of the diagonal of $\mathbf{LSL}^{\mathrm{T}}$. CI estimates, z-statistics, and p-values follow from these. Statistical inference can be performed given **a**, **S**, **L**, and optionally a vector of point null hypothesis values against which to test using an easy-to-use web application available at https://blakemcshane.shinyapps.io/lcwald/.

Table 5 SAT coaching data. The full data are available on MLMVmeta

Condition description	Dependent variable	School ID	Group ID	y	v	Hours
Control	Math	1	1	514.5	156.8	0
Control	Verbal	1	1	533.8	176.3	0
Treatment	Math	1	2	540.3	292.1	7.5
Treatment	Verbal	1	2	540.8	224.7	7.5
Control	Math	2	3	473.2	160.0	0
Control	Verbal	2	3	497.8	154.4	0
Treatment	Math	2	4	487.9	191.2	5.0
Treatment	Verbal	2	4	509.3	143.6	5.0
Control	Math	3	5	497.0	55.2	0
Control	Verbal	3	5	493.5	53.5	0
Treatment	Math	3	6	522.3	50.3	5.0
Treatment	Verbal	3	6	514.0	44.2	5.0
Control	Math	4	7	504.3	40.4	0
Control	Verbal	4	7	503.5	43.3	0
Treatment	Math	4	8	517.1	38.8	7.5
Treatment	Verbal	4	8	517.1	37.1	7.5
Control	Math	5	9	520.3	91.5	0
Control	Verbal	5	9	496.9	101.5	0
Treatment	Math	5	10	551.2	85.5	20.0
Treatment	Verbal	5	10	539.3	82.5	20.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 6 SAT coaching sampling variance-covariance matrix. The full matrix is available on MLMVmeta

156.8	87.4	0	0	0	0	0	0	…
87.4	176.3	0	0	0	0	0	0	…
0	0	292.1	189.4	0	0	0	0	…
0	0	189.4	224.7	0	0	0	0	…
0	0	0	0	160.0	66.0	0	0	…
0	0	0	0	66.0	154.4	0	0	…
0	0	0	0	0	0	191.2	91.2	…
0	0	0	0	0	0	91.2	143.6	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

$$\alpha_i = \alpha_{0,c_i,d_i} + \alpha_{1,c_i,d_i}\text{Hours}_i.$$

Because Hours_i is by definition zero for observations which are assigned to the control condition, we set the $\alpha_{1,1,d}$ to zero.

The model further assumes that the quadruplets of β_i for each school are independently and identically distributed quadvariate normal with mean zero and variance-covariance matrix specified according to the MMCS model specification with variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ to be discussed below. It also assumes that the pairs of ϵ_i for each group of students are independently distributed bivariate normal with mean zero and variance-covariance matrix given by the relevant entries of the estimate of the sampling variance-covariance matrix discussed above and that the β_i and ϵ_j are independent. As is standard in meta-analysis, the estimate of the sampling variance-covariance matrix is assumed known, and the goal is to estimate the fixed effect parameters $\alpha_{p,c,d}$ and the variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$.

The MMCS model specification implies that

$$\text{var}(y_i) = \text{var}(\beta_i) + \text{var}(\epsilon_i) = (\tau^2_{2,d_i} + \tau^2_{3,d_i}) + v_i$$

and

$$\begin{aligned}\text{cov}(y_i, y_j) &= \text{cov}(\beta_i, \beta_j) + \text{cov}(\epsilon_i, \epsilon_j) \\ &= (\rho_{2,d_i,d_j}\tau_{2,d_i}\tau_{2,d_j}m_{2,i,j} + \rho_{3,d_i,d_j}\tau_{3,d_i}\tau_{3,d_j}m_{3,i,j}) + v_{i,j}\end{aligned}$$

where the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ respectively give the variances and correlations that model the variation and covariation among

Table 7 SAT coaching MMCS and EAMMCS model performance

Model specification	Variance component parameters	REML LL	AIC
MMCS			
Fixed effects	0	-835.69	1671.39
Equal variance, zero correlation	2	-757.06	1518.12
Unequal variance, zero correlation	4	-749.58	1507.16
Unconstrained	6	-746.03	1504.07
EAMMCS			
Fixed effects	0	-835.69	1671.39
Equal variance, zero correlation	2	-757.06	1518.12
Unequal variance, zero correlation	3	-749.62	1505.24
Unconstrained	4	-746.31	1500.62

the β_i induced by level k of the nesting structure (here level 2 denotes the group of students and level 3 denotes the school); $m_{k,i,j}$ is one if observations i and j are nested in the same group at level k and zero otherwise; and the v_i and $v_{i,j}$ are respectively the assumed known sampling variances and covariances of the observations.

In addition to the MMCS model specification, we also consider a special case of MMCS termed equal allocation multilevel multivariate compound symmetry (EAMMCS) that constrains the fractional allocation of the variation and covariation induced by the nesting structure to be equal across all dependent variable pairs. In particular, if we let $\omega_{k,d,d'} = \rho_{k,d,d'}\tau_{k,d}\tau_{k,d'}$, $\omega_{d,d'} = \sum_{k=2}^{K}\omega_{k,d,d'}$, and $\pi_{k,d,d'} = \omega_{k,d,d'}/\omega_{d,d'}$ where K is the number of levels in the nesting structure (here K is 3), EAMMCS holds when we constrain the $\pi_{k,d,d'}$ (i.e., the fractional allocation of the covariation among dependent variables d and d' induced by level k of the nesting structure) to be equal to π_k for all $d, d' \in \{1, ..., D\}$ where D is the number of dependent variables (here D is 2). Under the EAMMCS specification, the original MMCS variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ where $k \in\{2,...,K\}$ and $d, d' \in \{1, ..., D\}$ are parameterized by the EAMMCS variance component parameters τ^2_d, $\rho_{d,d'}$, and π_k where the π_k are constrained to be between zero and one and to sum to one. Specifically, $\tau^2_{k,d} = \pi_k\tau^2_d$, which can be seen by noting that $\omega_{k,d,d} = \tau^2_{k,d}$ and $\omega_{d,d} = \tau^2_d$. Further, $\rho_{k,d,d'} = \rho_{d,d'}$, which can be seen by noting that

$$\rho_{k,d,d'} = \frac{\omega_{k,d,d'}}{\tau_{k,d}\tau_{k,d'}} = \frac{\pi_k\omega_{d,d'}}{\sqrt{\pi_k\tau^2_d}\sqrt{\pi_k\tau^2_{d'}}} = \frac{\omega_{d,d'}}{\tau_d\tau_{d'}}$$

does not depend on k. We consider the EAMMCS model specification for reasons of estimability, interpretability, and parsimony.

In addition to these unconstrained MMCS and EAMMCS model specifications, we also consider a sequence of nested simplifications. For MMCS, we consider a fixed effects specification that sets the $\tau^2_{k,d} = 0$ for both k and both d such that the $\rho_{k,d,d'}$ are irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_{k,d} = \tau^2_k$ for both d and the $\rho_{k,d,d'} = 0$ for $d \neq d'$; and an unequal variance, zero correlation specification that sets the $\rho_{k,d,d'} = 0$ for $d \neq d'$. For EAMMCS, we consider a fixed effects specification that sets the $\tau^2_d = 0$ for both d such that $\rho_{d,d'}$ and the π_k are irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_d = \tau^2$ for both d and $\rho_{d,d'} = 0$ for $d \neq d'$; and an unequal variance, zero correlation specification that sets $\rho_{d,d'} = 0$ for $d \neq d'$. The appendix details all model specifications implemented by MLMVmeta and cases in which they are equivalent.

