






The Impact of Ignoring Random Slopes in a $1 \rightarrow 1 \rightarrow 1$ Mediation Model in the Presence of Upper-Level Mediator-Outcome Confounding

Jasper Bogaert¹ , Wen Wei Loh² , Beatrijs Moerkerke¹ , Yves Rosseel¹ ,
Tom Loeys¹ 

[1] *Department of Data Analysis, Ghent University, Ghent, Belgium.* [2] *Department of Methodology & Statistics, Maastricht University, Maastricht, the Netherlands.*

Methodology, 2025, Vol. 21(4), 313–345, <https://doi.org/10.5964/meth.18093>

Received: 2025-05-19 • **Accepted:** 2025-10-13 • **Published (VoR):** 2025-12-18

Handling Editor: Eduardo Estrada, Autonomous University of Madrid, Madrid, Spain.

Corresponding Author: Tom Loeys, Department of Data Analysis, Ghent University, Henri Dunantlaan 1, 9000 Ghent, Belgium. tom.loeys@ugent.be.

Supplementary Materials: Code, Data, Materials [see [Index of Supplementary Materials](#)]



Abstract

In behavioral sciences, researchers frequently employ mediation analysis with longitudinal data. A common scenario involves the $1 \rightarrow 1 \rightarrow 1$ mediation model, where the Predictor X , Mediator(s) M , and Outcome Y are all measured at different occasions (Level 1) within individuals (Level 2). The standard $1 \rightarrow 1 \rightarrow 1$ mediation model approach fits two multilevel models, one for the mediator and one for the outcome, with two random intercepts and three random slopes in total. However researchers often exclude random slopes from multilevel models and only include random intercepts to account for non-independence across observations of the same individual. We demonstrate that ignoring random slopes in the $1 \rightarrow 1 \rightarrow 1$ mediation model can result in biased average indirect effect estimators, as well as underestimated standard errors. We provide code from open source and free statistical software that can be used by practitioners to fit the $1 \rightarrow 1 \rightarrow 1$ mediation model.

Keywords

multilevel mediation, lower level mediation, unmeasured confounding, random slopes

In behavioral sciences, researchers are often interested in investigating the underlying causal mechanisms of treatment or intervention effects on an outcome through media-



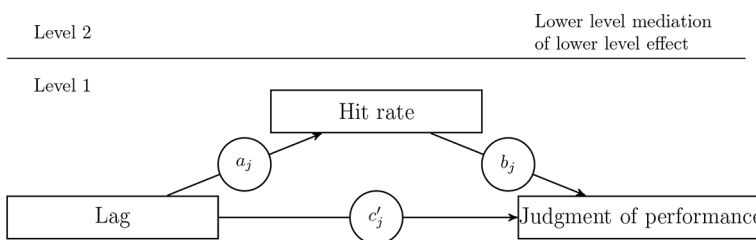
This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), [CC BY 4.0](#), which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

tion analysis. The total effect, encoding the effect of the intervention X on the outcome Y , is then decomposed into a direct and an indirect effect. The direct effect represents the part of the effect of the predictor variable X on the outcome Y that is not transmitted by the mediator M , while the indirect effect is the part of the effect of X on Y that is transmitted through M (MacKinnon, 2012).

Realistic research settings in behavioral sciences often involve hierarchical data structures such as, (i) data nested within groups (clustered data), and (ii) repeated measures from the same subject (longitudinal data) (West et al., 2022). Multilevel mediation analysis occurs when conducting mediation analysis with such hierarchical data. Various multilevel mediation exist depending on the level at which the predictor, mediator(s) and outcome are measured¹. A common setting is a $1 \rightarrow 1 \rightarrow 1$ mediation model, where the intervention (X), mediator (M), and outcome (Y) are all measured at Level 1 but clustered within Level 2. Examples of studies including such a model are Fitzsimmons-Craft et al. (2016), Hill et al. (2021), Tan et al. (2022), and Yu and Hu (2022). Figure 1 illustrates a $1 \rightarrow 1 \rightarrow 1$ mediation model, first presented by Vuurre and Bolger (2018) for a computer game study by Metcalfe and Greene (2007). Participants played a computer game catching Xs and avoiding Os while rating their perceived performance. The experiment manipulated the game by introducing a brief delay (250 ms) between mouse movement and cursor response (intervention X). The mediation question is whether this manipulation affects performance ratings indirectly, by first reducing actual performance, i.e: the hit rate (mediator M), which then influences participants' judgments of performance (outcome Y). All 3 variables were repeatedly measured within $N = 43$ independent individuals ($j = 1, \dots, N$), four times in the lag condition, and four times in a control condition. Since we have longitudinal data within individuals, we can allow for heterogeneity in the effects between individuals (and hence the subscript j).

Figure 1

Computer Game Example of Metcalfe and Greene (2007): A $1 \rightarrow 1 \rightarrow 1$ Mediation Model, First Presented by Vuurre and Bolger (2018)



1) We refer to Bauer et al. (2006) and Preacher et al. (2010) for an overview.

How can the mediation model depicted in [Figure 1](#) be formalized? Throughout this paper, we will assume linear relationships. For convenience, we will only consider the simple setting where no other (measured) variables are considered, although in practice, adjusting for confounders for the different relationships may be necessary to obtain causal effects. Following the approach of [Kenny et al. \(2003\)](#), the $1 \rightarrow 1 \rightarrow 1$ mediation model can be written as two Level 1 equations:

$$\begin{aligned} M_{ij} &= d_{Mj} + a_j X_{ij} + e_{Mij} \\ Y_{ij} &= d_{Yj} + b_j M_{ij} + c'_j X_{ij} + e_{Yij} \end{aligned} \quad (1)$$

where the subscript i refers to the lower-level units (i.e., the measurement occasions), the subscript j refers to the upper level units (i.e., the individuals), and e_{Mij} and e_{Yij} are respectively the Level 1 residuals for M and Y . The other five terms in the model are intercepts and slope coefficients, which can be interpreted almost similarly as in the standard linear regression setting with independent observations. Here, however, the parameters are random (as shown by the subscript j), implying that they can vary across the individuals. As in the single level model (but now with subscript j added), the direct effect in individual j is denoted by c'_j , the effect from X on M by a_j , and the effect from M on Y by b_j . In the next section, we provide an overview of the specific assumptions that must be fulfilled in order to obtain unbiased estimates and ensure valid inference for the effects in this model.

Model (1) includes two random intercepts and three random slopes. These random slopes are depicted in [Figure 1](#) as circles. It is important to note that both [Kenny et al. \(2003\)](#) and [Bauer et al. \(2006\)](#) consider all five random effects of Model (1). Hence, it could be argued that fitting all random effects is the default way to analyze a $1 \rightarrow 1 \rightarrow 1$ model. However, in multilevel modeling, researchers often exclude random slopes from the model and only include random intercepts to account for the non-independence across observations of the same individual ([Heisig et al., 2017](#); [Schielzeth & Forstmeier, 2009](#)). Several reasons for ignoring random slopes can be found in the literature. First, there may be no expected variation in the effect between Level 2 units, which implies that there is no random slope. However, [LaHuis et al. \(2020\)](#) suggested that it is rarely the case that the variance for a slope is exactly zero when working with hierarchical data. [Heisig et al. \(2017\)](#) also demonstrated that the assumption of invariant effects of lower-level control variables is often implausible. Second, there may be insufficient information to consider all random effects, for example when limited data is available. Challenges related to limited sample sizes can emerge at distinct levels. While small cluster sizes (i.e. a small number of measurement occasions) are problematic, especially with small to medium samples sizes at the higher level, the primary concern in multilevel modeling typically pertains to the number of individuals or clusters ([Hox & McNeish, 2020](#); [McNeish, 2014](#)). Increasing the number of individuals or clusters is often not

straightforward, as it may entail a higher financial cost and more time (Maas & Hox, 2005). When limited information is available from the data, convergence issues may occur (Bryan & Jenkins, 2016; Schielzeth & Forstmeier, 2009). Additionally, fitting (multiple) random slopes could be hindered by software limitations or insufficient computational resources (Heisig & Schaeffer, 2018). A third explanation may be that researchers have limited knowledge of multilevel modeling and are unaware of the consequences of misspecifying all the necessary random components. As noted by Schielzeth and Forstmeier (2009), there is a prevalent misconception that utilizing a multilevel model with only a random intercept adequately addresses all instances of non-independence in the data.

