

Multilevel meta-analysis of multiple regression coefficients from single-case experimental studies

Laleh Jamshidi, Lies Declercq, Belén Fernández-Castilla, John M. Ferron, Mariola Moeyaert, S. Natasha Beretvas & Wim Van den Noortgate

Behavior Research Methods

e-ISSN 1554-3528

Behav Res

DOI 10.3758/s13428-020-01380-w



Your article is protected by copyright and all rights are held exclusively by The Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Multilevel meta-analysis of multiple regression coefficients from single-case experimental studies

Laleh Jamshidi^{1,2} · Lies Declercq¹ · Belén Fernández-Castilla¹ · John M. Ferron³ · Mariola Moeyaert⁴ · S. Natasha Beretvas⁵ · Wim Van den Noortgate¹

© The Psychonomic Society, Inc. 2020

Abstract

The focus of the current study is on handling the dependence among multiple regression coefficients representing the treatment effects when meta-analyzing data from single-case experimental studies. We compare the results when applying three different multilevel meta-analytic models (i.e., a univariate multilevel model avoiding the dependence, a multivariate multilevel model ignoring covariance at higher levels, and a multivariate multilevel model modeling the existing covariance) to deal with the dependent effect sizes. The results indicate better estimates of the overall treatment effects and variance components when a multivariate multilevel model is applied, independent of modeling or ignoring the existing covariance. These findings confirm the robustness of multilevel modeling to misspecifying the existing covariance at the case and study level in terms of estimating the overall treatment effects and variance components. The results also show that the overall treatment effect estimates are unbiased regardless of the underlying model, but the between-case and between-study variance components are biased in certain conditions. In addition, the between-study variance estimates are particularly biased when the number of studies is smaller than 40 (i.e., 10 or 20) and the true value of the between-case variance is relatively large (i.e., 8). The observed bias is larger for the between-case variance estimates compared to the between-study variance estimates when the true between-case variance is relatively small (i.e., 0.5).

Keywords Single-case experimental design · Multilevel meta-analysis · Multivariate multilevel model · Robust variance estimator

Introduction

Single-case experimental designs (SCEDs) have been frequently used in many different fields such as psychology, education, social sciences, and medical sciences to establish evidence-based practices. In this kind of design, the outcome

of interest is measured repeatedly over time prior to and during or after the treatment. Afterwards, scores under these conditions (pre and post treatment) can be compared to evaluate treatment effectiveness (Horner et al., 2005; Kratochwill et al., 2010). An inherent SCED characteristic is that each identity (e.g., an individual or a group of individuals) serves as its own control. There are a variety of SCED types including withdrawal or reversal designs, alternating treatment designs, changing criterion designs, and multiple baseline designs (MBDs). In the current study, we focus on MBDs in which the introduction of the treatment is staggered across different time points for multiple cases, outcomes, or settings. The reason for this choice was that MBDs are not only the most frequently used SCED design, but they are also internally and externally more valid compared to other SCED design types (Kratochwill et al., 2010; Onghena & Edgington, 2005).

The data from several MBDs can be synthesized by applying multilevel meta-analytic modeling proposed by Van den Noortgate and Onghena (2003a, 2003b, 2008). A three-level model can be used to model the hierarchical structure of the data: the measurement occasions (i.e., the first-level units) are

✉ Laleh Jamshidi
laleh.jamshidi@uregina.ca

¹ Faculty of Psychology and Educational Sciences & ITEC, imec research group at KU Leuven, KU Leuven, University of Leuven, Leuven, Belgium

² Canadian Institute for Public Safety Research and Treatment (CIPSRT), University of Regina, Regina, Canada

³ Department of Educational Measurement and Research, University of South Florida, Tampa, FL, USA

⁴ Department of Educational and Counseling Psychology, University at Albany (State University of New York), Albany, NY, USA

⁵ Department of Educational Psychology, University of Texas at Austin, Austin, TX, USA

nested within cases (i.e., the second-level units), which in turn are nested within studies (i.e., the third-level units). Multilevel modeling not only can yield improved estimates of the treatment effects for individual cases, but also an average treatment effect estimate across cases for each study and an overall treatment effect estimate across all studies. The variance at the first level, that is the variance between measurement occasions within cases, can be modeled as follows:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}T_{ijk} + \beta_{2jk}D_{ijk} + \beta_{3jk}T_{ijk}D_{ijk} + e_{ijk} \quad \text{with } e_{ijk} \sim N(0, \sigma_e^2) \quad (1)$$

In this equation, i is an indicator for the measurement occasion ($i = 0, 1, \dots, I$), j for the case ($j = 0, 1, \dots, J$), and k for the study ($k = 0, 1, \dots, K$). Therefore, Y_{ijk} is the outcome on the i^{th} measurement occasion from the j^{th} case within study k . The dummy variable D_{ijk} indicates the corresponding phase for each measurement and equals 0 if the measurement belongs to the baseline phase and 1 if it belongs to the treatment phase. T_{ijk} is a time indicator (e.g., days, the session number). If we center the time indicator such that it is equal to 0 at the start point of the treatment phase, β_{0jk} expresses the expected baseline level at the start point of the treatment phase (where $T_{ijk} = 0$) and β_{2jk} the immediate treatment effect on the outcomes. The coefficient β_{1jk} refers to the linear time trend in the outcome during the baseline phase, and β_{3jk} indicates the treatment effect on the time trend. The sampling error, e_{ijk} , is assumed normally, identically and independently distributed with a zero mean and a sampling variance of σ_e^2 . The indices j and k in the regression coefficients mean that these coefficients are case- and study-specific, respectively. The variation over cases (second level) is modeled using additional equations. In a simple model, the case-specific regression coefficients are assumed to deviate randomly from the study mean coefficients. The following equations show that the baseline level, time effect in the baseline phase, the immediate treatment effect, and the treatment effect on the time trend, respectively, vary randomly around the study-specific means:

$$\begin{cases} \beta_{0jk} = \theta_{00k} + u_{0jk} \\ \beta_{1jk} = \theta_{10k} + u_{1jk} \\ \beta_{2jk} = \theta_{20k} + u_{2jk} \\ \beta_{3jk} = \theta_{30k} + u_{3jk} \end{cases} \quad \text{with } \begin{pmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{pmatrix} \sim MVN(\mathbf{0}, \Sigma_u)$$

$$\Sigma_u = \begin{pmatrix} \sigma_{u0}^2 & & & \\ \sigma_{u1u0} & \sigma_{u1}^2 & & \\ \sigma_{u2u0} & \sigma_{u2u1} & \sigma_{u2}^2 & \\ \sigma_{u3u0} & \sigma_{u3u1} & \sigma_{u3u2} & \sigma_{u3}^2 \end{pmatrix} \quad (2)$$

Each case-specific coefficient β_{jk} (i.e., β_{0jk} , β_{1jk} , β_{2jk} , or β_{3jk}) equals a study-specific mean $\theta_{,0k}$ (i.e., θ_{00k} , θ_{10k} , θ_{20k} , or θ_{30k}) plus a random case-specific deviation from this mean. The covariance matrix Σ_u has on its diagonal variances

between cases within the same study and the off-diagonal elements the covariances between cases within the same study. For instance, σ_{u2}^2 and σ_{u3}^2 are the variances between cases from the same study of the immediate treatment effect and the treatment effect on the time trend; σ_{u3u2} represents the covariance between the immediate treatment effect and the treatment effect on the time trend.