We present the number of variance component parameters estimated by, the REML LL of, and the AIC of each of these model specifications in Table 7. As can be seen, the unconstrained EAMMCS model specification performs best in terms of AIC. Consequently, we present estimates from it in Tables 8 and 9. The estimates indicate that coaching results in improved test performance relative to no coaching, a comparison to which we return at the end of this case study. In addition, they indicate that there is much greater variation among the β_i for math scores than for verbal scores (i.e., 168.31 versus 39.03). Further, they indicate that there is much greater variation and covariation among the β_i at the school level than at the group level (i.e., 0.73 versus 0.27). Finally, they indicate that there is substantial correlation (i.e., 0.53) between the β_i for the two scores at the school level and the group level.

Given the estimates of the EAMMCS variance component parameters τ^2_d, $\rho_{d,d'}$, and π_k presented in Table 9, it is trivial to obtain estimates of the MMCS variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$. In particular, one can estimate the $\tau^2_{k,d}$ by multiplying the estimates of π_k and τ^2_d; for example, the estimate of the Level 3 math dependent variable variance

Table 8 SAT coaching unconstrained EAMMCS model fixed effect estimates

Dependent variable & condition description	Estimate	Standard error
Math; control	500.04	2.20
Math; treatment	499.30	4.62
Math; treatment × hours	2.20	0.38
Verbal; control	501.94	1.48
Verbal; treatment	500.13	3.40
Verbal; treatment × hours	1.65	0.28

Table 9 SAT coaching unconstrained EAMMCS model variance component estimates

Dep. variable	τ^2 Estimate	ρ Estimate	Level	π Estimate
Math	168.31	0.53	School	0.73
Verbal	39.03		Group	0.27

component is given by $0.73 \cdot 168.31 \approx 123.10$. Further, one can estimate the $\rho_{k,d,d'}$ by the $\rho_{d,d'}$; for example, the estimate of the Level 3 correlation variance component is given by 0.53. These estimates are presented in Table 10.

In addition to considering multilevel multivariate meta-analytic model specifications, we also present estimates from the basic random effects model in Table 11. In particular, we present estimates from three variants of this model: Model 1 is the basic random effects model with a single fixed effect parameter α and a single variance component parameter τ^2, Model 2 generalizes Model 1 to allow for fixed effect parameters that accommodate differences between the experimental conditions, and Model 3 generalizes Model 2 to further allow for fixed effect parameters that accommodate differences between the dependent variables. We note that REML LL and AIC are comparable only across models that include the same fixed effect parameters; consequently, only Model 3 can be compared to the model specifications presented in Table 7. The REML LL and AIC of Model 3 are -780.43 and 1562.85, respectively. Therefore, it performs worse in terms of AIC as compared to the multilevel multivariate meta-analytic model specification discussed above.

From the perspective of theory, these basic random effects model variants are inappropriate because they ignore differences between experimental conditions and dependent variables. Further, from the perspective of the data structure, they are inappropriate because they ignore the fact that observations nested within groups of subjects and schools and treat all observations as independent. Both of these have important implications for the estimates obtained from these models.

Table 10 SAT coaching unconstrained EAMMCS model variance component estimates presented as MMCS variance component estimates

Dep. variable	School (Level 3)		Group (Level 2)	
	τ^2	ρ	τ^2	ρ
	Estimate	Estimate	Estimate	Estimate
Math	123.10	0.53	45.21	0.53
Verbal	28.54		10.48	

First, the estimate of the fixed effect from Model 1 is rather nonsensical and uninterpretable because it ignores differences between experimental conditions and dependent variables. As a consequence, the estimate of the variance component is inflated relative to those of the multilevel multivariate meta-analytic model.

Second, the estimates of the fixed effects from both Model 2 and Model 3 are reasonable and are not dissimilar to those of the multilevel multivariate meta-analytic model. However, this fact only holds for Model 2 as a direct consequence of the fact that the estimates of the fixed effects are similar for both dependent variables. Had they differed—which they may well have empirically—the estimates of the fixed effects from Model 2 would not necessarily have been reasonable. Further, because the basic random effects model has only a single variance component parameter and thus assumes that there are only two levels in the nesting structure and equality across dependent variables, the Model 2 and Model 3 estimates of the variance component (i.e., 102.64 and 100.14, respectively) are rather nonsensical and uninterpretable because they ignore differences between dependent variables; indeed, the multilevel multivariate meta-analytic model estimates indicate substantial differences between the dependent variables (i.e., 168.31 for math and 39.03 for verbal). Finally, Model 2 and Model 3 assume that none of the variation is at the school level, all

Table 11 SAT coaching basic random effects model estimates

Model	Dependent variable & condition description	α		τ^2
		Estimate	Std. err.	Estimate
Model 1	Not applicable	510.82	1.28	247.30
Model 2	Control	500.99	1.33	102.64
	Treatment	500.38	3.18	
	Treatment × hours	1.90	0.27	
Model 3	Math; control	500.12	1.87	100.14
	Math; treatment	499.46	4.48	
	Math; treatment × hours	2.21	0.38	
	Verbal; control	501.86	1.87	
	Verbal; treatment	501.34	4.44	
	Verbal; treatment × hours	1.57	0.37	

Table 12 SAT coaching unconstrained EAMMCS model effect estimates

Dependent variable & hours	Estimate [95% CI]	Standard error	z-statistic	p-value
Math; 5 hours	10.26 [4.22, 16.31]	3.08	3.33	< 0.001
Math; 10 hours	21.26 [16.96, 25.57]	2.20	9.68	< 0.001
Math; 20 hours	43.27 [35.26, 51.27]	4.08	10.59	< 0.001
Verbal; 5 hours	6.46 [1.70, 11.21]	2.43	2.66	0.008
Verbal; 10 hours	14.73 [11.21, 18.25]	1.80	8.20	< 0.001
Verbal; 20 hours	31.28 [25.13, 37.43]	3.14	9.97	< 0.001

is at the group level, and there is no covariation between the two dependent variables at any level; on the other hand, the multilevel multivariate meta-analytic model estimates indicate that 73% of the variation is at the school level, 27% is at the group level, and the covariation is substantial (i.e., correlated at 0.53).

Consequently, these basic random effects model variants are unsuitable for conducting statistical inference on the fixed effect parameters or linear combinations of them (i.e., because statistical inference depends on the estimate of the variance-covariance matrix of the estimates of the fixed effects which in turn depends on the estimates of the variance components; here, for example, the estimates of the standard errors of the estimates of the fixed effects for Model 3 are inflated relative to those of the multilevel multivariate meta-analytic model for verbal and deflated for math and the estimates of the correlations of the estimates of the fixed effects for Model 3 are deflated relative to those of the multilevel multivariate meta-analytic model). Therefore, we conduct statistical inference for such linear combinations—in particular, the effect of various hours of coaching on math and verbal scores—for only the multilevel multivariate meta-analytic model. Specifically, we conduct statistical inference for the linear combinations

$$\alpha_{0,\text{Treatment,Math}} + \alpha_{1,\text{Treatment,Math}}\text{Hours} - \alpha_{0,\text{Control,Math}}$$

and

$$\alpha_{0,\text{Treatment,Verbal}} + \alpha_{1,\text{Treatment,Verbal}}\text{Hours} - \alpha_{0,\text{Control,Verbal}}.$$

Estimates, 95% CI estimates, estimates of standard errors of the estimates, and z-statistics and p-values against the point null hypothesis of zero are presented in Table 12. The estimates indicate that SAT coaching results in modest to moderate improvements in math and verbal scores depending on the number of hours of coaching.