The consequences of ignoring random slopes have already been studied in the multilevel modeling literature. Several scholars have shown that falsely assuming fixed slopes rather than random slopes results in biased standard errors and overly narrow confidence intervals (i.e., anti-conservative inference) for the fixed effects (Barr et al., 2013; Bell et al., 2019; Berkhof & Kampen, 2004; Bryan & Jenkins, 2016; Heisig et al., 2017; Hoffman & Walters, 2022; Schmidt-Catran & Fairbrother, 2016). To the best of our knowledge, the impact of omitting random slopes in the more complex $1 \rightarrow 1 \rightarrow 1$ model has not been studied yet. There are two important aspects of this model that have not been thoroughly investigated in the literature under the considered scenario. First, this setting includes jointly fitting two models with five possibly correlated random effects. Second, as we will explain later, the expression for the average indirect in the $1 \rightarrow 1 \rightarrow 1$ model includes an additional term involving the covariance of random slopes. Therefore, the aim of this article is to examine in detail the impact of ignoring random slopes in the $1 \rightarrow 1 \rightarrow 1$ mediation model. More specifically, simulation studies based on the joint modeling approach described by Bauer et al. (2006) will be used to study the impact of omitting the random slopes. In the simulation study, we first assess the impact of model complexity on the convergence rate when using the default options in open source software. Second, given the mediation context, we focus on the bias, precision, efficiency, and coverage probability of the average indirect effect estimator.

The outline of this article is as follows. First, we provide a detailed overview of the joint modeling approach proposed by Bauer et al. (2006), and explain how the average indirect effect can be derived in the $1 \rightarrow 1 \rightarrow 1$ setting. Next, we describe the results from the simulation study. Finally, we summarize the key findings and provide some concluding remarks.

Joint Modeling Approach

In this section, we present a detailed overview of the joint modeling approach proposed by Bauer et al. (2006). We begin by outlining the $1 \rightarrow 1 \rightarrow 1$ mediation model and its underlying assumptions. Next, we discuss how to restructure the data and fit the model

using free statistical software. Finally, we describe the closed-form expressions for the indirect and total effects.

Model and Assumptions

Consider $1 \rightarrow 1 \rightarrow 1$ model presented in Equation 1. Bauer et al. (2006) make the following assumptions:

1. The predictors are uncorrelated with the random effects and the residuals, both within and across equations.
2. Both Level-1 residuals e_{Mij} and e_{Yij} are assumed to be uncorrelated² and normally distributed with a mean of zero. The residuals in each equation are expected to be independent and homoscedastic across individual observations (i) within clusters (j).
3. The random effects are normally distributed with mean their respective average population effect (or zero if notation in deviation form is used). The random effects may covary with each other.
4. The Level-1 residuals are uncorrelated with the random effects both within and across equations (i.e., e_{Mj} and e_{Yj} are uncorrelated with d_{Mj} , a_j , d_{Yj} , b_j , and c_j).
5. It follows from Assumptions 2 and 3 that the distributions of M and Y are conditionally normal. More specifically, M is normally distributed conditional on X ; Y is normally distributed conditional on M and X .

Kenny et al. (2003) separately estimate the two Level-1 models presented in Equation 1. However, as pointed out by Bauer et al. (2006), this approach results in a loss of essential information. Indeed, when fitting both models separately, the correlation between the random effects is not modeled explicitly (cfr. Assumption 3). It is important to understand the consequences of this. The random intercepts d_{Mj} and d_{Yj} in Model 1 represent unmeasured Level-2 heterogeneity in the mediator and outcome, respectively. In a standard multilevel model random intercepts are assumed to be independent from the predictors. Hence, the unmeasured heterogeneity in Y is assumed to be independent from X and M , and cannot capture unmeasured upper level common causes of M and Y . When assuming correlated random intercepts though, we explicitly allow for unmeasured upper level common causes of M and Y . To eliminate upper level confounding, it is sometimes argued to separate within and between effects. By doing so, the cluster-mean centered predictors will be independent from any upper level variable. While this is an interesting alternative, we will further built on the framework of Bauer et al. (2006), and capture any upper level confounding of M and Y by allowing the random intercepts to be correlated. In the next section, we explain how those two models can be fitted jointly.

2) As noted by one of the reviewers, the assumption of independent residuals could be relaxed. For example, in our illustration there might be residual temporal correlation, see Bauer et al. (2006).

In this paper, we focus on the 1 → 1 → 1 mediation model under the assumption that (unmeasured) upper-level confounding exists only for the mediator–outcome relationship. In other words, we do not account for (unmeasured) upper-level confounding of the exposure–mediator or exposure–outcome relationships. In our computer game example, the exposure is randomized, which satisfies this assumption. This condition is also commonly met in alternating treatment designs (Manolov et al. 2022), used in single-subject research (Shadish et al., 2013) for example, but it may be violated in observational settings where the exposure is not randomized. The primary aim of the paper is to show that ignoring random slopes in 1 → 1 → 1 mediation models can lead to biased estimates of the average indirect effect. In the discussion, we outline possible extensions that relax this simplifying assumption by allowing for upper-level confounding beyond the mediator–outcome relationship.

Reorganizing the Data

Bauer et al. (2006) proposed a solution that involves formulating the model jointly in a single Level 1 equation using selection variables. This requires creating a new outcome variable Z by stacking Y and M for each observation i within cluster j . To distinguish between M and Y , two selection variables (S_M and S_Y) are created. If Z pertains to M , then $S_M = 1$ and $S_Y = 0$. Otherwise, if Z pertains to Y , then $S_M = 0$ and $S_Y = 1$. S_M is associated with the model for the mediator M , while S_Y is associated with the model for the outcome Y . In the stacked data set, the predictor and mediator are retained, along with the new outcome variable and the two selection variables.