The variation of the study-specific means, θ 's, are modeled at the third level as the overall regression coefficients, γ 's, plus a random study-specific deviation from these overall means over all studies.

$$\begin{cases} \theta_{00k} = \gamma_{000} + v_{00k} \\ \theta_{10k} = \gamma_{100} + v_{10k} \\ \theta_{20k} = \gamma_{200} + v_{20k} \\ \theta_{30k} = \gamma_{300} + v_{30k} \end{cases} \quad \text{with } \begin{pmatrix} v_{00k} \\ v_{10k} \\ v_{20k} \\ v_{30k} \end{pmatrix} \sim MVN(\mathbf{0}, \Sigma_v)$$

$$\Sigma_v = \begin{pmatrix} \sigma_{v0}^2 & & & \\ \sigma_{v1v0} & \sigma_{v1}^2 & & \\ \sigma_{v2v0} & \sigma_{v2v1} & \sigma_{v2}^2 & \\ \sigma_{v3v0} & \sigma_{v3v1} & \sigma_{v3v2} & \sigma_{v3}^2 \end{pmatrix} \quad (3)$$

Residuals at case and study levels are assumed to be multivariate normally distributed. Researchers are mainly interested in the overall treatment effects (i.e., the overall immediate treatment effect, γ_{200} , and the overall treatment effect on time trend, γ_{300}) and the (co)variation in these effects over cases and studies. These (and other) model parameters can be estimated by applying maximum likelihood procedures commonly used in multilevel modeling. In addition, one can also obtain improved estimates of the case- and study-specific treatment effects (the β 's and θ 's) by using empirical Bayes techniques (Raudenbush & Bryk, 2002). Multilevel modeling has been applied in the context of SCED data in previous studies, and its performance has been assessed in several simulation studies (e.g., Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2013; Owens & Ferron, 2012; Ugille, Moeyaert, Beretvas, Ferron, & Van den Noortgate, 2012, 2014). In general, the findings have indicated appropriate performance in terms of statistical properties such as relative bias, MSE of this method, both for synthesizing the standardized and unstandardized SCED data.

The model can be modified in various ways, according to the data characteristics and the research interests. For instance, predictor variables can be added at each level to study their effect on the outcome, or to study how they moderate the treatment effects. In addition, it is possible to use alternative specifications of the structure of variances and covariances at each level.

The model can also be adapted to combine effect sizes, rather than raw data from multiple SCED studies. If we obtain the estimates for two effect sizes of interest (i.e., the immediate treatment effect and the treatment effect on the time trend) for each case separately by fitting a simple linear model (Eq. 1),

we can then synthesize the case-specific ordinary least squares (OLS) estimates (i.e., b_{2jk} and b_{3jk}) by using either a separate univariate three-level meta-analysis or a multivariate three-level meta-analysis (Van den Noortgate & Onghena, 2003b, 2008). For instance, b_{2jk} is equal to β_{2jk} , the immediate treatment effect for the case j from the study k , plus a random deviation that is due to sampling variance. At the second and third levels, the equations for immediate effects are the same as for the raw data model (see Eqs. 2 and 3). We can rewrite this three-level model in a single equation (Eq. 4):

$$\begin{aligned}
 b_{2jk} &= \gamma_{200} + u_{2jk} + v_{20k} + r_{2jk} \\
 r_{2jk} &\sim N\left(0, \sigma_{r_{2jk}}^2\right), u_{2jk} \sim N\left(0, \sigma_{u_2}^2\right), v_{2jk} \sim N\left(0, \sigma_{v_2}^2\right)
 \end{aligned}
 \tag{4}$$

in which this estimate of the immediate effect for case j from study k is equal to the overall immediate treatment effect, γ_{200} , plus a case-specific random deviation, u_{2jk} , a study-specific random deviation, v_{20k} , and a residual deviation due to sampling variance, r_{2jk} . This model provides the overall treatment effect, γ_{200} , as well as the corresponding between-case and between-study variance components, $\sigma_{u_2}^2$ and $\sigma_{v_2}^2$, respectively. From the initial OLS regression analysis, we already have an estimate of the sampling variance, $\sigma_{r_{2jk}}^2$, so in the meta-analysis it is assumed to be known. We can use the same model for synthesizing the estimates of the treatment effect on the time trend, b_{3jk} by using Eq. 5:

$$\begin{aligned}
 b_{3jk} &= \gamma_{300} + u_{3jk} + v_{30k} + r_{3jk} \\
 r_{3jk} &\sim N\left(0, \sigma_{r_{3jk}}^2\right), u_{3jk} \sim N\left(0, \sigma_{u_3}^2\right), v_{3jk} \sim N\left(0, \sigma_{v_3}^2\right)
 \end{aligned}
 \tag{5}$$

Alternatively, another way to analyze the multiple effect sizes, taking into account the dependence among these effect sizes, is the use of the multivariate mixed model proposed by Kalaian and Raudenbush (1996), Eq. 6. This model uses the estimated sampling covariance matrix of multivariate effect sizes to obtain the optimal weights for the effect sizes and to estimate model parameters and their corresponding standard errors rather than using sampling variance estimates of the effect sizes, which is the case in ordinary meta-analysis. One of the most important advantages of this approach is that all available information is used in a single analysis in order to estimate any treatment effects (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013).

$$\begin{aligned}
 \begin{cases} b_{2jk} = \gamma_{200} + u_{2jk} + v_{20k} + r_{2jk} \\ b_{3jk} = \gamma_{300} + u_{3jk} + v_{30k} + r_{3jk} \end{cases} \\
 \begin{pmatrix} r_{2jk} \\ r_{3jk} \end{pmatrix} \sim MVN\left(\mathbf{0}, \hat{\sigma}^2(\mathbf{b})\right), \begin{pmatrix} u_{2jk} \\ u_{3jk} \end{pmatrix} \sim MVN(\mathbf{0}, \Sigma_u), \begin{pmatrix} v_{2jk} \\ v_{3jk} \end{pmatrix} \sim MVN(\mathbf{0}, \Sigma_v)
 \end{aligned}
 \tag{6}$$

The sampling covariance matrix, $\hat{\sigma}^2(\mathbf{b})$, the between-case and between-study covariance matrices Σ_u and Σ_v , are elements

of $\mathbb{R}^{2 \times 2}$. Combining these case-specific effect sizes using a bivariate three-level meta-analytic model, Eq. 6, results in the estimates of two fixed effects (γ_{200} and γ_{300}) and six unique (co)variance components (the elements of Σ_u and Σ_v).

The main advantage of the multivariate model over separate univariate three-level models is its ability to provide more information about the potential covariances between effect sizes. This strategy explicitly models the correlations among the effect size estimates and needs information regarding the covariance structure of the effect size estimation errors. Getting knowledge about this structure requires the information about the dependence structure of the raw data in each study (Hedges, Tipton, & Johnson, 2010). If the raw data are available, an estimate of the sampling covariance between effect sizes can indeed be obtained from the OLS regression. A major drawback of this multivariate approach is that when the raw data are not available, this approach is often not applicable because it is often possible to reconstruct the sampling variances (the squared standard errors, that are typically reported or can be reconstructed based on the results of statistical significance tests), but not the sampling covariances (Van den Noortgate et al., 2013).