Case study 2 revisited: SAT coaching standardized

The estimates presented in Case Study 2 are on the SAT score scale. However, it is common in meta-analysis to present estimates on a standardized scale such as the correlation scale as in Case Study 1 or the standardized mean difference (or Cohen's d) scale. To achieve this, the data is converted to the standardized scale prior to analysis such that the resulting estimates are on the standardized scale. For comparison purposes, we also present estimates on a standardized scale, specifically the standardized mean difference scale.

In Table 13, we present the data presented in Table 5 on the standardized mean difference scale. Each study reports an estimate of the standardized mean difference (labeled y in the table) in the performance on the math and verbal sections of the SAT of a group of treated students provided coaching and a group of control students provided no coaching in a given school and an estimate of the sampling variance of the estimate of the standardized mean difference (labeled v in the table). The table also reports the number of hours of coaching provided to the treated groups of students. On

Table 13 SAT coaching standardized data. The full data as well as the estimate of the sampling variance-covariance matrix are available on MLMVmeta

Dep. variable	School ID	y	v	Hours
Math	1	0.2509	0.0425	7.5
Verbal	1	0.0713	0.0417	7.5
Math	2	0.1556	0.0393	5
Verbal	2	0.1318	0.0391	5
Math	3	0.2466	0.0100	5
Verbal	3	0.2082	0.0101	5
Math	4	0.1293	0.0081	7.5
Verbal	4	0.1367	0.0081	7.5
Math	5	0.3254	0.0196	20
Verbal	5	0.4369	0.0195	20
⋮	⋮	⋮	⋮	⋮

Table 14 SAT coaching standardized MMCS model performance

Model specification	Variance component parameters	REML LL	AIC
Fixed effects	0	55.16	-110.31
Equal variance, zero correlation	1	58.49	-114.99
Unequal variance, zero correlation	2	59.81	-115.62
Unconstrained	3	61.22	-116.44

MLMVmeta, we present the estimate of the sampling variance-covariance matrix of the estimates of the standardized mean differences.

In the absence of the continuous covariate, the multilevel multivariate meta-analytic model specification for this data is given by

$$y_i = \alpha_{d_i} + \beta_i + \varepsilon_i$$

where i indexes the observations in Table 13; the α_d are treated as fixed effects that model the overall average for each dependent variable; the d_i denote the dependent variable $d \in \{1,2\}$ (here denoting math and verbal, respectively) which each observation measures; the β_i are treated as random effects for each observation; and the ε_i are random errors for each observation.

To account for the number of hours of coaching provided to the treated group of students, the model replaces the α_d with observation-specific α_i which are specified as

$$\alpha_i = \alpha_{0,d_i} + \alpha_{1,d_i}\text{Hours}_i.$$

The model further assumes that the pairs of β_i for each school are independently and identically distributed bivariate normal with mean zero and variance-covariance matrix specified according to the MMCS model specification with variance component parameters $\tau^2_{2,d}$ and $\rho_{2,d,d'}$ to be discussed below. It also assumes that the pairs of ε_i for each school are independently distributed bivariate normal with mean zero and variance-covariance matrix given by the relevant entries of the estimate of the sampling variance-covariance matrix discussed above and that the β_i and ε_j are independent. As is standard in meta-analysis, the estimate of the sampling variance-covariance matrix is assumed known, and the goal is to estimate the fixed effect parameters $\alpha_{p,d}$ and the variance component parameters $\tau^2_{2,d}$ and $\rho_{2,d,d'}$.

The MMCS model specification implies that

$$\text{var}(y_i) = \text{var}(\beta_i) + \text{var}(\varepsilon_i) = \tau^2_{2,d_i} + v_i$$

and

$$\text{cov}(y_i, y_j) = \text{cov}(\beta_i, \beta_j) + \text{cov}(\varepsilon_i, \varepsilon_j) = \rho_{2,d_i,d_j}\tau_{2,d_i}\tau_{2,d_j}m_{2,i,j} + v_{i,j}$$

where the $\tau^2_{2,d}$ and $\rho_{2,d,d'}$ respectively give the variances and correlations that model the variation and covariation among the β_i induced by level 2 of the nesting structure (here level 2 denotes the school); $m_{2,i,j}$ is one if observations i and j are nested in the same group at level 2 (i.e., are from the same school) and zero otherwise; and the v_i and $v_{i,j}$ are respectively the assumed known sampling variances and covariances of the observations. We note that this model specification is equivalent to that used in Case Study 1 with three exceptions: the dependent variables here are standardized mean differences in the performance on the math and verbal sections of the SAT rather than correlations of pairs of dizygotic and monozygotic twins as in Case Study 1, the model here accounts for the continuous covariate which was absent in Case Study 1, and the model here accounts for covariation within the pairs of ε_i for each school which was absent in Case Study 1.

In addition to this unconstrained model specification, we consider a sequence of nested simplifications: a fixed effects specification that sets the $\tau^2_{2,d} = 0$ for both d such that $\rho_{2,d,d'}$ is irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_{2,d} = \tau^2_2$ for both d and $\rho_{2,d,d'} = 0$ for $d \neq d'$; and an unequal variance, zero correlation specification that sets $\rho_{2,d,d'} = 0$ for $d \neq d'$.

We present the number of variance component parameters estimated by, the REML LL of, and the AIC of each of these model specifications in Table 14. As can be seen, the unconstrained model specification performs best in terms of AIC. Consequently, we present estimates from it in Table 15. The estimates indicate that coaching results in improved test performance relative to no coaching, a comparison to which we return at the end of this case study. In addition, they indicate that there is much greater variation among the β_i for math scores than for verbal scores (i.e., 0.011 versus

Table 15 SAT coaching standardized unconstrained MMCS model estimates

Dependent variable	α		τ^2	ρ
	Estimate	Std. error	Estimate	Estimate
Math	-0.021	0.054	0.011	0.768
Math × hours	0.024	0.005		
Verbal	-0.064	0.044	0.003	
Verbal × hours	0.021	0.004		

Table 16 SAT coaching standardized basic random effects model estimates

Model	Dependent variable	α		τ^2
		Estimate	Std. error	Estimate
Model 1	Not applicable	0.200	0.019	0.020
Model 2	Math	-0.026	0.049	0.007
	Math × hours	0.024	0.004	
	Verbal	-0.068	0.049	
	Verbal × hours	0.022	0.004	

0.003). Finally, they indicate that there is substantial correlation (i.e., 0.768) between the β_i for the two scores at the school level.

In addition to considering multilevel multivariate meta-analytic model specifications, we also present estimates from the basic random effects model in Table 16. In particular, we present estimates from two variants of this model: Model 1 is the basic random effects model with a single fixed effect parameter α and a single variance component parameter τ^2 while Model 2 generalizes Model 1 to allow for fixed effect parameters that accommodate differences between the dependent variables and the number of hours of coaching provided to the treated group of students. We note that REML LL and AIC are comparable only across models that include the same fixed effect parameters; consequently, only Model 2 can be compared to the model specifications presented in Table 14. The REML LL and AIC of Model 2 are 48.37 and -94.74, respectively. Therefore, it performs worse in terms of AIC as compared to the multilevel multivariate meta-analytic model specification discussed above.

From the perspective of theory, these basic random effects model variants are inappropriate because they ignore differences between dependent variables. Further, from the perspective of the data structure, they are inappropriate because they ignore the fact that observations are nested within schools and treat all observations as independent. Both of these have important implications for the estimates obtained from these models.