Fitting the Joint Model

By using the selection variables, the 1 → 1 → 1 model (see Equation 1) can be specified in a single equation:

$$Z_{ij} = S_{M_{ij}}(d_{M_j} + a_j X_{ij}) + S_{Y_{ij}}(d_{Y_j} + b_j M_{ij} + c'_j X_{ij}) + e_{Z_{ij}}. \quad (2)$$

As explained by Bauer et al. (2006), the joint model for Z (see Equation 2) can be fitted by specifying a model for Z without a common fixed intercept, but with random intercepts for S_M and S_Y (denoted as d_{M_j} and d_{Y_j} , respectively) and with random slopes for the product variables $S_M X$, $S_Y M$, and $S_Y X$ (a_j , b_j , and c'_j , respectively).³ Furthermore, the residual variance (denoted as $\text{Var}(e_{Z_{ij}})$) is allowed to differ depending on one of the selection variables (e.g., S_M) (Bauer et al., 2006). By fitting the joint model, the 1 → 1 → 1

3) As noted by an anonymous reviewer, the common fixed intercept is removed in this joint model. Instead the main effects of the selection variables ($S_{M_{ij}}$ and $S_{Y_{ij}}$) become the fixed intercepts for each outcome. Similarly, there is no common random intercept or common random slope variance, but instead the random effect variances are outcome-specific.

mediation models (see [Equation 1](#)) can be estimated simultaneously, and both the complete covariance matrix of the random effects and the complete covariance matrix of the fixed effects can be directly estimated ([Bauer et al., 2006](#)). As we will show in the next section, all the necessary information to estimate the average indirect and total effects becomes available with this stacked approach.

Estimating and Evaluating the Indirect and Total Effects

This section discusses the expressions for the average indirect and total effects as presented by [Bauer et al. \(2006\)](#) and [Kenny et al. \(2003\)](#), as well as their precision (i.e., the standard error) and expressions for their 95% confidence intervals.

The Average Indirect and Total Effects

While cluster-specific indirect and total effects can easily be obtained from [Equation 1](#), expressions for the average indirect and total effects require more attention. The average indirect effect ([Goodman, 1960](#), p. 712) can be derived as:

$$E(a_j b_j) = ab + \sigma_{a_j b_j}, \quad (3)$$

where a denotes the average effect of X on M , b denotes the average effect of M on Y , and $\sigma_{a_j b_j}$ represents the covariance between the two random slopes. The average total effect is given by ([Kenny et al., 2003](#)):

$$E(a_j b_j + c'_j) = ab + \sigma_{a_j b_j} + c', \quad (4)$$

with c' representing the average direct effect of X on Y .

Since a_j and b_j are not necessarily independent (cfr. the third assumption), the average indirect and total effect do not only depend on the average effect of X on M (a) and the average effect of M on Y (b), but also on the covariance between the two random slopes $\sigma_{a_j b_j}$ ([Kenny et al., 2003](#)).⁴ This implies that ignoring the presence of correlated random slopes may lead to biased estimators for the average indirect and total effects ([Tofighi et al., 2013](#)). In practice, sample estimates of the quantities in [Equation 3](#) and [Equation 4](#) are used as substitutes for their population counterparts.

Precision of the Estimated Indirect and Total Effects

This section aims to quantify the precision of the estimated average indirect and total effects. As was pointed out by [Bauer et al. \(2006\)](#) the precision of the estimated average effects should not be confused with the heterogeneity in the effects. The precision refers

4) As explained by [Tofighi et al. \(2013\)](#), a non-zero covariance between the random slopes may be a result of an omitted Level-2 variable that serves as a moderator of both the $M-X$ and the $M-Y$ associations.

to the variability in the estimators of the average effects, while heterogeneity pertains to the variability across Level-2 units. In this article, we focus on the precision of the average effects estimators. Bauer et al. (2006) showed that, assuming multivariate normality of the estimators, an expression for the variance of the estimated average indirect effect is given by

$$\text{Var}(\hat{a}\hat{b} + \hat{\sigma}_{a,b_j}) = b^2\text{Var}(\hat{a}) + a^2\text{Var}(\hat{b}) + \text{Var}(\hat{a})\text{Var}(\hat{b}) + 2ab\text{Cov}(\hat{a}, \hat{b}) + \text{Cov}(\hat{a}, \hat{b})^2 + \text{Var}(\hat{\sigma}_{a,b_j}), \tag{5}$$

whereas the variance of the estimated total effect is given by

$$\text{Var}(\hat{a}\hat{b} + \hat{\sigma}_{a,b_j} + \hat{c}') = b^2\text{Var}(\hat{a}) + a^2\text{Var}(\hat{b}) + \text{Var}(\hat{a})\text{Var}(\hat{b}) + 2ab\text{Cov}(\hat{a}, \hat{b}) + \text{Cov}(\hat{a}, \hat{b})^2 + \text{Var}(\hat{c}') + 2b\text{Cov}(\hat{a}, \hat{c}') + 2a\text{Cov}(\hat{b}, \hat{c}') + \text{Var}(\hat{\sigma}_{a,b_j}). \tag{6}$$

The (co)variances in Equation 5 and Equation 6 represent the (co)variances of the fixed effect estimates \hat{a} , \hat{b} , \hat{c}' , and the covariance estimate $\hat{\sigma}_{a,b_j}$ for the random slopes. These (co-)variances can be estimated by substituting the population values in Equation 5 and Equation 6 with their sample estimates (Raudenbush & Bryk, 2002). Again, the covariance between fixed effect estimators residing in two different models, $\text{Cov}(\hat{a}, \hat{b})$, is only possible by simultaneously estimating the 1 → 1 → 1 mediation model (Bauer et al., 2006). Additionally, it should be noted that Equation 5 and Equation 6 include the term $\text{Var}(\hat{\sigma}_{a,b_j})$.

Confidence Interval of the Average Indirect and Total Effects

Bauer et al. (2006) also presented expressions for 95% confidence intervals (CIs) for the average indirect and total effects. These expressions assume that the sampling distributions of the estimators are normal. Under this normality assumption, the 95% CIs for the average indirect effect and the average total effect can be calculated as:

$$(\hat{a}\hat{b} + \hat{\sigma}_{a,b_j}) \pm 1.96\sqrt{\widehat{\text{Var}}(\hat{a}\hat{b} + \hat{\sigma}_{a,b_j})} \tag{7}$$

and

$$(\hat{a}\hat{b} + \hat{\sigma}_{a,b_j} + \hat{c}') \pm 1.96\sqrt{\widehat{\text{Var}}(\hat{a}\hat{b} + \hat{\sigma}_{a,b_j} + \hat{c}')}, \tag{8}$$

with $\widehat{\text{Var}}$ denoting the estimated sampling variances obtained by plugging in sample estimates for their unknown population values (see Equation 5 and Equation 6). Importantly, when correlated random slopes are present but by ignoring $\hat{\sigma}_{a,b_j}$, one may obtain

shifted and narrowed CIs due to bias in the estimated average indirect effect and the underestimated variability.

Furthermore, [Bauer et al. \(2006\)](#) noted that a Wald-based null hypothesis test could be performed similarly by dividing the estimate of the effect by its standard error. However, since $\hat{a}\hat{b}$ is a product of normally distributed estimators, and the product of the normally distributed variables does typically not follow a normal distribution ([Aroian et al., 1978](#)), the normality assumption required for hypothesis testing and construction of the CIs may not hold exactly ([Shrout & Bolger, 2002](#)). However, the deviation from normality may be small enough that the CIs or significance tests will still be reasonably accurate ([Bauer et al., 2006](#)). In the next paragraph we present an alternative approach for the construction of confidence intervals that was also suggested by [Bauer et al. \(2006\)](#).