To understand the sampling covariance between both effects (i.e., immediate treatment effect and treatment effect on the time trend) at the lowest level, consider the situation where there is a positive time trend in the baseline and a positive immediate treatment effect and a positive effect on the time trend. If by coincidence the first measurement in the treatment phase happens to be unexpectedly large, the immediate effect is likely to be overestimated, while the effect on the trend will be underestimated. In general, a negative correlation between both coefficients can be expected. But also at the level of participants or studies, both kinds of effect might be dependent: for instance, both coefficients can be negatively correlated at the case level, if, for example, some cases show immediately a full effect of the treatment, while for other cases, the effect is building up gradually. In this situation, a non-zero covariance between residuals at the case and study level can be expected. Still, in most previous studies of multilevel modeling to synthesize the standardized and unstandardized SCED data (e.g., Moeyaert et al., 2013; Ugille et al., 2012, 2014), it was assumed that each random effect is independently, identically, and normally distributed with the mean of zero and a diagonal covariance matrix at the second and third level. Moeyaert, Ugille, Ferron, Beretvas, and Van den Noortgate (2016) investigated the impact of covariance matrix misspecification on the quality of the estimates of fixed and random effects in the situation where the residuals covary at the case and study level. They modeled the condition in which there were two regression coefficients (i.e., the baseline level and the treatment level) in the multilevel modeling. They found that misspecifying the existing covariance has no significant impact on the estimates of treatment effect. Moreover,

the results indicated that modeling or ignoring the covariance leads to relatively unbiased between-case variance estimates, whereas the between-study variance estimates are biased in both situations.

The aim of current study is to compare different approaches to deal with dependent effect sizes in SCED research. We focus on the situation where you have two effect sizes per case: one referring to the immediate treatment effect and one referring to the effect on a time trend. To accommodate these dependent effect sizes, we use three different approaches: a traditional univariate random effects approach (for each type of coefficients separately) avoiding the existing dependence, a multivariate random effects meta-analytic approach proposed by Kalaian and Raudenbush (1996) ignoring the existing covariance at higher levels, and a multivariate multilevel meta-analysis modeling the covariance at higher levels. In the last two approaches, we also study the performance of the sandwich estimator (Huber, 1967; White, 1982), also known as the robust covariance matrix estimator, the empirical covariance matrix estimator, or robust variance estimator (RVE; Hedges et al., 2010). This procedure is assumed to yield asymptotically consistent covariance matrix estimates, even if the underlying models are incorrect or misspecified. We would like to understand whether treating the existing dependence in different ways has any significant impact on the quality of the overall treatment effects estimates and the between-case and between-study (co)variance estimates. A difference with the study of Moeyaert et al. (2016), is that we focus on the meta-analysis of effect sizes rather than on the meta-analysis of SCED raw data, and we compare the results of a multivariate approach with those of univariate analyses. A difference with the studies of Ugille et al. (2012, 2014), is that we explicitly assume that effect sizes are correlated at higher levels and that we do not only use univariate models to analyze the data.

Method

A simulation study was conducted in order to handle the dependence among multiple regression coefficients in the multilevel meta-analysis of SCED studies when freely estimating covariances between residuals at higher levels (i.e., case-level and study-level). Several models were fitted to the generated data to evaluate their performance in parameter recovery.

Data generation and analyses were done using SAS 9.4 (Fan, Felsovalyi, Sivo, & Keenan, 2002; SAS Institute Inc., 2017).

Data generation

SCED raw data were generated using the three-level model from Eqs. 1 to 3 for MBDs across participants. The

coefficients of covariates in the baseline phase (i. e. , γ_{000} and γ_{100}) were set to 0, whereas the coefficients of covariates related to the treatment (i.e., γ_{200} and γ_{300}) were equal to 2 and 0.2, respectively. We manipulated the between-cases variability and the between-studies variability. In addition, we varied the size of simulated datasets by varying the number of measurement occasions I ($I = 10, 20,$ or 40), the number of cases within studies J ($J = 3, 4,$ or 7), and the number of studies meta-analyzed K ($K = 10, 20,$ or 40). The values for I were chosen based on the findings of several meta-analyses and systematic reviews of SCEDs (Ferron, Farmer, & Owens, 2010; Shadish & Sullivan, 2011). The values for J were selected based on recommendation of several studies on SCEDs (Farmer, Owens, Ferron, & Allsopp, 2010; Ferron et al., 2010; Moeyaert et al., 2013; Shadish & Sullivan, 2011) to find out how small the number of cases per study could be to return unbiased estimates in combination with other conditions. For K , we chose the values based on a systematic review of SCED meta-analyses (Jamshidi et al., 2020).

Because the data were generated for MBDs, the timing of treatment introduction was staggered across subjects. Start points of the treatment for each case (Table 1) were chosen in line with Ugille et al. (2012) and Moeyaert, Ugille, Ferron, Beretvas, and Van den Noortgate (2014).

Previous studies have shown that the between-study and between-case variances are significant factors affecting the estimates of treatment effects and the variances of these estimates (Moeyaert et al., 2013, 2014). The level-1 variance of the raw data was set to 1 and assumed to be homogeneous across phases. For the between-case and between-study covariance matrices (i.e., Σ_u and Σ_v , the covariance matrices at the second and third level), we chose relatively small or relatively large amounts of variance, and therefore the variances (i.e., the diagonal elements in the matrices) were chosen to be either (8, 0.08, 8, 0.08) or (0.5, 0.05, 0.5, 0.05). Because we were interested in the treatment effect estimates and not in the baseline levels, we solely assumed a non-zero covariance between the immediate treatment effect and treatment effect on the time trend at both levels (i.e., σ_{u2u3} and σ_{v2v3}). We considered either no correlation ($\rho = 0$) or a large negative correlation ($\rho = - .7$,

Table 1 Starting occasion of treatment introduction

I	J			
		3 (case 1 to case 3)	4 (case 1 to case 4)	7 (case 1 to case7)
10	4, 6, 8		4, 5, 7, 8	4, 5, 5, 6, 7, 7, 8
20	7, 11, 15		7, 10, 12, 15	7, 9, 9, 11, 13, 13, 15
40	11, 21, 31		11, 18, 24, 31	11, 15, 15, 21, 27, 27, 31

Notes: I = number of measurement occasions; J = number of cases

corresponding to a covariance of -0.56 or -0.11 , depending on the chosen population variances) between these treatment effects. This rather large correlation was chosen to make a possible effect of this factor clearly visible. The value chosen was negative in order to simulate the situation described in the Introduction, where one may either show a larger immediate effect or a gradually increasing effect. Therefore, in Σ_u and Σ_v , all the off-diagonal elements of the covariance matrices were set to 0, except for σ_{u2u3} and σ_{v2v3} for which the values of 0, -0.56 , or -0.11 were used. Crossing the values of the diagonal and off-diagonal elements led to four combinations for the between-case and four for the between-study covariance matrix (i.e., relatively large variance with no covariance, relatively large variance with covariance, relatively small variance with no covariance, and relatively small variance with covariance).