First, the estimate of the fixed effect from Model 1 is rather nonsensical and uninterpretable because it ignores differences between dependent variables and hours of coaching. As a consequence, the estimate of the variance component is inflated relative to those of the multilevel multivariate meta-analytic model.

Second, the estimates of the fixed effects from Model 2 are reasonable and are not dissimilar to those of the multilevel multivariate meta-analytic model. However, because the basic random effects model has only a single variance component parameter and thus assumes equality across dependent variables, the Model 2 estimate of the variance component (i.e., 0.007) is rather nonsensical and uninterpretable because it ignores differences between dependent variables; indeed, the multilevel multivariate meta-analytic model estimates indicate substantial differences between the dependent variables (i.e., 0.011 for math and 0.003 for verbal). Further, Model 2 assumes that there is no correlation in this variation between the two dependent variables; on the other hand, the multilevel multivariate meta-analytic model estimates indicate that the correlation is substantial (i.e., 0.768).

Consequently, these basic random effects model variants are unsuitable for conducting statistical inference on the fixed effect parameters or linear combinations of them (e.g., the estimates of the standard errors of the estimates of the fixed effects for Model 2 are inflated relative to those of the multilevel multivariate meta-analytic model for verbal and deflated for math and the estimates of the correlations of the estimates of the fixed effects for Model 2 are deflated relative to those of the multilevel multivariate meta-analytic model). Therefore, we conduct statistical inference for such linear combinations—in particular, the effect of various hours of coaching on math and verbal scores—for only the multilevel multivariate meta-analytic model. Specifically, we conduct statistical inference for the linear combinations

$$\alpha_{0,\text{Math}} + \alpha_{1,\text{Math}}\text{Hours}$$

and

$$\alpha_{0,\text{Verbal}} + \alpha_{1,\text{Verbal}}\text{Hours}.$$

Estimates, 95% CI estimates, estimates of the standard errors of the estimates, and z-statistics and p-values against the point null hypothesis of zero are presented in Table 17. The estimates indicate that SAT coaching results in improvements in math and verbal scores depending on the number of hours of coaching.

Because the estimates in Table 17 are in standard deviation units, it is difficult to assess their magnitude. In contrast, the estimates in Table 12 from the initial analysis are in SAT score units and thus are more interpretable.

The initial analysis also has another advantage relative to this analysis. Specifically, it can assess variation and covariation that is common to the groups of students within a school as well as that which is idiosyncratic to each group of students. Indeed, the estimates in Table 9 indicate that there is much greater variation at the school level than at the group level (i.e., 0.73 versus 0.27). In contrast, due to the differencing involved in conversion to the standardized mean difference scale, any variation and covariation that is common to the groups of students within a school is lost and only variation and covariation which is idiosyncratic to groups of students can be

Table 17 SAT coaching standardized unconstrained MMCS model effect estimates

Dependent variable & hours	Estimate [95% CI]	Standard error	*z*-statistic	*p*-value
Math; 5 hours	0.096 [0.028,0.165]	0.035	2.76	0.006
Math; 10 hours	0.214 [0.169,0.260]	0.023	9.24	< 0.001
Math; 20 hours	0.449 [0.355,0.544]	0.048	9.35	< 0.001
Verbal; 5 hours	0.041 [-0.015,0.097]	0.029	1.44	0.149
Verbal; 10 hours	0.146 [0.109,0.183]	0.019	7.69	< 0.001
Verbal; 20 hours	0.356 [0.279,0.433]	0.039	9.09	< 0.001

assessed. Consequently, the estimates in Table 15 do not assess this substantial source of variation and covariation.

We also note two cautions that are due when converting to a common scale, standardized or otherwise. First, one should be careful that the same or a similar construct is assessed across studies and that any differences are solely in the scales used for the dependent variable across studies. Second, one should be careful that the conversion employed is reasonable, in particular that any and all differences in how individuals might respond to differences in the scales are accounted for by the conversion.

In addition, we note that the historical motivation for conversion to a standardized scale was to adjust for differences in the scales used for the dependent variable across studies. However, such adjustment is not necessary in this case study (i.e., because the same scale was used for each dependent variable across all studies). Further, it is preferable in meta-analysis, as in statistical analysis more generally, to present estimates on the original scale when possible (e.g., when the scale used for the dependent variable across studies is the same; Tukey (1969), Greenland, Schlesselman, and Criqui (1986), Wilkinson (1999), Bond, Wiitala, and Richard (2003), and Baguley (2009)).

We finally note that when the scale used for a dependent variable across studies is the same as in this case study, it is also common in meta-analysis to convert the data to the raw (or unstandardized) difference scale prior to analysis such that the resulting estimates are estimates of differences (i.e., as are those estimates presented in Table 12). While this approach avoids any concerns about conversion, it is nonetheless disadvantageous for a reason discussed above, namely that due to the differencing any variation and covariation that is common to the groups of students within a school is lost and only variation and covariation which is idiosyncratic to groups of students can be assessed.

The appendix details the degree to which the model specifications implemented by MLMVmeta accommodate differences in the scales used across both dependent variables and studies.

Multilevel multivariate model specification

Before proceeding to the third and most complex case study, we introduce the multilevel multivariate meta-analytic model specification in full generality.

The general model specification is given by

$$y_i = \alpha_{c_i,d_i} + \beta_i + \varepsilon_i$$

where i indexes the observations; the $\alpha_{c,d}$ are treated as fixed effects that model the overall average for each experimental condition and dependent variable; the c_i denote the experimental condition $c \in \{1,...,C\}$ to which each observation is assigned; the d_i denote the dependent variable $d \in \{1,...,D\}$ which each observation measures; the β_i are treated as random effects for each observation; and the ε_i are random errors for each observation.

The model further assumes that the vector containing the β_i is distributed multivariate normal with mean zero and variance-covariance matrix specified according to the MMCS model specification with variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ to be discussed below. It also assumes that the vector containing the ε_i is distributed multivariate normal with mean zero and variance-covariance matrix given by the estimate of the sampling variance-covariance matrix of the y_i and that the β_i and ε_j are independent. As is standard in meta-analysis, the estimate of the sampling variance-covariance matrix is assumed known, and the goal is to estimate the fixed effects parameters $\alpha_{c,d}$ and the variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$.

The MMCS model specification implies that

$$\text{var}(y_i) = \text{var}(\beta_i) + \text{var}(\varepsilon_i) = \left(\sum_{k=2}^{K} \tau^2_{k,d_i}\right) + v_i$$

and

$$\begin{aligned}\text{cov}(y_i, y_j) &= \text{cov}(\beta_i, \beta_j) + \text{cov}(\varepsilon_i, \varepsilon_j) \\ &= \left(\sum_{k=2}^{K} \rho_{k,d_i,d_j} \tau_{k,d_i} \tau_{k,d_j} m_{k,i,j}\right) + v_{i,j}\end{aligned}$$

where K is the number of levels in the nesting structure; the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ respectively give the variances and correlations that model the variation and covariation among the β_i induced by level k of the nesting structure; $m_{k,i,j}$ is one if observations i and j are nested in the same group at level k and zero otherwise; and the v_i and $v_{i,j}$ are respectively the assumed known sampling variances and covariances of the observations. As such, the β_i can be decomposed as

$$\beta_i = \sum_{k=2}^{K} \beta_{k,\mathrm{ID}_{k,i},d_i}$$

where the $\mathrm{ID}_{k,i}$ denote the identity of each observation at each level and the vectors containing the $\beta_{k,\mathrm{ID},1},...,\beta_{k,\mathrm{ID},D}$ for each ID at each level are assumed to be independently distributed multivariate normal with mean zero and variance-covariance matrix specified by the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$.