Monte Carlo Method to Construct Confidence Intervals

A viable alternative for constructing CIs is the Monte Carlo (MC) method, which does not assume that the product $\hat{a}\hat{b}$ follows a normal distribution. Instead, the method only relies on the assumption that the estimators \hat{a} and \hat{b} follow a bivariate distribution ([MacKinnon et al., 2004](#)). Other distribution-free alternatives, such as non-parametric and residual bootstrapping, are also available. However, there is no consensus on the best way to perform bootstrapping methods in the multilevel context, and multilevel modeling software does not always facilitate their application ([Preacher et al., 2010](#); [Preacher & Selig, 2012](#)). We therefore opted to present the MC method here since it is computationally faster and has been shown to perform comparable to other approaches ([Preacher & Selig, 2012](#)). The MC method for a $1 \rightarrow 1 \rightarrow 1$ mediation model is implemented in a free online R calculator, see [Preacher and Selig, \(2010\)](#).

Simulation Study

Simulation Model and Conditions

We examined the impact of ignoring random slopes in the $1 \rightarrow 1 \rightarrow 1$ mediation model in the presence of unmeasured upper-level mediator-outcome confounding. We focused on conditions commonly encountered in studies involving multilevel (mediation) models, using varying cluster sizes ($N_1 = 20, 50$) and numbers of clusters ($N_2 = 20, 50, 100$), along with different values for the intraclass correlation coefficient (ICC) to represent different clustering effects. Three scenarios were considered regarding the ICC, involving both equal and unequal values for the ICC of X , M , and Y . Specifically, we examined: (i) a lower ICC for the predictor ($ICC_X = 0.1, ICC_M = 0.2, ICC_Y = 0.3$), (ii) equal ICC values ($ICC_X = 0.2, ICC_M = 0.2, ICC_Y = 0.2$), and (iii) a higher ICC for the predictor ($ICC_X = 0.3, ICC_M = 0.2, ICC_Y = 0.1$). The ICC was computed by dividing the cluster-level random intercept variances of X , M , and Y by the total variance for each variable

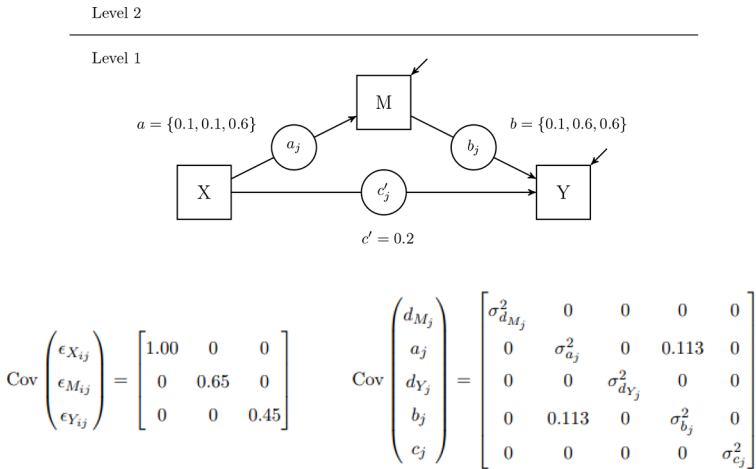
(i.e., the sum of the Level 1 residual variances and the cluster level intercept variance), following the approach of Zigler and Ye (2019). These values for the number of clusters, cluster sizes, and ICC values have been widely used in the literature and cover a range of conditions used in previous work (see e.g., Baird & Maxwell, 2016; Hox & Maas, 2001; Kim & Hong, 2020; Preacher et al., 2011; Zhang et al., 2009; Zigler & Ye, 2019).

The data generating model for the simulation study is represented by Model (1) and used various population values for the fixed and random effects (co-)variances. The fixed effects for a and b were assumed to be equal, and both set to either 0.1 or 0.6. Additionally, we considered a scenario in which the fixed effects were unequal, with $a = 0.1$ and $b = 0.6$. The population value for the fixed effect c' (i.e., the direct effect) was set to 0.2. The variances of the random slopes of the a and b paths ($\sigma_{a_j}^2$ and $\sigma_{b_j}^2$) were both chosen to be 0.01, 0.15, and 0.50. Additionally, we examined a scenario in which the variances of the random slopes were unequal, with $\sigma_{a_j}^2 = 0.15$ and $\sigma_{b_j}^2 = 0.50$. To evaluate the performance of the approaches in a model without random slopes, we included a condition with minimal non-zero variances for $\sigma_{a_j}^2$ and $\sigma_{b_j}^2$ (i.e., $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.01$), which we refer to as the 'no random slopes' condition. The covariance between the random slopes of both paths (σ_{a_j, b_j}) was set to -0.113 , 0 , and 0.113 .⁵ However, when the data generating model contained no random slopes for the a and b paths (i.e., $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.01$), then σ_{a_j, b_j} was set to zero. For simplicity, all other random effects were assumed to be uncorrelated, while the variances of the other random effects were based on the values used by Bauer et al. (2006). An overview of all parameters in the data generation models can be found in Figure 2.

5) When $\sigma_{a_j, b_j} = 0.113$, the resulting values for the random slope correlation (ρ_{a_j, b_j}) were 0, 0.75, 0.41, and 0.22. Given that the amount of bias in the average indirect when ignoring random slopes depends on the covariance between the random slopes of both paths (σ_{a_j, b_j}), we focused on the covariance rather than the correlation.

Figure 2

Simulation Model and Covariance Matrices of Level 1 Residual Variances (left) and Random Effects (right)



Note. The values of $\sigma_{d_M}^2$ and $\sigma_{d_Y}^2$ were chosen based on the clustering effect. The random slope variances σ_a^2 and σ_b^2 were set to 0.01, 0.15 and 0.5.

For each of the 648 conditions (i.e., 2 different cluster sizes \times 3 different numbers of clusters \times 3 scenarios for the ICC values \times 3 scenarios for a and $b \times$ 4 scenarios for $\sigma_{a_j}^2$ and $\sigma_{b_j}^2 \times$ 3 different values for the random slope covariance), 1000 samples were generated and analyzed using the statistical software program R (Version 4.2.1) (R Core Team, 2022) and the lme4 package (Bates, Mächler et al., 2015). In order to examine the impact of ignoring random slopes, we fitted a joint model with random slopes (further abbreviated *RS*) and without random slopes (further abbreviated *RI*). To assess the convergence rate, we initially used the default optimizer and the default number of iterations from lme4 for both the *RI* and *RS* approaches. However, the *RS* approach exhibited high non-convergence rates under various conditions when using the default optimizer. To address this, we tested the bobyqa optimizer, which demonstrated substantially higher convergence rates.⁶ The results presented in the manuscript are therefore based on the bobyqa optimizer.

Additionally, we identified singular solutions for the *RS* approach. Below, we present results for the 1000 samples (hereafter referred to as *RS all*) and for the subset excluding

6) When we changed to the bobyqa optimizer for the *RS* approach, a higher convergence rate was observed in 69 out of the 72 considered conditions. In the worst-case scenario, the bobyqa optimizer achieved a convergence rate that was 6% lower, while in the best-case scenario, it achieved a convergence rate that was 60% higher compared to the default optimizer. The largest differences between the two optimizers were observed when the cluster size ($N_1 = 50$), number of clusters, and random slope variances were large, or in the absence of random slopes.

these singular fits (hereafter referred to as RS). As explained by Bates, Kliegl et al. (2015); Bates, Mächler et al. (2015), singular fits are common with complex covariance models and occur when a random effect variance is estimated near zero.⁷ As it has been recommended that robust standard errors may be used to overcome model misspecification in multilevel models McNeish (2023), we considered both the regular (hereafter referred to as RI) and robust standard errors (hereafter referred to as *RI robust*) for the RI approach. The robust standard errors were obtained using the clubSandwich package (Pustejovsky, 2023). An illustration of the examined estimation approaches is provided in Appendix A.