In total, we evaluated $3 \times 3 \times 3 \times 4 \times 4 = 432$ conditions and for each combination, we simulated 2000 data sets, resulting in 864,000 data sets to be analyzed.

Data analysis

We fitted an OLS model to the generated raw data of each individual case, using Eq. 1, to estimate the regression coefficients representing the treatment effect sizes (i.e., the immediate effect and the effect on time trend, b_{2jk} and b_{3jk}) and the residual's standard deviation ($\hat{\sigma}_{ejk}$). These case-specific intervention effect size estimates were then standardized by dividing them by the estimated residual standard deviation (Van den Noortgate & Onghena, 2003b).

$$b'_{jk} = \frac{b_{jk}}{\hat{\sigma}_{ejk}} \tag{7}$$

For group-comparison designs, Hedges (1981) proposed to multiply observed standardized mean differences with an approximate bias correction factor, in order to reduce the small sample bias from standardized mean differences:

$$c(m) = 1 - \frac{3}{4m-1} \tag{8}$$

with m equal to the degrees of freedom, which is the number of participants minus two.

In a previous simulation study, we found that applying Hedges' (1981) bias correction to the standardized effect sizes leads to better estimates of overall treatment effects and variance components, especially for small I (Jamshidi et al., 2019). Therefore, in the current simulation, we synthesized the bias-corrected standardized regression coefficients. For SCED data, the degrees of freedom, m , is equal to the number of measurement occasions, I , minus the number of predictors, p , in the regression model (Eq. 1)

minus one (i.e., $m = I - p - 1$). According to Eq. 1, the degrees of freedom equals $I - 4$. The bias-corrected standardized effect size is:

$$\left(b'_{jk}\right)^c = b'_{jk} \left(1 - \frac{3}{4m-1}\right) \tag{9}$$

We also applied the bias correction factor to the standard error estimates and therefore to the sampling variance corresponding with each coefficient:

$$\left(\hat{\sigma}_{b'}^c\right)^2 = \widehat{\sigma}_{b'}^2 \left(1 - \frac{3}{4m-1}\right)^2 \tag{10}$$

We analyzed the bias-corrected standardized regression coefficients, i.e., $\left(b'_{2jk}\right)^c$ and $\left(b'_{3jk}\right)^c$ with three different models including a separate univariate three-level meta-analytic model for each type of effect, a multivariate three-level meta-analytic model ignoring the covariance structure at the case and study level, and a multivariate three-level meta-analytic model including the covariance structure at the case and study level. We used the last two models without and with RVE approach, so in total, five approaches were evaluated. In the following, "Model 1" refers to the two separate univariate three-level meta-analytic models for the immediate effect and the effect on time trend, using Eqs. 4 and 5. By analyzing the two types of effects separately, we avoid the issue of modeling dependence among these effect sizes in our analysis.

In the second approach, we applied a multivariate three-level meta-analytic model (later referred to as "Model 2"), proposed by Kalaian and Raudenbush (1996), Eq. 6, but ignoring the covariance structure at the case and study level. For doing so, first the Cholesky transformation (Kalaian & Raudenbush, 1996) was used to orthogonalize the regression coefficients for each case at the first level. Doing so, we could estimate the model parameters, using the sampling variance and covariance estimates obtained from the OLS regression analysis. A bivariate multilevel meta-analysis on these regression coefficients was then used, estimating the overall effect on level and on the time trend, as well as their variances at the case and study level. In a third approach (later referred to as "Model 2 with RVE"), we used the sandwich estimator for the standard errors of the overall treatment effect estimates. The possible advantage of using RVE is that it can return asymptotically consistent covariance matrix estimates, even if the underlying models are incorrect or misspecified (Huber, 1967; White, 1982). We were interested in evaluating the functionality of adding this estimator to our model. RVE only adapts the standard errors, not the estimates of the fixed parameters and variance components. Therefore, in the results section we only provide separate results for the second and third approach for the relative bias of standard errors estimates.

In the fourth approach, we used again a three-level multivariate model, but this time including covariances at the case and study level (“Model 3”). In the fifth and final approach (referred to as “Model 3 with RVE”), we used the same model but applied the sandwich estimator to the standard errors estimates of the overall treatment effects.

For all approaches, we made use of the restricted maximum likelihood procedure implement in PROC MIXED, using the Kenward-Roger approach to approximate the degrees of freedom for the t -tests of the fixed effects. We selected this method for estimating degrees of freedom, because it was found to perform better in synthesizing multiple-baseline studies through multilevel modeling with a complex error structure (Ferron et al., 2010).

To evaluate the performance of the five approaches for estimating and testing the overall treatment effects parameters (i.e., γ_{200} and γ_{300}), we used various criteria. The amount of absolute bias was approximated by subtracting the true value from the mean estimated value. We calculated the mean squared error (MSE) in order to check the accuracy of the estimates. Moreover, we looked at the estimates of the standard error (SE) of the overall treatment effects' estimates, that can be used to construct confidence intervals. The standard deviation of the treatment effects estimates can be assumed as an appropriate approximation of the true standard errors. The bias of the standard errors was therefore evaluated by the difference between the median standard error estimates and the standard deviation of the estimates of the treatment effects. Because the distributions of the standard errors tend to be skewed, we used the median rather than the mean of the standard errors. The coverage proportions (CPs) of the 95% confidence intervals of the overall treatment effects estimates were constructed to evaluate the treatment effects estimates and the estimated standard errors. Given that for each condition 2000 datasets are generated, the standard error of the estimate of the coverage proportion is approximately .005; therefore, a fairly good coverage proportion would be expected to have a value between .94 and .96. An appropriate coverage proportion of the intervals would be a good indicator of a lack of bias in both the parameter estimates and the corresponding standard errors.

We also looked at the performance of three-level models in estimating the variance components estimates ignoring and modeling the covariance at the case and study level. This way, we could evaluate the impact of the analysis approach on the between-case variance and between-study variance of the treatment effects in terms of accuracy of estimates (i.e., bias) and the precision of the estimates (i.e., MSE).

Afterwards, we calculated the relative bias of each parameter estimates (treatment effects, standard errors, and variance components) by dividing the estimated bias by its true value. We considered the cutoff proposed by Hoogland and Boomsma (1998) to evaluate the values of relative bias.

Values less than .05 are considered as acceptable for the fixed effect estimates and the values less than .10 are acceptable for the variance component estimates.

Comparing the results of these analyses in terms of bias, MSE, standard errors, and confidence interval coverage proportions can clarify the added value of performing a multivariate multilevel meta-analysis over separate univariate multilevel meta-analyses for each kind of regression coefficients. We did not expect major differences between these methods in terms of bias or confidence interval coverage proportions, but we did anticipate better precision (i.e., smaller MSE's and standard errors) when performing a multivariate meta-analysis as was found in Van den Noortgate et al. (2013).

Finally, we conducted ANOVA's to investigate which simulation factors affect the bias, MSE, the standard error, and the CP. In these analyses, main effect and first order interaction effects were included. In our discussion, we only focused on effects with eta squared (η^2) higher than .01. We chose this cutoff because these effects are relatively large compared to effects of other design factors, although based on the rule of thumb suggested by Cohen (1988) η^2 equal to .01 is considered as a small effect.