As can be seen, this model generalizes the basic random effects model in several important respects. In particular, the basic random effects model holds when $K=2$ and $m_{2,i,j}=0$ for all $i \neq j$ such that the $\rho_{2,d,d'}$ are irrelevant and the $v_{i,j}=0$, the $\alpha_{c,d} = \alpha$ for all c and d, and the $\tau^2_{2,d} = \tau^2$ for all d.

Like the basic random effects model, this model can, as illustrated in Case Study 2, be generalized to accommodate covariates by replacing the $\alpha_{c,d}$ with observation-specific α_i which are specified as $\alpha_i = \alpha_{0,c,d} + \sum_{p=1}^{P} \alpha_{p,c,d} x_{i,p}$ where $x_{i,p}$ is the p^{th} covariate for observation i and P is the number of covariates.

While there are multiple ways to estimate this model as well as the basic random effects model, estimation on MLMVmeta proceeds as follows: (i) the sampling variance-covariance matrix is supplied by the user (typically an estimate of this matrix based on conventional formulae is supplied, and, as is standard in meta-analysis, this estimate is assumed known); (ii) the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ are estimated using REML conditional on the (estimate of the) sampling variance-covariance matrix; and (iii) the $\alpha_{c,d}$ and their variance-covariance matrix is estimated using generalized least squares conditional on the estimates of the sampling variance-covariance matrix and the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ (Harville, 1977; Robinson, 1991).

Case study 3: Choice overload

Studies of the choice overload hypothesis, the conjecture that an increase in the number of options from which to choose can result in adverse consequences such as a decrease in the likelihood of making a choice or a decrease in the satisfaction with a choice, are an important source of data in consumer psychology. In Table 18, we present hypothetical data typical of studies of choice overload. Each study reports an estimate of the mean (labeled y in the table) of one or more of three dependent variables (confidence, regret, and satisfaction; regret has been reverse-coded for consistency with the other two dependent variables) in either two or four of a total of ten experimental conditions arising from the manipulation of one or two of a total of three experimental factors (assortment size, decision goal, and decision task difficulty) and an estimate of the sampling variance of mean of the dependent variable (labeled v in the table). The table also reports the identity of the paper from which the study came, the study itself, the group(s) of subjects in the study, and the experimental conditions in the study. As can be seen, these studies vary considerably in terms of their experimental factors and dependent variables; examine multiple conditions that result from the manipulation of those experimental factors and give rise to multiple dependent effects of interest; and employ a mix of study designs. On MLMVmeta, we present the estimate of the sampling variance-covariance matrix of the means of the dependent variables.

The multilevel multivariate meta-analytic model specification for this data is given by

$$y_i = \alpha_{c_i,d_i} + \beta_i + \varepsilon_i$$

where i indexes the observations in Table 18; the $\alpha_{c,d}$ are treated as fixed effects that model the overall average for each experimental condition and dependent variable; the c_i denote the experimental condition $c \in \{1,...,10\}$ (here denoting "Decision task difficulty low; assortment size small," "Decision task difficulty low; assortment size large," etc.) to which each observation is assigned; the d_i denote the dependent variable $d \in \{1,2,3\}$ (here denoting confidence, regret, and satisfaction, respectively) which each observation measures; the β_i are treated as random effects for each observation; and the ε_i are random errors for each observation.

The model further assumes that the vector containing the β_i is distributed multivariate normal with mean zero and variance-covariance matrix specified according to the MMCS model specification with variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ to be discussed below. It also assumes that the vector containing the ε_i is distributed multivariate normal with mean zero and variance-covariance matrix given by the estimate of the sampling variance-covariance matrix discussed above and that the β_i and ε_j are independent. As is standard in meta-analysis, the estimate of the sampling variance-covariance matrix is assumed known, and the goal is to estimate the fixed effect parameters $\alpha_{c,d}$ and the variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$.

The MMCS model specification implies that

$$\mathrm{var}(y_i) = \mathrm{var}(\beta_i) + \mathrm{var}(\varepsilon_i) = \sum_{k=2}^{5} \tau^2_{k,d_i} + v_i$$

Table 18 Choice overload data. The full data as well as the estimate of the sampling variance-covariance matrix are available on MLMVmeta

Condition description	Dependent variable	Paper ID	Study ID	Group ID	Condition ID	y	v
Decision task difficulty low; assortment size small	Confidence	1	1	1	1	7.9310	0.0666
Decision task difficulty low; assortment size large	Confidence	1	1	1	2	8.3303	0.0595
Decision task difficulty high; assortment size small	Confidence	1	1	2	3	7.9951	0.0533
Decision task difficulty high; assortment size large	Confidence	1	1	2	4	7.3869	0.0455
Decision task difficulty high; assortment size small	Confidence	1	2	3	5	7.8693	0.0669
Decision task difficulty high; assortment size large	Confidence	1	2	3	6	8.2601	0.0709
Decision task difficulty high; assortment size small	Confidence	1	2	4	7	8.0077	0.0578
Decision task difficulty high; assortment size large	Confidence	1	2	4	8	7.0553	0.0875
Decision task difficulty high; assortment size small	Satisfaction	1	2	3	5	8.4830	0.0658
Decision task difficulty high; assortment size large	Satisfaction	1	2	3	6	8.5876	0.0568
Decision task difficulty high; assortment size small	Satisfaction	1	2	4	7	8.1818	0.0562
Decision task difficulty high; assortment size large	Satisfaction	1	2	4	8	8.0902	0.0555
Decision task difficulty high; assortment size small	Confidence	1	3	5	9	7.8656	0.0534
Decision task difficulty high; assortment size large	Confidence	1	3	6	10	8.1689	0.0547
Decision task difficulty high; assortment size small	Confidence	1	3	7	11	7.7538	0.0403
Decision task difficulty high; assortment size large	Confidence	1	3	8	12	7.1776	0.0557
Decision task difficulty high; assortment size small	Satisfaction	1	3	5	9	6.6317	0.0803
Decision task difficulty high; assortment size large	Satisfaction	1	3	6	10	7.9451	0.0581
Decision task difficulty high; assortment size small	Satisfaction	1	3	7	11	7.7166	0.0557
Decision task difficulty high; assortment size large	Satisfaction	1	3	8	12	6.9272	0.0593
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

and

$$\begin{aligned} \text{cov}(y_i, y_j) &= \text{cov}(\beta_i, \beta_j) + \text{cov}(\varepsilon_i, \varepsilon_j) \\ &= \sum_{k=2}^{5} \rho_{k,d_i,d_j} \tau_{k,d_i} \tau_{k,d_j} m_{k,i,j} + v_{i,j} \end{aligned}$$

where the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ respectively give the variances and correlations that model the variation and covariation among the β_i induced by level k of the nesting structure (here level 2 denotes the experimental condition, level 3 denotes the group of subjects, level 4 denotes the study, and level 5 denotes the paper); $m_{k,i,j}$ is one if observations i and j are nested in the same group at level k and zero otherwise; and the v_i and $v_{i,j}$ are respectively the assumed known sampling variances and covariances of the observations.