Performance Criteria

Several performance criteria were employed to evaluate the impact of ignoring random slopes in the 1 → 1 → 1 mediation model. The first criterion was non-convergence, which occurs when the model-fitting algorithm fails to achieve a solution within the pre-defined number of iterations or stopping criteria (Park et al., 2020). Possible convergence issues might arise for the RS approach, especially in conditions with a smaller cluster size and/or number of clusters. The convergence rate (CR) for each condition is calculated as:

$$CR = \frac{\text{\#converged solutions}}{R}, \quad (9)$$

with R the number of generated data sets.

Second, we assessed the bias of the estimated average indirect effect. Since boxplots of the estimated average indirect effect (see Appendix B) revealed several right-skewed distributions and numerous outliers, the median bias was computed. The median bias (MB) is expressed as the difference between the median of the parameter estimates and the true parameter value:

$$MB = \text{Med}(\hat{\theta}) - \theta, \quad (10)$$

where $\hat{\theta}$ and θ denote the estimated and the true parameter value(s) of the average indirect effect, respectively. Furthermore, we investigated the median bias in the estimated standard errors by contrasting the median estimated standard error with the empirical standard error.

Fourth, the efficiency of the estimated average indirect effect was evaluated. Typically, the root mean squared error is used to examine the bias-variance trade-off. However, we utilized its robust counterpart the median absolute error (MAE):

⁷ As noted by an anonymous reviewer, McNeish and Bauer (2022) recently proposed using a factor analytic structure for the random effect covariance matrix to avoid non-positive definiteness. This approach was first suggested by Jennrich and Schluchter (1986) and was extended by Miyazaki and Frank (2006). For more information, we refer readers to McNeish and Bauer, (2022).

$$\text{MAE} = \text{Med}|\hat{\theta} - \theta|. \quad (11)$$

Fifth, the coverage probability (CP) of the 95% confidence intervals was inspected. The coverage probability is the proportion of 95% confidence intervals (CIs) containing the true parameter value:

$$\text{CP} = P(T_u \leq \theta \leq T_v), \quad (12)$$

with T_u and T_v the lower and upper bound of the 95% CI, respectively.

Finally, we examined the use of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for model comparison and selection (West et al., 2022). The AIC and BIC are derived from the log-likelihood (LL) statistic and can be computed as follows:

$$\text{AIC} = -2\log(L) + 2p \quad \text{and} \quad \text{BIC} = -2\log(L) + 2p\log(n), \quad (13)$$

where p is the number of estimated parameters and n is the total sample size. To evaluate the effectiveness of these information criteria for model selection, we calculated the model selection rate of the RS all approach over the RI approach. This was done by computing the proportion of instances in which the RS all approach yielded lower AIC or BIC compared to the RI approach within each condition.

Hypotheses

Based on the available multilevel literature, we anticipated several findings. First, we expected lower convergence rates for the RS approach compared to the RI approach, especially for a small number of clusters and/or when $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.01$ (i.e., when there were no random slopes in the data generating model). Second, we anticipated that the RI approach would show downward or upward biased estimates for the average indirect effect, depending on whether $\sigma_{a_j b_j}^2$ is negative or positive, and would underestimate the standard errors, except when $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.01$. Finally, we expected a poor coverage probability for the RI approach in most conditions, which may be improved when a robust estimator is used for the standard error (i.e., robust RI). For transparency and reproducibility, the code to perform the analysis and all the results are available at Bogaert et al. (2025).

Results

To compare the performance of the estimation models in terms of convergence rate, median bias (of both the indirect effect and the standard error), median absolute error, coverage probability, and the model selection rate, an ANOVA was conducted for each of those criteria. The independent variables included the employed estimation models

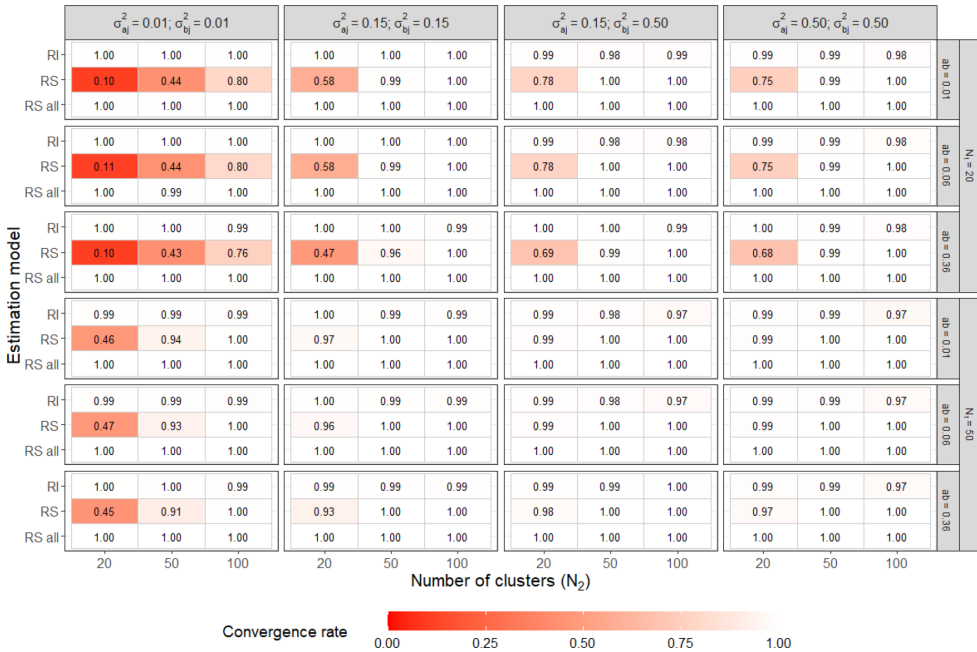
and the simulation conditions: cluster size (N_1), number of clusters (N_2), value of the indirect effect (ab), variance of the random slopes (σ_{rs}^2), and ICC. All two-way interactions between the independent variables were included in the analysis. The results, presented in [Table C1](#) in [Appendix C](#), indicated that cluster size, number of clusters, value of the indirect effect, random slope variance, and estimation models had a statistically significant effect on the performance criteria. In contrast, neither the ICC nor its interaction terms exhibited a significant effect, suggesting a limited impact on the outcomes under the tested conditions. Therefore, we focus on the conditions with balanced ICC values ($ICC_X = 0.2$, $ICC_M = 0.2$, $ICC_Y = 0.2$). Although various values for the random slope covariance were examined, we report only the results for the condition where $\sigma_{a_p, b_j} = 0.113$. The results showed mostly similar patterns across different values of the random slope covariance. However, when $\sigma_{a_p, b_j} = -0.113$, the bias in the average indirect effect was positive (instead of negative). When $\sigma_{a_p, b_j} = 0$, there was no bias in the average indirect effect, but the standard error was underestimated, which led to unacceptable low coverage probabilities. A full overview of the results can be found at [Bogaert et al. \(2025\)](#).

Convergence Rate

As expected, the RI approach outperformed the RS approach, with convergence rates consistently at or above 0.97. The RS approach exhibited poor performance in conditions with a small cluster size ($N_1 = 20$) and number of clusters ($N_2 = 20$), particularly when no random slopes were included in the data generating model (i.e., $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.01$), yielding convergence rates as low as 0.10. However, as both the cluster size and number of clusters increased, the RS approach achieved convergence rates as high as 1.00. The RS all approach demonstrated perfect convergence rates, suggesting the occurrence of numerous singular solutions in the case of a small sample size and/or when no random slopes were present in the data generating model. Overall, convergence rates remained largely similar across different values of the indirect effect. An overview of the convergence rates is provided in [Figure 3](#).