Results

Bias and MSE of the overall treatment effect estimates

We looked at the absolute bias of the overall treatment effects estimates (i.e., immediate treatment effect and treatment effect on time trend). According to an ANOVA, none of the design factors had a substantial impact on the bias of overall treatment effects estimates. We looked at the relative bias of the immediate effect estimates and it was close to zero across all models. The highest relative bias was found for Model 1 when the overall population effect is 2, resulting in a relative bias of 0.0010. The relative bias for Model 2 and Model 3 was 0.0008 and 0.0007 respectively.

We also looked at the MSE of the average treatment effect estimates, as an indicator of precision providing important information about bias and variance of the estimates. The smaller the values of MSE, the better the estimates. The ANOVA revealed that the number of studies at the highest level ($\eta^2 = .050$) and the between-study variance ($\eta^2 = .046$), were the design factors with substantial impact on the MSE of the estimated treatment effect and to a smaller extent the interaction of these factors ($\eta^2 = .013$). An overview of how the estimated MSE values is associated with the relevant simulation factors is shown in Table 2. There were no substantial differences between the underlying models in terms of MSE estimates. As we can see in Table 2, the largest MSE were obtained in conditions with large between-study variance and

Table 2 Mean squared error (MSE) of the estimated immediate effect ($\hat{\gamma}_{200}$)

	$\sigma_{v_2}^2$	$K = 10$	$K = 20$	$K = 40$
Model 1	0.5	0.196	0.097	0.049
	8	0.953	0.478	0.239
Models 2 and 3	0.5	0.191	0.095	0.048
	8	0.948	0.476	0.238

small number of studies. As the number of studies increased and the between-study variance decreased, the MSE estimates dropped dramatically. As the number of studies increased from 10 to 40, the values of MSE became four times smaller. The MSE estimates range from 0.016 up to 1.232 with the largest value when ten measurement occasions, three cases, ten studies, a large value of between-case variance (i.e., $\sigma_{u_2}^2 = 8$), and a large value of between-study variance (i.e., $\sigma_{v_2}^2 = 8$) are included.

K = number of studies; $\sigma_{v_2}^2$ = between-study variance; Model 1 = univariate three-level meta-analysis; Model 2 = multivariate three-level meta-analysis ignoring covariance; Model 3 = multivariate three-level meta-analysis modeling covariance.

Relative standard error bias and confidence interval coverage proportion

An ANOVA indicated that the relative standard error bias was mostly affected by the underlying model ($\eta^2 = .296$) and the number of studies meta-analyzed ($\eta^2 = .259$). Negative values of the relative bias indicate that the standard error estimates were slightly smaller than expected. The results in Table 3 reveal that in none of the conditions, the relative bias exceeds the .10 or -.10 cut-off score. The standard errors are more underestimated when the sandwich estimator procedure is applied, particularly when the number of studies is small. The results also indicate as the number of studies increases to 40,

Table 3 Relative difference between the median of the standard error estimates and the standard deviation of the estimated effects

	$K = 10$	$K = 20$	$K = 40$
Model 1	-0.024	-0.009	-0.003
Model 2	-0.022	-0.008	-0.002
Model 2 with RVE	-0.063	-0.023	-0.008
Model 3	-0.022	-0.009	-0.003
Model 3 with RVE	-0.063	-0.023	-0.008

Notes. K = number of studies; Model 1 = univariate three-level meta-analysis; Model 2 = multivariate three-level meta-analysis ignoring covariance; Model 3 = multivariate three-level meta-analysis modeling covariance

the relative bias is close to zero and negligible across all analysis models no matter how many cases and how many measurements are included.

We also looked at the confidence interval coverage proportions to evaluate the estimated standard errors. The estimated coverage proportions range from .93 to .96. As can be seen in Table 4, there are slight differences in coverage proportion of the 95% confidence intervals between different approaches. In general, the CPs are close to the nominal values in most of the conditions across different models, except in models applying sandwich estimator with a small number of studies ($K = 10$), where the CP is slightly below the nominal value due to higher underestimation of standard errors in this condition.

Variance components estimates

We evaluated the quality of the variance components estimates (i.e., the between-study and between-case variance) for both the immediate treatment effect and the treatment effect on time trend. We only report the results associated to the immediate treatment effect because similar results were obtained. The ANOVA indicated that none of the factors had a substantial effect on the bias of the between-study variance estimates. However, there was notable variation across conditions. The median relative bias across all design conditions was -0.059, -0.052, and -0.055 for Models 1, 2, and 3, respectively.

Table 5 gives an overview of the results across conditions. A negative relative bias in the estimated between-study variance was found in all conditions. From this table, we can deduce that the relative bias ranges from -0.017 to -0.388 when univariate multilevel meta-analysis applied. This range equals to -0.011 to -0.390 for the approach that ignored the covariances at the case and study level and -0.017 to -0.355 for the approach that modeled these covariances. These results reveal that the estimates are unbiased as the number of studies increase to 40 and the number of cases is either 4 or 7. Additionally, the estimates are almost unbiased when the

Table 4 Coverage proportion for the estimated immediate effect ($\hat{\gamma}_{200}$)

	$K = 10$	$K = 20$	$K = 40$
Model 1	0.9538	0.9523	0.9506
Model 2	0.9532	0.9524	0.9515
Model 2 with RVE	0.9399	0.9448	0.9471
Model 3	0.9531	0.9519	0.9508
Model 3 with RVE	0.9397	0.9447	0.9471

Notes. K = number of studies; Model 1 = univariate three-level meta-analysis; Model 2 = multivariate three-level meta-analysis ignoring covariance; Model 3 = multivariate three-level meta-analysis modeling covariance

Coverage proportions between .94 and .96 are in boldface

Table 5 Median of relative deviation of the between-study variance estimates of γ_{200} compared to the true value

	$\sigma_{u_2}^2$	$\sigma_{v_2}^2$	$K = 10$			$K = 20$			$K = 40$		
			$J = 3$	$J = 4$	$J = 7$	$J = 3$	$J = 4$	$J = 7$	$J = 3$	$J = 4$	$J = 7$
Model 1	0.5	0.5	-0.188	-0.163	-0.114	-0.093	-0.069	-0.055	-0.042	-0.039	-0.021
		8	-0.087	-0.082	-0.082	-0.048	-0.038	-0.039	-0.019	-0.019	-0.017
	8	0.5	-0.388	-0.360	-0.233	-0.174	-0.174	-0.124	-0.120	-0.079	-0.058
		8	-0.114	-0.100	-0.092	-0.064	-0.050	-0.045	-0.026	-0.024	-0.019
Model 2	0.5	0.5	-0.137	-0.130	-0.099	-0.064	-0.051	-0.041	-0.027	-0.027	-0.011
		8	-0.079	-0.080	-0.079	-0.044	-0.034	-0.036	-0.013	-0.014	-0.014
	8	0.5	-0.390	-0.378	-0.225	-0.172	-0.174	-0.121	-0.120	-0.083	-0.060
		8	-0.110	-0.099	-0.090	-0.062	-0.048	-0.042	-0.022	-0.021	-0.017
Model 3	0.5	0.5	-0.138	-0.137	-0.107	-0.073	-0.062	-0.050	-0.039	-0.039	-0.021
		8	-0.079	-0.082	-0.081	-0.047	-0.038	-0.039	-0.019	-0.019	-0.017
	8	0.5	-0.321	-0.355	-0.224	-0.154	-0.167	-0.122	-0.123	-0.085	-0.059
		8	-0.105	-0.095	-0.092	-0.064	-0.050	-0.045	-0.025	-0.025	-0.019

Notes. J = number of cases; K = number of studies; $\sigma_{u_2}^2$ = between-case variance; $\sigma_{v_2}^2$ = between-study variance; Model 1 = univariate three-level meta-analysis; Model 2 = multivariate three-level meta-analysis ignoring covariance; Model 3 = multivariate three-level meta-analysis modeling covariance. The values in boldface are larger than the .10 cutoff proposed by Hoogland and Boomsma (1998)

between-study variance is larger (i.e., $\sigma_{v_2}^2 = 8$) independent of the number of cases, the number of studies, and the value of between-case variance. The smallest values for the relative bias were found in conditions with a larger number of studies ($K = 40$), independent of other design factors. Across all models, the smaller relative bias was obtained when the between-study variance is larger.