Table 19 Choice overload MMCS and EAMMCS model performance

Model specification	Variance component parameters	REML LL	AIC
MMCS			
Fixed effects	0	-262.79	525.58
Equal variance, zero correlation	4	-116.50	241.00
Unequal variance, zero correlation	12	-97.65	219.31
Unequal variance, single correlation	16	-89.06	210.13
EAMMCS			
Fixed effects	0	-262.79	525.58
Equal variance, zero correlation	4	-116.50	241.00
Unequal variance, zero correlation	6	-104.50	221.00
Unequal variance, single correlation	7	-97.45	208.89

In addition to this MMCS specification parameterized by the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ where $k \in \{2,...,5\}$ and $d, d' \in \{1,2,3\}$, we as in Case Study 2 again consider the EAMMCS model specification parameterized by the τ^2_d, $\rho_{d,d'}$, and π_k and again consider a sequence of nested simplifications to these unconstrained MMCS and EAMMCS model specifications. For MMCS, we consider a fixed effects specification that sets the $\tau^2_{k,d} = 0$ for all k and d such that the $\rho_{k,d,d'}$ are irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_{k,d} = \tau^2_k$ for all d and the $\rho_{k,d,d'} = 0$ for all k and $d \neq d'$; an unequal variance, zero correlation specification that sets the $\rho_{k,d,d'} = 0$ for all k and $d \neq d'$; and an unequal variance, single correlation specification that sets the $\rho_{k,d,d'} = \rho_k$ for all $d \neq d'$. For EAMMCS, we consider a fixed effects specification that sets the $\tau^2_d = 0$ for all d such that the $\rho_{d,d'}$ and π_k are irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_d = \tau^2$ for all d and the $\rho_{d,d'} = 0$ for all $d \neq d'$; an unequal variance, zero correlation specification that sets the $\rho_{d,d'} = 0$ for all $d \neq d'$; and an unequal variance, single correlation specification that sets the $\rho_{d,d'} = \rho$ for all $d \neq d'$. For reasons of estimability, interpretability, and parsimony, we do not consider unconstrained MMCS and EAMMCS model specifications.

We present the number of variance component parameters estimated by, the REML LL of, and the AIC of each of these model specifications in Table 19. As can be seen, the single correlation EAMMCS model specification performs best in terms of AIC. Consequently, we present estimates from it in Tables 20 and 21. The estimates indicate that the effect of large versus small assortment sizes varies according to the moderator variable (i.e., decision goal and decision task difficulty) level and dependent variable, a comparison to which we return at the end of this case study. In addition, they indicate that there is much greater variation among the β_i for satisfaction as compared to confidence and regret (i.e., 0.47 versus 0.05 and 0.17, respectively). Further, they indicate that the majority (i.e., 0.40 + 0.32 = 0.72) of the variation and covariation among the β_i is at the paper and study levels although all levels arguably provide a meaningful source of variation and covariation. Finally, they indicate that there is substantial correlation (i.e., 0.63) among the β_i for the three dependent variables at each level.

Given the estimates of the EAMMCS variance component parameters τ^2_d, $\rho_{d,d'}$, and π_k presented in Table 21, it is trivial to obtain estimates of the MMCS variance component parameters $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ as illustrated in Case Study 2.

In addition to considering multilevel multivariate meta-analytic model specifications, we also present estimates from the basic random effects model in Table 22. In particular, we present estimates from four variants of this model: Model 1 is the basic random effects model with a single fixed effect parameter α and a single variance component parameter τ^2, Model 2 generalizes Model 1 to allow for fixed effect parameters that accommodate differences between small and large assortment sizes, Model 3 generalizes Model 2 to further allow for fixed effect parameters that accommodate differences between all experimental conditions, and Model 4 generalizes Model 3 to further allow for fixed effect parameters that accommodate differences between the dependent variables. We note that REML LL and AIC are comparable only across models that include the same fixed effect parameters; consequently, only Model 4 can be compared to the model specifications presented in Table 19. The REML LL and AIC of Model 4 are -159.8 and 321.59, respectively. Therefore, it performs worse in terms of AIC as compared to the multilevel multivariate meta-analytic model specification discussed above.

From the perspective of theory, these basic random effects model variants are inappropriate because they ignore differences between experimental conditions and dependent variables. Further, from the perspective of the data structure, they are inappropriate because they ignore the fact that observations are nested within papers, studies, groups of subjects, and experimental conditions and

Table 20 Choice overload unequal variance, single correlation EAMMCS model fixed effect estimates

Dependent variable & condition description	Estimate	Standard error
Confidence; unmoderated; assortment size small	7.19	0.26
Confidence; unmoderated; assortment size large	6.05	0.26
Confidence; decision goal low; assortment size small	5.27	0.17
Confidence; decision goal low; assortment size large	6.81	0.16
Confidence; decision goal high; assortment size small	5.74	0.16
Confidence; decision goal high; assortment size large	6.65	0.16
Confidence; decision task difficulty low; assortment size small	7.66	0.11
Confidence; decision task difficulty low; assortment size large	7.84	0.11
Confidence; decision task difficulty high; assortment size small	7.65	0.11
Confidence; decision task difficulty high; assortment size large	6.98	0.11
Regret; unmoderated; assortment size small	8.81	0.28
Regret; unmoderated; assortment size large	6.85	0.28
Regret; decision goal low; assortment size small	2.37	0.20
Regret; decision goal low; assortment size large	5.73	0.20
Regret; decision goal high; assortment size small	2.81	0.20
Regret; decision goal high; assortment size large	5.46	0.20
Regret; decision task difficulty low; assortment size small	7.96	0.18
Regret; decision task difficulty low; assortment size large	8.02	0.17
Regret; decision task difficulty high; assortment size small	8.19	0.18
Regret; decision task difficulty high; assortment size large	7.22	0.18
Satisfaction; unmoderated; assortment size small	7.78	0.65
Satisfaction; unmoderated; assortment size large	7.27	0.64
Satisfaction; decision goal low; assortment size small	5.39	0.35
Satisfaction; decision goal low; assortment size large	7.02	0.35
Satisfaction; decision goal high; assortment size small	5.84	0.35
Satisfaction; decision goal high; assortment size large	6.89	0.35
Satisfaction; decision task difficulty low; assortment size small	7.56	0.26
Satisfaction; decision task difficulty low; assortment size large	7.75	0.26
Satisfaction; decision task difficulty high; assortment size small	7.80	0.26
Satisfaction; decision task difficulty high; assortment size large	7.01	0.26

treat all observations as independent. Both of these have important implications for the estimates obtained from these models.

First, the estimate of the fixed effect from Model 1 is rather nonsensical and uninterpretable because it ignores differences between experimental conditions and dependent variables. As a consequence, the estimate of the variance component is inflated relative to those of the multilevel multivariate meta-analytic model.

Table 21 Choice overload unequal variance, single correlation EAMMCS model variance component estimates

Dependent variable	τ^2 Estimate	ρ Estimate	Level	π Estimate
Confidence	0.05	0.63	Paper	0.40
Regret	0.17		Study	0.32
Satisfaction	0.47		Group	0.18
			Condition	0.10

Second, and similarly, the estimates of the fixed effects from Model 2 (Model 3) are rather nonsensical and uninterpretable because they ignore differences between experimental conditions and dependent variables (dependent variables). As a consequence, the estimate of the variance component is inflated relative to those of the multilevel multivariate meta-analytic model.