Figure 3

Convergence Rates for the Considered Estimation Models

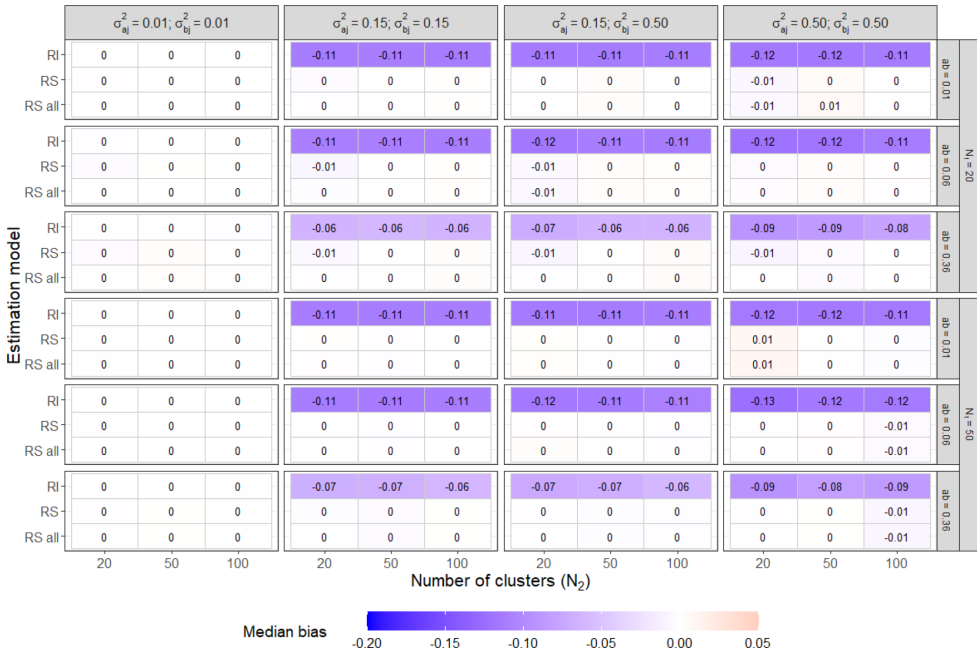


Median Bias of Average Indirect Effect Estimates

The RS approach performed well, even if the cluster size and number of clusters were small. Furthermore, only negligible differences were found between the RS and RS all approaches, indicating that the singular solutions yielded reasonable estimates for the average indirect effect. The RI approach demonstrated substantially biased estimates for the average indirect effect in all conditions, except when $\sigma_{a_i}^2 = \sigma_{b_j}^2 = 0.01$. In this case, there were no random slopes and therefore the covariance between the random slopes was zero (i.e., $\sigma_{a_i, b_j} = 0.01$). The RI approach showed the same absolute bias but smaller relative bias when the fixed indirect effect was larger. Finally, the cluster size, number of clusters, and ICC did not influence the overall performance regarding bias. An overview of the median bias can be found in [Figure 4](#).

Figure 4

Median Bias of the Average Indirect Effect Estimates for the Considered Estimation Models



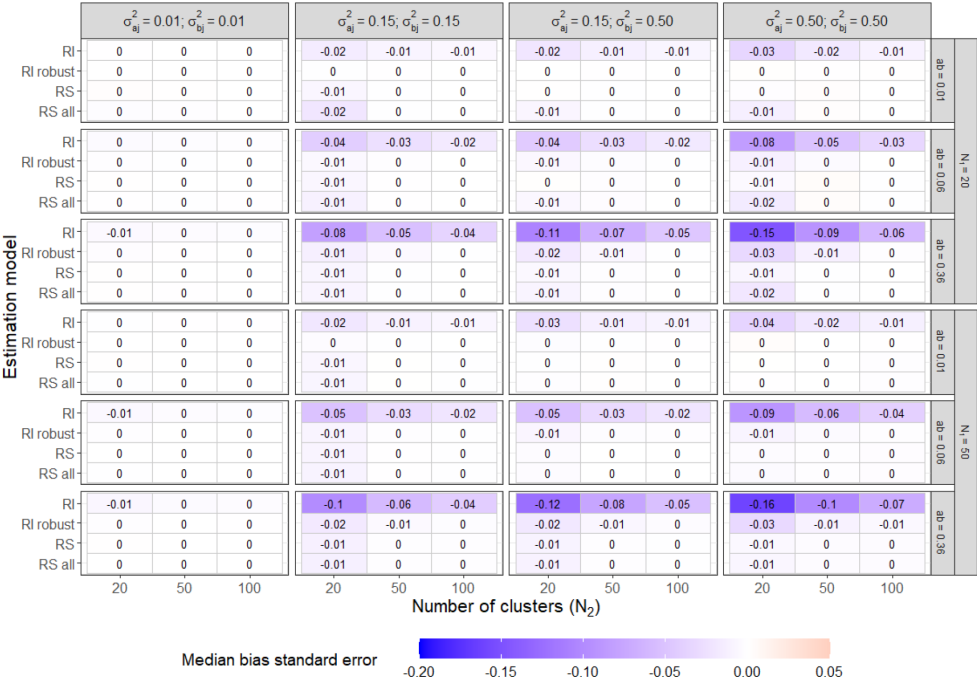
Median Bias of Standard Error

The RS approach was overall the best performing estimation model, showing only minor underestimation of standard errors if the cluster size and number of clusters is small. Furthermore, only negligible differences could be found between the RS and RS all approaches, except in the condition with the smallest cluster size ($N_1 = 20$) and number of clusters ($N_2 = 20$), in which case the singular solutions resulted in somewhat underestimated standard errors.

The RI approach showed considerable underestimated standard errors across all conditions, except when the data-generating model did not include random slopes. The underestimation increased substantially when the fixed indirect effect and variance of the random slopes became larger, as well as when the cluster size and number of clusters decreased. However, the underestimated standard error largely diminished when the RI approach is combined with a robust estimator, indicating that the RI robust approach performed comparable to RS in most conditions. Finally, the ICC did not influence the bias in the standard errors. An overview of the median bias can be found in Figure 5.

Figure 5

Median Bias of the Standard Error of the Average Indirect Effect Estimates for the Considered Estimation Models