The ANOVA showed that the MSE of between-study variance estimates are mainly affected by the number of studies ($\eta^2 = .021$) and the between-study variance ($\eta^2 = .033$), and to a minor extent by the interaction of these two factors ($\eta^2 = .011$). Table 6 displays the MSE estimates in terms of these effective design factors. The MSE estimates are particularly large in conditions with a larger between-study variance (i.e., $\sigma_{v_2}^2 = 8$). Closer inspection of the results indicates that the MSE decreases significantly as the number of studies K increases and the between-case and between-study variances decrease. The largest MSE was obtained in $K = 10, J = 4, I = 10, \sigma_{u_2}^2 = 8, \sigma_{v_2}^2 = 8$ condition. We also found that the smallest MSE occurs in conditions with a large number of studies, a large number of cases, more measurement occasions, and a small between-case and between-study variability. The results also reveal very similar MSE estimates across different approaches. This means that there is no clear advantage of modeling the covariances at higher levels, in terms of the precision of the estimated between-study variance.

We also looked at the relative bias in the estimate of the between-case variance of the immediate treatment effect. The preliminary ANOVA showed that the number of measurement occasions ($\eta^2 = .098$), the between-case variance ($\eta^2 = .052$), and the interaction of between-study variance and the number

of measurement occasions ($\eta^2 = .040$) are the most influential design factors explaining variability in between-case variance estimates. As shown in Table 7, the relative bias exceeds the cut-off of .10 when the population between-case variance is smaller (i.e., $\sigma_{u_2}^2 = 0.5$) regardless the values of number of measurement occasions and between-study variance. Specifically, the relative bias is highly larger in condition with smaller number of measurement occasions ($I = 10$). Furthermore, the relative bias estimates are smaller in

Table 6 Mean squared error (MSE) of the estimated between-study variance

	$\sigma_{u_2}^2$	$\sigma_{v_2}^2$	$K = 10$	$K = 20$	$K = 40$
Model 1	0.5	0.5	0.220	0.114	0.059
		8	17.545	8.185	4.013
	8	0.5	1.635	0.791	0.414
		8	25.844	12.308	6.091
Model 2	0.5	0.5	0.209	0.107	0.055
		8	17.462	8.137	4.004
	8	0.5	1.598	0.775	0.406
		8	25.794	12.293	6.084
Model 3	0.5	0.5	0.215	0.106	0.054
		8	17.626	8.099	3.972
	8	0.5	1.690	0.777	0.403
		8	26.029	12.236	6.037

Notes. K = number of studies; $\sigma_{u_2}^2$ = between-case variance; $\sigma_{v_2}^2$ = between-study variance; Model 1 = univariate three-level meta-analysis; Model 2 = multivariate three-level meta-analysis ignoring covariance; Model 3 = multivariate three-level meta-analysis modeling covariance

Table 7 Median of relative deviation of the between-case variance estimates of γ_{200}

	$\sigma_{u_2}^2$	$\sigma_{v_2}^2$	$I = 10$	$I = 20$	$I = 40$
Model 1	0.5	0.5	0.271	0.191	0.094
		8	1.928	0.692	0.308
	8	0.5	0.083	0.017	-0.005
Model 2	0.5	8	0.213	0.065	0.022
		0.5	2.083	0.532	0.177
	8	0.5	0.215	0.046	0.002
Model 3	0.5	8	0.347	0.095	0.029
		0.5	2.102	0.508	0.170
	8	0.5	0.201	0.037	-0.000
		8	0.331	0.085	0.026

Notes. I = number of measurement occasions; $\sigma_{u_2}^2$ = between-case variance; $\sigma_{v_2}^2$ = between-study variance; Model 1 = univariate three-level meta-analysis; Model 2 = multivariate three-level meta-analysis ignoring covariance; Model 3 = multivariate three-level meta-analysis modeling covariance

The values in boldface are larger than the .10 cutoff proposed by Hoogland and Boomsma (1998)

conditions with larger between-case variance and a number of measurement occasions larger than 10 (i.e., $I = 20$ or 40). In general, the univariate approach led to better estimation of the between-case variance compared to other approaches. The approaches of ignoring and modeling the covariances in second and third levels obtained very similar results.

According to the ANOVA, there were no design factors that substantially affected the MSE for between-case variance estimates. The highest η^2 was observed for the number of measurement occasions ($\eta^2 = .004$). As Table 8 shows, the MSE is noticeably large when the number of measurement occasions is 10. The MSE's highly drop as the number of measurement occasions increases. As was the case for the between-study variance, the MSE for the between-case variance was smallest when the number of observations, cases and studies was largest and the variances at the higher levels smaller.

Bias and MSE of between-study covariance and between-case covariance

The overall absolute bias of the estimated between-case covariance (i.e., σ_{u2u3}) equals 0.024 and ranges from - 0.008 to 0.090 depending on the condition. We did not look at relative bias for covariances because of having value of 0 for covariances and it was not possible to calculate relative bias for some conditions. According to the ANOVA, the number of measurement occasions ($\eta^2 = .013$) is related to the bias (with a higher I corresponding to less bias). The MSE values ranged

from zero to 0.679. The largest MSE was observed in $K = 10$, $J = 4$, $I = 10$, $\sigma_{u_2}^2 = 8$, $\sigma_{v_2}^2 = 8$ condition. Similar to the bias, the number of measurement occasions ($\eta^2 = .006$) appeared to be the factor most influencing variability of MSE estimates. As can be seen in Fig. 1, the bias and MSE is greatly reduced as the number of measurement occasions increases from 10 to 20.

The mean bias of the between-study covariance was equal to - 0.003 with the range between - 0.023 and 0.007 representing unbiased estimates. For the estimated between-study covariance, no substantial effects of the design factors on the bias were found. The MSE values ranged from 0.006 to 0.837 with the largest value in $K = 10$, $J = 4$, $I = 10$, $\sigma_{u_2}^2 = 8$, $\sigma_{v_2}^2 = 8$ condition. The ANOVA showed that the number of studies ($\eta^2 = .032$), the number of measurement occasions ($\eta^2 = .016$), and between-study variance ($\eta^2 = .016$) have significant effects on MSE estimates. Table 9 indicates that the smallest MSE values were observed in condition with the number of measurement occasions larger than 10 (i.e., $I = 20$ or 40) and the smaller between-study variance (i.e., $\sigma_{v_2}^2 = 0.5$).