Third, the estimates of the fixed effects from Model 4 are reasonable and are not dissimilar to those of the multilevel multivariate meta-analytic model. However, because the basic random effects model has only a single variance component parameter and thus assumes that there are only two levels in the nesting structure and equality across dependent variables, the Model 4 estimate of the variance component (i.e., 0.17) is rather nonsensical and uninterpretable because it ignores differences among

Table 22 Choice overload basic random effects model estimates

Model	Dependent variable & condition description	α		τ^2
		Estimate	Standard error	Estimate
Model 1	Not applicable	6.77	0.10	2.38
Model 2	Assortment size small	6.54	0.14	2.34
	Assortment size large	7.00	0.14	
Model 3	Unmoderated; assortment size small	8.20	0.35	0.65
	Unmoderated; assortment size large	6.73	0.35	
	Decision goal low; assortment size small	4.07	0.18	
	Decision goal low; assortment size large	6.39	0.18	
	Decision goal high; assortment size small	4.43	0.18	
	Decision goal high; assortment size large	6.17	0.18	
	Decision task difficulty low; assortment size small	7.71	0.14	
	Decision task difficulty low; assortment size large	7.86	0.14	
	Decision task difficulty high; assortment size small	7.87	0.14	
	Decision task difficulty high; assortment size large	7.06	0.14	
Model 4	Confidence; unmoderated; assortment size small	7.23	0.35	0.17
	Confidence; unmoderated; assortment size large	6.09	0.35	
	Confidence; decision goal low; assortment size small	5.28	0.22	
	Confidence; decision goal low; assortment size large	6.80	0.21	
	Confidence; decision goal high; assortment size small	5.69	0.22	
	Confidence; decision goal high; assortment size large	6.59	0.22	
	Confidence; decision task difficulty low; assortment size small	7.60	0.13	
	Confidence; decision task difficulty low; assortment size large	7.85	0.13	
	Confidence; decision task difficulty high; assortment size small	7.64	0.13	
	Confidence; decision task difficulty high; assortment size large	6.99	0.13	
	Regret; unmoderated; assortment size small	8.89	0.28	
	Regret; unmoderated; assortment size large	6.89	0.28	
	Regret; decision goal low; assortment size small	2.38	0.15	
	Regret; decision goal low; assortment size large	5.71	0.15	
	Regret; decision goal high; assortment size small	2.78	0.15	
	Regret; decision goal high; assortment size large	5.41	0.15	
	Regret; decision task difficulty low; assortment size small	7.82	0.14	
	Regret; decision task difficulty low; assortment size large	7.82	0.14	
	Regret; decision task difficulty high; assortment size small	8.01	0.14	
	Regret; decision task difficulty high; assortment size large	7.03	0.14	
	Satisfaction; unmoderated; assortment size small	8.04	0.52	
	Satisfaction; unmoderated; assortment size large	7.52	0.50	
	Satisfaction; decision goal low; assortment size small	5.42	0.17	
	Satisfaction; decision goal low; assortment size large	7.00	0.17	
	Satisfaction; decision goal high; assortment size small	5.74	0.17	
	Satisfaction; decision goal high; assortment size large	6.87	0.17	
	Satisfaction; decision task difficulty low; assortment size small	7.70	0.13	
	Satisfaction; decision task difficulty low; assortment size large	7.90	0.13	
	Satisfaction; decision task difficulty high; assortment size small	7.95	0.13	
	Satisfaction; decision task difficulty high; assortment size large	7.14	0.13	

dependent variables; indeed, the multilevel multivariate meta-analytic model estimates indicate substantial differences among the dependent variables (i.e., 0.05 for confidence, 0.17 for regret, and 0.47 for satisfaction). Further, Model 4 assumes that none of the variation is at the paper level, none is at the study level, none is at the group of subjects level, all is at the experimental condition level, and there is no covariation among the three dependent

Table 23 Choice overload unequal variance, single correlation EAMMCS model simple effect estimates

Dependent variable & moderator variable level	Estimate [95% CI]	Standard error	*z*-statistic	*p*-value
Confidence; unmoderated	-1.14 [-1.59,-0.69]	0.23	-4.98	< 0.001
Confidence; decision goal low	1.53 [1.24,1.83]	0.15	10.15	< 0.001
Confidence; decision goal high	0.91 [0.63,1.20]	0.15	6.22	< 0.001
Confidence; decision task difficulty low	0.18 [0.01,0.35]	0.09	2.09	0.036
Confidence; decision task difficulty high	-0.67 [-0.85,-0.50]	0.09	-7.52	< 0.001
Regret; unmoderated	-1.96 [-2.40,-1.52]	0.22	-8.74	< 0.001
Regret; decision goal low	3.36 [3.11,3.61]	0.13	26.68	< 0.001
Regret; decision goal high	2.65 [2.41,2.88]	0.12	22.09	< 0.001
Regret; decision task difficulty low	0.06 [-0.16,0.28]	0.11	0.55	0.584
Regret; decision Task difficulty high	-0.97 [-1.19,-0.75]	0.11	-8.68	< 0.001
Satisfaction; unmoderated	-0.50 [-1.38,0.37]	0.45	-1.13	0.259
Satisfaction; decision goal low	1.63 [1.27,1.99]	0.18	8.93	< 0.001
Satisfaction; decision goal high	1.05 [0.70,1.41]	0.18	5.77	< 0.001
Satisfaction; decision task difficulty low	0.19 [-0.08,0.45]	0.14	1.38	0.166
Satisfaction; decision task difficulty high	-0.78 [-1.05,-0.52]	0.13	-5.89	< 0.001

Table 24 Choice overload EAMMCS unequal variance, single correlation model interaction effect estimates

Dependent variable & moderator variable	Estimate [95% CI]	Standard error	*z*-statistic	*p*-value
Confidence; decision goal	-0.62 [-1.03,-0.21]	0.21	-2.94	< 0.001
Confidence; decision task difficulty	-0.86 [-1.10,-0.61]	0.12	-6.87	< 0.001
Regret; decision goal	-0.71 [-1.05,-0.37]	0.17	-4.09	< 0.001
Regret; decision task difficulty	-1.03 [-1.34,-0.72]	0.16	-6.55	< 0.001
Satisfaction; decision goal	-0.58 [-1.08,-0.07]	0.26	-2.24	0.025
Satisfaction; decision task difficulty	-0.97 [-1.34,-0.60]	0.19	-5.12	< 0.001

variables at any level; on the other hand, the multilevel multivariate meta-analytic model estimates indicate that 40% of the variation is at the paper level, 32% is at the study level, 19% is at the group of subjects level, 10% is at the experimental condition level, and the covariation is substantial (i.e., correlated at 0.63).

Consequently, these basic random effects model variants are unsuitable for conducting statistical inference on the fixed effect parameters or linear combinations of them (e.g., the estimates of the standard errors of the estimates of the fixed effects for Model 4 are inflated relative to those of the multilevel multivariate meta-analytic model for confidence and deflated for regret and satisfaction and the estimates of the correlations of the estimates of the fixed effects for Model 4 are deflated relative to those of the multilevel multivariate meta-analytic model). Therefore, we conduct statistical inference for such linear combinations—in particular, the simple effect of large versus small assortment sizes at each level of the moderator variables for each dependent variable and the assortment size × moderator variable interaction effect for each dependent variable—for only the multilevel multivariate meta-analytic model. Estimates, 95% CI estimates, estimates of the standard errors of the estimates, and *z*-statistics and *p*-values against the point null hypothesis of zero are presented in Tables 23 and 24, respectively. The estimates indicate that choice overload results for all three dependent variables in unmoderated studies and when the decision task difficulty moderator variable is set to the high level; however, it is reversed when the decision task difficulty moderator variable is set to the low level and when decision goal is the moderator variable. Further, the estimates of the effects on confidence and satisfaction appear similar. Finally, the estimates of the interaction effects are consistently negative in line with the choice overload hypothesis.