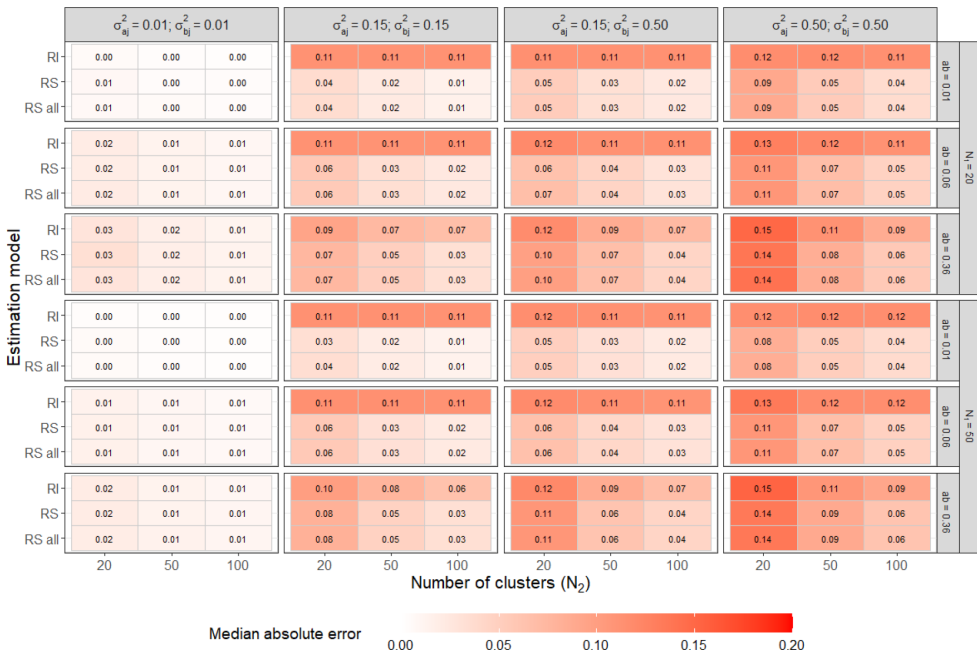


Bias-Variance Trade-Off: Median Absolute Error

The findings indicated that the RS (all) approach was the most efficient, as it exhibited the lowest median absolute error in all conditions, except when sigma_a_j^2 = sigma_b_j^2 = 0.01. No (considerable) differences in median absolute error values were observed between the RS and RS all approaches, even when the number of clusters and cluster size were small. The efficiency of all estimation models improved as the number of clusters increased and the random slope variance decreased. The RI approach also demonstrated a lower median absolute error when the fixed indirect effect was larger, in which case the differences between the estimation models became smaller. Additionally, the ICC and cluster size did not substantially influence efficiency. An overview of the median absolute error is provided in Figure 6.

Figure 6

Median Absolute Error of the Average Indirect Effect Estimates for the Considered Estimation Models



Coverage Probability

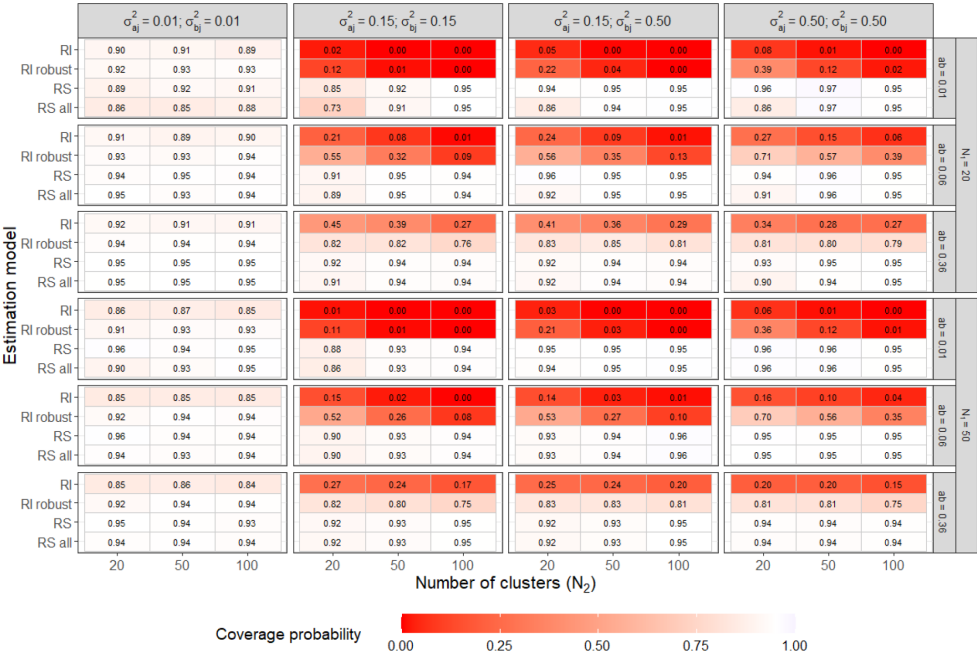
Overall, the RI approach demonstrated low coverage probabilities. It performed best when no random slopes were present in the data generating model, with coverage probabilities ranging from 0.84 and 0.92. The lowest coverage was observed with small cluster sizes. However, when random slopes were present, the RI approach exhibited coverage rates as low as 0.00 in multiple conditions. The RI robust approach showed better coverage than RI in most conditions but still performed poorly when random slopes were present in the data-generating model (coverage probabilities ranging from 0.00 to 0.85). The RI robust approach also performed best when $\sigma_{a_i}^2 = \sigma_{b_j}^2 = 0.01$, consistently achieving coverage probabilities near the nominal level of 0.95. Importantly, the performance of RI (robust) approach deteriorated as the cluster size and number of clusters increased. Although the RI robust approach exhibited improved coverage rates when the fixed indirect effect was large, its overall performance remained unsatisfactory.

The RS (all) approach demonstrated overall acceptable coverage probabilities. However, the coverage rate decreased when the cluster size and number of clusters were low, and when the fixed indirect effect was small. In these conditions, the RS all approach displayed somewhat lower coverage probabilities compared to RS, implying that in the

case of singular solutions lower coverage probability may also occur. Notably, no large differences in coverage probability were observed when $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.01$, where most singular solutions occurred. The ICC did not effect the coverage probability, as was the case for RI (robust) approach. An overview of the coverage probabilities is provided in Figure 7.

Figure 7

Coverage Probability of Considered Estimation Models for the 95% CI of the Average Indirect Effect

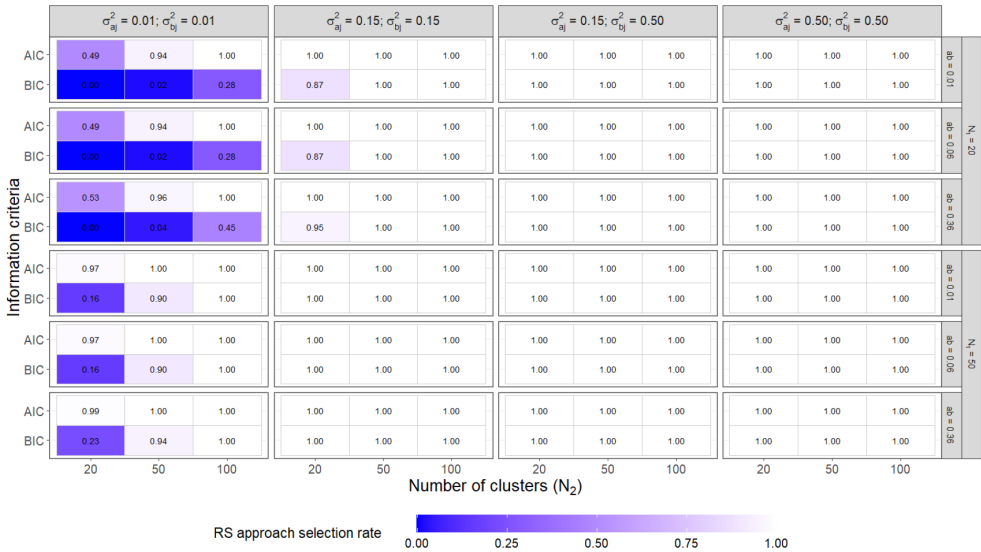


Model Selection Rate

When random slopes were included in the model, both the AIC and BIC favored the RS all approach over the RI approach under these conditions. When no random slopes were included, the RS all approach was increasingly favored over the RI approach as cluster size, number of clusters, and the magnitude of the indirect effect increased. Selection rates also varied depending on the information criterion used. Specifically, the BIC tended to favor the RI approach more frequently than the AIC, which only favored the RI approach in scenarios with small cluster sizes and a low number of clusters. Additionally, the ICC did not affect the model selection rate. An overview of the selection rates is provided in Figure 8.