In sum, the results indicate that the between-case and between-study covariance are well estimated in terms of accuracy and precision. Also estimating the between-case and especially the between-study variance, results in small and even negligible bias in almost all conditions. Our results also indicate that the approach modeling the covariances does not significantly outperform the approach ignoring the covariances at the case and study level.

Discussion

In the current simulation study, we studied the meta-analysis of multiple regression coefficients from SCEDs. We simulated SCED raw data with the covariance between the immediate treatment effect and the treatment effect on the time trend at the case and study level. Afterwards, we used five main approaches to handle the existing dependence between the main treatment effects (i.e., the immediate treatment effect and the treatment effect on the time trend).

Table 8 Mean squared error (MSE) of the estimated between-case variance

	$I = 10$	$I = 20$	$I = 40$
Model 1	10.850	2.241	1.527
Model 2	13.947	2.437	1.544
Model 3	13.586	2.344	1.531

Notes. I = number of measurement occasions; Model 1 = univariate three-level meta-analysis; Model 2 = multivariate three-level meta-analysis ignoring covariance; Model 3 = multivariate three-level meta-analysis modeling covariance

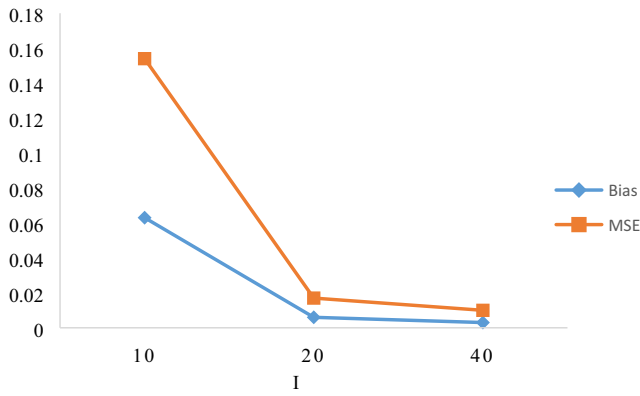


Fig. 1 Bias and MSE of the estimated between-case covariance estimates; I = number of measurement occasions

In the first approach, we ignored the dependence among the treatment effects by running a separate univariate three-level meta-analysis for each treatment effect. In the second approach, we used a multivariate model but ignored the existing simulated covariance at the case and study level. In the third approach, we used the same model but used a sandwich estimator for the standard errors of the effect estimates. The fourth approach used again a multivariate model but included the covariance between the treatment effects at higher levels. In the fifth approach, a sandwich estimator was applied for the standard errors for this multivariate model. The question was whether handling the dependency among multiple regression coefficients for each case within SCED studies in a three-level meta-analysis with these approaches affects the estimation of statistical properties of fixed effects and variance components.

Based on the reported results, we did not find substantial differences between the underlying models in terms of the bias and MSE of the treatment effect estimates. However, the univariate model (i.e., Model 1) led to slightly larger estimates of the overall treatment effects compared to other models. None of the simulation design factors had a substantial impact on the treatment effects estimates. In contrast, the number of studies, the between-study and the between-case variance were found

to have significant effects on MSE. The highest MSE values were obtained in the conditions with a larger between-study variance and a small number of studies. In terms of the standard errors of the treatment effects, the underlying models along with the number of studies and the interaction of number of measurement occasions, number of cases, and number of studies played a significant role in the variability of the estimates. The results indicate a greater underestimation when the sandwich estimator was applied to results from both approaches of modeling and when ignoring the covariances at higher levels. This in turn led to smaller coverage proportions for results from the models paired with the sandwich estimator, which was slightly smaller than the nominal coverage proportion values. In general, the overall treatment effects are estimated precisely and accurately when the covariance exists at the case and study level, independent of the underlying model fitted. The results indicate that the estimates of the treatment effects are robust even when the random effects in the model are misspecified through ignoring the existing covariances.

The between-study and between-case variances are not estimated precisely and accurately in certain conditions. We found underestimated between-study variances across all combinations of conditions. However, unbiased estimates were obtained for the between-study variance in most conditions. Biased between-study variance estimates were observed particularly in scenarios with 10 or 20 studies and large between-case variances. On the other hand, more bias was observed in the between-case variance estimates. Conditions with a larger between-case variance and 20 or 40 measurement occasions resulted in unbiased between-case variance estimates. Biased estimates were obtained in other conditions. In general, modeling the covariances at higher levels led to slightly less biased estimations of the between-case and between-study variances in conditions with a larger between-case variance. The reported results indicate that the approach that models the existing covariances does not outperform the approach that ignores these covariances. Also, in this simulation, we obtained biased and less precise estimates for the between-case variance compared to the estimates for the between-study variance.

These results are not completely in line with the findings in the study of Moeyaert et al. (2016), who found more bias for the between-study variance than the between-case variance. An explanation is that we simulated the covariance between the immediate treatment effect and the treatment effect on the time trend at the higher levels, whereas Moeyaert et al. (2016) generated the covariance between the baseline level and the treatment effect not between two dependent treatment effect sizes. Our results show that the degree of bias does not differ substantially across the models (i.e., the misspecified model that ignores the covariances at higher levels and the correctly specified model), and unbiased estimates are obtained for the

Table 9 Mean squared error (MSE) of the estimated between-study covariance

I	$\sigma_{v_2}^2$	$K = 10$	$K = 20$	$K = 40$
10	0.5	0.281	0.171	0.130
	8	0.443	0.301	0.227
20	0.5	0.092	0.081	0.077
	8	0.173	0.158	0.151
40	0.5	0.081	0.076	0.073
	8	0.161	0.153	0.148

Note. I = number of measurement occasions; K = number of studies; $\sigma_{v_2}^2$ = between-study variance

between-case and the between-study covariance in different conditions.

Limitations and recommendations

The current study suffers from some limitations. The generalizability of the findings in this study is one of the major concerns, although we generated the data considering certain realistic conditions that seem representative for the single-case experimental studies. We limited the simulated three-level data to data from MBDs. We also did not include a possible autocorrelation between the residuals at the first level although the existence of autocorrelation might be expected when a case is repeatedly measured. Furthermore, we only included the correlation between the regression coefficients associated with the treatment effects and set the correlation between other coefficients to zero to explore the performance of different models in simple scenarios. We only generated conditions with the covariances at the case and study level together. We did not simulate conditions with the covariances at only one of these levels.

Moreover, in this study we evaluated the performance of different models in terms of handling dependence among the treatment effects at higher levels applying the three-level meta-analytic approach. This is only one possible source of dependence among effect sizes. Dependence among effect sizes might occur due to several other reasons. Multiple effect sizes might be obtained by comparison of different treatments, by using multiple measures for the same construct or by using multiple outcome variables (Cheung, 2014; Van den Noortgate et al., 2013). Dependent effect sizes are less informative compared to independent effect sizes because they are calculated from common or related data and therefore give partly the same information. Thus, including dependent effect sizes in the meta-analysis as independent effects leads to using the same information multiple times that in turn, would overestimate confidence in the results of the meta-analysis. By treating dependent effect sizes this way, an underestimation of the standard errors is likely, resulting in too small confidence interval (Van den Noortgate et al., 2013). It might be interesting to investigate other sources of dependence among effect sizes from SCED studies in future studies.