Discussion

In this paper, we aimed to help overcome challenges so that multilevel multivariate meta-analytic models will be more often employed in practice. We did so by introducing MLMVmeta, an easy-to-use web application that implements multilevel multivariate meta-analytic methodology that is both specially tailored to contemporary psychological research and easily estimable, interpretable, and parsimonious. We are optimistic that MLMVmeta will facilitate the use of multilevel multivariate meta-analytic models in practice.

Appendix

To reproduce the results reported in the case studies using MLMVmeta, first click the "Examples" tab and download the data. Next, click the "Input" tab and load the principal data (for all three case studies) and the sampling variance-covariance matrix (for the second and third case studies). Then, click the "Click to Estimate Model" button. It is important to wait until the data has fully loaded before clicking this button; completion of loading is indicated by the fact that the principal data and, when applicable, the sampling variance-covariance matrix appear in the main panel of the tab.

After the progress bar indicates that estimation is complete, results can be obtained by clicking the "Results" tab. This tab provides the number of variance component parameters estimated by, the REML LL of, and the AIC of each of the model specifications as well as the estimates from all model specifications. These results may be downloaded as an .RData file by scrolling to the bottom of the tab and clicking the link.

To analyze other data using MLMVmeta, users must prepare the principal data and optionally the sampling variance-covariance matrix (e.g., as a .csv file). The first two columns to appear in the principal data file must be titled ConditionDescription and DependentVariable; these columns respectively provide the experimental condition and dependent variable for each observation (when there are no experimental conditions, the entries in this column can simply be "Not Applicable"). Next, one or more columns that provide the identity of each observation at each level of the nesting structure may optionally appear in the principal data file. The next two columns to appear in the principal data file must be titled y and v; the former provides a single statistic that summarizes individual-level data associated with each observation and the latter provides an estimate of the sampling variance of the statistic. Finally, one or more columns that provide covariates for each observation may optionally appear in the principal data file. The sampling variance-covariance matrix provides (typically an estimate of) the sampling variance-covariance matrix of y; it is, as noted, optional to prepare this matrix but it is recommended to do so when the covariances are nonzero.

In addition to the unconstrained MMCS model specification parameterized by the $\tau^2_{k,d}$ and $\rho_{k,d,d'}$ where $k \in \{2,...,K\}$ and $d, d' \in \{1, ..., D\}$ and discussed in the section entitled "Multilevel multivariate model specification," MLMVmeta implements several simplifications: a fixed effects specification that sets the $\tau^2_{k,d} = 0$ for all k and d such that the $\rho_{k,d,d'}$ are irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_{k,d} = \tau^2_k$ for all d and the $\rho_{k,d,d'} = 0$ for all k and $d \neq d'$; an equal variance, single correlation specification that sets the $\tau^2_{k,d} = \tau^2_k$ for all d and the $\rho_{k,d,d'} = \rho_k$ for all $d \neq d'$; an unequal variance, zero correlation specification that sets the $\rho_{k,d,d'} = 0$ for all k and $d \neq d'$; and an unequal variance, single correlation specification that sets the $\rho_{k,d,d'} = \rho_k$ for all $d \neq d'$.

In addition to the unconstrained EAMMCS model specification parameterized by the τ^2_d, $\rho_{d,d'}$, and π_k where $d, d' \in \{1, ..., D\}$ and $k \in \{2,...,K\}$ and discussed in Case Study 2, MLMVmeta implements several simplifications: a fixed effects specification that sets the $\tau^2_d = 0$ for all d such that the $\rho_{d,d'}$ and π_k are irrelevant; an equal variance, zero correlation specification that sets the $\tau^2_d = \tau^2$ for all d and the $\rho_{d,d'} = 0$ for all $d \neq d'$; an equal variance, single correlation specification that sets the $\tau^2_d = \tau^2$ for all d and the $\rho_{d,d'} = \rho$ for all $d \neq d'$; an unequal variance, zero correlation specification that sets the $\rho_{d,d'} = 0$ for all $d \neq d'$; and an unequal variance, single correlation specification that sets the $\rho_{d,d'} = \rho$ for all $d \neq d'$.

We note that (i) all model specifications other than the fixed effects ones are equivalent when the number of dependent variables is equal to one; (ii) MMCS and EAMMCS model specifications are equivalent when the number of dependent variables is equal to one, when the number of levels is equal to two, for the fixed effects model specifications, and for the equal variance, zero correlation model specifications; (iii) the unconstrained model specifications are estimated only when the number of dependent variables is equal to one or when the number of levels is less than or equal to three; and (iv) only the unconstrained MMCS model specification allows negative correlations. We also note that only EAMMCS model specifications are estimated by default; MMCS model specifications may be estimated at the discretion of the user but require more time for estimation. We finally note that the MMCS equal variance, single correlation model specification; the MMCS unequal variance, single correlation model specification; and the EAMMCS equal variance, single correlation model specification are novel relative to McShane and Böckenholt (2018).

MLMVmeta does not accommodate differences in the scales used for a given dependent variable across studies; when studies use different scales for a given dependent variable, conversion to a common scale (heeding the cautions discussed in Case Study 2 Revisited) prior to analysis is necessary. MLMVmeta does accommodate differences in the scales used for different dependent variables; however, equal variance model specifications may be rather nonsensical and perform poorly when there are such differences. Finally, dependent variables should be reverse-coded as necessary for consistency with one another prior to analysis; otherwise, single correlation model specifications and the unconstrained EAMMCS model specification (which allow only nonnegative correlations) may be rather nonsensical and perform poorly.

References

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*(3), 603–617.

Becker, B.J. (2000) *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. San Diego: Academic Press. Chap Multivariate Meta-analysis.

Berkey, C.S., Hoaglin, D.C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G.A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, *17*(22), 2537–2550.

Bond, C.F., Jr., Wiitala, W. L., & Richard, F. D. (2003). Meta-analysis of raw mean differences. *Psychological Methods, 8*(4), 406.

Chernev, A., Böckenholt, U, & Goodman, J. (2015). Choice overload: a conceptual review and meta-analysis. *Journal of Consumer Psychology*, *25*(2), 333–358.

Cheung, M.W.L. (2015) *Meta-Analysis: A Structural Equation Modeling Approach*. West Sussex, England: Wiley.

Cooper, H., Hedges, L.V., & Valentine J.C. (2019). The handbook of research synthesis and meta-analysis. Russell Sage Foundation.

Greenland, S., Schlesselman, J.J., & Criqui, M.H. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology*, *123*(2), 203–208.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), 320–338.

Kalaian, H.A., & Raudenbush, S.W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, *1*(3), 227–235.

McShane, B.B., & Böckenholt, U. (2018). Multilevel multivariate meta-analysis with application to choice overload. *Psychometrika*, *83*(1), 255–271.

McShane, B.B., & Böckenholt, U. (2020). Enriching meta-analytic models of summary data: a thought experiment and case study. *Advances in Methods and Practices in Psychological Science*, *3*(1), 81–93.

Polderman, T.J., Benyamin, B., De Leeuw, C.A., Sullivan, P.F., Van Bochoven, A., Visscher, P.M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709.

Raudenbush, S.W., & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Sage, Thousand Oaks, CA.

Robinson, G.K. (1991). That blup is a good thing: the estimation of random effects. *Statistical Science*, *6*(1), 15–32.

Schmid, C.H., Stijnen, T., & White, I. (2020). Handbook of Meta-analysis. CRC Press.

Tipton, E., Pustejovsky, J.E., & Ahmadi, H. (2019). Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods*, *10*(2), 180–194.

Tukey, J.W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*(2), 83–91.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604.

Open Practices Statement As detailed in the Appendix, the hypothetical case study data is available on MLMVmeta, and it can be used to reproduce the results reported in the case studies.