This simulation further generated data by sampling residuals from normally distributed and homogeneous population distributions, and assuming that time trends are linear. Finally, we meta-analyzed only treatment effect sizes calculated as regression coefficients. It might be interesting if in future research other parametric and non-parametric effect sizes (Manolov & Moeyaert, 2017) are synthesized applying a multilevel meta-analytic approach. Some non-overlap indices such as percentage of non-overlapping data (PND; Parker, Vannest, & Davis, 2011; Scruggs, Mastropieri, & Casto, 1987), percentage of all non-overlapping data (PAND;

Parker, Hagan-Burke, & Vannest, 2007; Parker et al., 2011), and percentage of zero data (PZD; Campbell, 2004) or a standardized mean difference (SMD; Gingerich, 1984; Maggin et al., 2011) can be calculated instead of regression coefficients as the treatment effect sizes.

In summary, the findings in the current study indicate no substantial differences between the results of applying different three-level meta-analytic models, confirming the robustness of the three-level approach in handling the dependence among the treatment effects at higher levels even if the covariance structure is misspecified. Although biased variance components estimates were obtained in some conditions, the misspecification of the covariance structure did not lead to worse estimates. We recommend SCED researchers use a multivariate multilevel approach to model the existing dependence among multiple regression coefficients (i.e., treatment effect sizes). Using this approach gives the opportunity to model the covariance at higher levels when combining the results from the primary SCEDs.

Author Notes This project was supported in part by the Institute of Education Sciences, U.S. Department of Education, under Grant R305D150007. All the content is solely the responsibility of the authors and do not represent views of the Institute of Education Sciences, U.S. Department of Education. For the simulation, we used equipment at the VSC–Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government, Department EWI.

References

- Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, 28(2), 234–246. <https://doi.org/10.1177/0145445503259264>
- Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods*, 19(2), 211–229. <https://doi.org/10.1037/a0032968>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates Publishers.
- Fan, X., Felsovalyi, A., Sivo, S. A., & Keenan, S. C. (2002). *SAS® for Monte Carlo studies: A guide for quantitative researchers*. Cary: SAS Institute Inc.
- Farmer, J. L., Owens, C. M., Ferron, J. M., & Allsopp, D. H. (2010). *A methodological review of single-case meta-analyses*. Denver: Paper presented at the annual meeting of the American Educational Research Association.
- Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel-modeling approaches. *Behavior Research Methods*, 42(4), 930–943. <https://doi.org/10.3758/BRM.42.4.930>
- Gingerich, W. J. (1984). Meta-analysis of applied time-series data. *The Journal of Applied Behavioral Science*, 20(1), 71–79.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.2307/1164588>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1, 39–65. <https://doi.org/10.1002/jrsm.5>

- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26(3), 329–367. <https://doi.org/10.1177/0049124198026003003>
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165–179. <https://doi.org/10.1177/001440290507100203>
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. LeCam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 221–233). Retrieved from https://projecteuclid.org/download/pdf_1/euclid.bsmmsp/1200512988
- Jamshidi, L., Declercq, L., Fernández-Castilla, B., Ferron, J. M., Moeyaert, M., Beretvas, S. N., & Van den Noortgate, W. (2019). Bias adjustment in multilevel meta-analysis of standardized single-case experimental data. *Journal of Experimental Education*, 1–18. <https://doi.org/10.1080/00220973.2019.1658568>
- Jamshidi, L., Heyvaert, M., Declercq, L., Fernández-Castilla, B., Ferron, J. M., Moeyaert, M., ... Van den Noortgate, W. (2020). A systematic review of single-case experimental design meta-analyses: Characteristics of study designs, data and analyses. *Evidence-based Communication Assessment and Intervention*.
- Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, 1(3), 227–235. <https://doi.org/10.1037/1082-989X.1.3.227>
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/www/pdf/www_scd.pdf
- Maggin, D. M., Swaminathan, H., Rogers, H. J., O’Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology*, 49, 301–321. <https://doi.org/10.1016/j.jsp.2011.03.004>
- Manolov, R., & Moeyaert, M. (2017). Recommendations for choosing single-case data analytical techniques. *Behavior Therapy*, 48(1), 97–114. <https://doi.org/10.1016/j.beth.2016.04.008>
- Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2013). The three-level synthesis of standardized single-subject experimental data: A Monte Carlo simulation study. *Multivariate Behavioral Research*, 48(5), 719–748. <https://doi.org/10.1080/00273171.2013.816621>
- Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2014). Three-level analysis of single-case experimental data: Empirical validation. *The Journal of Experimental Education*, 82(1), 1–21. <https://doi.org/10.1080/00220973.2012.745470>
- Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2016). The misspecification of the covariance structures in multilevel models for single-case data: A Monte Carlo simulation Study. *Journal of Experimental Education*, 84(3), 473–509. <https://doi.org/10.1080/00220973.2015.1065216>
- Onghena, P., & Edgington, E. S. (2005). Customization of pain treatments: Single-case design and analysis. *The Clinical Journal of Pain*, 21(1), 56–68; discussion 69–72. <https://doi.org/10.1097/0002508-200501000-00007>
- Owens, C. M., & Ferron, J. M. (2012). Synthesizing single-case studies: A Monte Carlo examination of a three-level meta-analytic model. *Behavior Research Methods*, 44(3), 795–805. <https://doi.org/10.3758/s13428-011-0180-y>
- Parker, R. I., Hagan-Burke, S., & Vannest, K. J. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, 40(4), 194–204. <https://doi.org/10.1177/00224669070400040101>
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: a review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303–322. <https://doi.org/10.1177/0145445511399147>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models : Applications and data analysis methods* (2nd ed.). London: SAGE Publications Inc.
- SAS Institute, I. (2017). *Base SAS® 9.4 procedures guide* (7th ed.). Cary, NC: SAS Institute Inc.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, 8(2), 24–33. <https://doi.org/10.1177/074193258700800206>
- Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43, 971–980. <https://doi.org/10.3758/s13428-011-0111-y>
- Ugille, M., Moeyaert, M., Beretvas, S. N., Ferron, J. M., & Van den Noortgate, W. (2012). Multilevel meta-analysis of single-subject experimental designs : A simulation study. *Behavior Research Methods*, 44(4), 1244–1254. <https://doi.org/10.3758/s13428-012-0213-1>
- Ugille, M., Moeyaert, M., Beretvas, S. N., Ferron, J. M., & Van den Noortgate, W. (2014). Bias corrections for standardized effect size estimates used with single-subject experimental designs. *The Journal of Experimental Education*, 82(3), 358–374. <https://doi.org/10.1080/00220973.2013.813366>
- Van den Noortgate, W., López-López, J. A., Marin-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45, 576–594. <https://doi.org/10.3758/s13428-012-0261-6>
- Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, 18(3), 325–346. <https://doi.org/10.1521/scpq.18.3.325.22577>
- Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, 35(1), 1–10. <https://doi.org/10.3758/BF03195492>
- Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention*, 2(3), 142–151. <https://doi.org/10.1080/17489530802505362>
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1–25.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.