Figure 8

Model Selection Rate of RS All Over RI Approach



Discussion

In this article, we investigated the impact of ignoring random slopes in a 1 → 1 → 1 mediation model. We examined the convergence rate, bias, efficiency, and coverage probability of the average indirect effect for an estimation model with and without random slopes. Various cluster sizes and number of clusters were used, along with different values for the intraclass correlation coefficient, the fixed effects and the variances of the random slopes.

The results revealed clear patterns that align with prior literature. First, we examined the convergence rate. As expected, the RI approach displayed very high convergence rates in all conditions, whereas the model including random slopes showed substantially lower convergence rates (i.e., many singular solutions) when the cluster size and number of clusters were small, particularly when no random slopes were included in the data generating model. These findings are consistent with those of Bauer et al. (2006), Bryan and Jenkins (2016), Park et al. (2020), and Schielzeth and Forstmeier (2009), who argued that the estimation of a multilevel model becomes more difficult as the number of random effects (and therefore the number of parameters to be estimated) increases. Importantly, when using the default optimizer of the lme4 package (instead of the bobyqa optimizer), the findings showed lower convergence rates for the RS approach.

We therefore recommend using the bobyqa optimizer in case of a $1 \rightarrow 1 \rightarrow 1$ mediation model.

The RS approach performed well in terms of bias of the estimated indirect effect, even when the data generating model contained no random slopes and/or in the smallest considered cluster size and number of clusters. In contrast, the RI approach displayed substantially biased indirect effect estimates in all conditions, except when there were no random slopes in the data generating model or the random slope covariance was zero (i.e., $\sigma_{a,b_j} = 0$). The bias decreased when the fixed indirect effect became larger, but still remained considerable. As discussed earlier, the RI approach did not provide unbiased estimates for the average indirect effect whenever $\sigma_{a,b_j} \neq 0$, or in other words, whenever correlated random slopes for paths a and b are present in the data generating model. This is a distinctive finding for the $1 \rightarrow 1 \rightarrow 1$ mediation model compared to a standalone multilevel model, where omitting random slopes usually does not result in biased estimators for the fixed effects (LaHuis et al., 2020; Schielzeth & Forstmeier, 2009).

A similar trend was observed for the estimated standard errors. The RS approach yielded appropriate standard errors, whereas the RI approach systematically underestimated the standard errors. This result coincides with findings of previous work Barr et al. (2013), Bell et al. (2019), Heisig et al. (2017), and LaHuis et al. (2020), who investigated the impact of omitting random slopes in the standalone multilevel model and reported underestimated standard errors for the fixed effects too. A better performance was observed for the RI approach when a robust estimator for the standard error was used, which is in line with previous research findings (Bell et al., 2019; King & Roberts, 2015).

The RS approach resulted in acceptable coverage probabilities for the average indirect effect in most conditions. In contrast, the RI approach performed extremely poorly. Its performance was best when no random slopes were included in the data generating model, yet it still exhibited too low coverage probabilities under these conditions. These findings align with the results of Barr et al. (2013), Bell et al. (2019), LaHuis et al. (2020), and Schielzeth and Forstmeier (2009), who concluded that falsely assuming fixed slopes can be detrimental for statistical inference (i.e., too narrow confidence intervals). Here, however, the low coverage probabilities result from both the bias in the estimated indirect effect and the underestimated standard error, which respectively shifts the confidence intervals away from the true average indirect effect and narrows the 95 % CIs.

Finally, the RI and RS all approaches were compared in terms of their performance regarding the bias-variance trade-off and model selection based on information criteria. The RS all approach exhibited smaller median absolute error values compared to the RI approach, except when $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.01$, and is therefore considered being more efficient. Both information criteria showed high model selection rates, indicating that the RS all approach was generally favored over the RI approach, particularly when random slopes were included in the model. In the absence of random slopes, the RS all approach was increasingly preferred as cluster size, number of clusters, and the magnitude of the

indirect effect increased. The BIC favored the RI approach more often than the AIC, but only in scenarios with small cluster sizes and a low number of clusters.

Overall, we make the following recommendations. When convergence or estimation issues arise and/or insufficient resources are available, a first recommendation is to try a different optimizer and/or increasing the number of iterations. A second recommendation is to fit the RI approach with robust standard errors. The RI approach with robust standard errors showed a better but still unsatisfactory performance (due to biased indirect effect estimates and low coverage probability). However, it can be used as a diagnostic procedure, as suggested by [King and Roberts \(2015\)](#). If there is a disparity between model-based and robust standard errors, it implies the presence of model misspecification. This misspecification could potentially involve the omission of random slopes in the model [Bell et al. \(2019\)](#). A third recommendation is to use a piecemeal approach to potentially leave out random slopes (e.g., for the direct effect) and/or correlation between the random effects. This may involve evaluating the magnitude of the estimated variance components and comparing models using likelihood ratio tests with the `varTestnlme` package in R for example or information criteria ([Bates, Kliegl et al., 2015](#); [Bell et al., 2019](#); [LaHuis et al., 2020](#); [Matuschek et al., 2017](#); [Park et al., 2020](#)). Finally, it is also possible to test the $\sigma_{a,b}$ parameter using a Wald test.

Like all simulation studies, this study was not without limitations. A first limitation is that we restricted upper-level confounding to the mediator-outcome relationship. Under the considered data generating mechanism in our simulation study the within- and between-effect of the exposure are equal. Though this assumption is only likely to hold in settings such as alternating treatment designs where the exposure is randomized, it is also unlikely to matter for the main conclusions in this paper. A second limitation is that the simulation study included only a limited number of conditions. Although the conditions that we considered here may be commonly found in the literature, different settings may lead to different results. A third limitation is that we have not considered the bootstrap or Monte Carlo approach for estimating the precision of the indirect effect since it is computationally very intensive. A fourth limitation is that we have not considered cross-level interactions (i.e., a Level 2 that moderates the relationship between two Level 1 variables). Previous research findings indicate that the negative consequences of falsely omitting random slopes may aggravate in this condition ([Heisig et al., 2017](#)). A fifth limitation in the presentation of the simulation results is that we only compared the performance of the more simple RI model with the performance of the more complex RS model. In practice, a researcher may use model selection based on the AIC. Once a model is selected by AIC, it could be interesting to look whether it actually gives good estimates of the parameters of interest.

In this work, we have employed the joint modeling approach by [Bauer et al. \(2006\)](#), as it was the most convenient approach to examine the consequences of ignoring random slopes in a $1 \rightarrow 1 \rightarrow 1$ mediation model in the traditional multilevel framework. However,

there are other methods for multilevel mediation analysis, such as MSEM (Preacher, 2011). The MSEM approach to multilevel mediation modeling has the advantage that it can easily handle multiple multilevel models simultaneously, but it often requires a larger sample size compared to the traditional multilevel approaches (McNeish, 2017; Preacher et al., 2011; Zigler & Ye, 2019). Indeed the latent decomposition in MSEM can be quite demanding. Because MSEM is more complex (modeling latent components, more parameters) and has greater sample-size and convergence demands, we stuck to the joint multilevel approach. The latter is easier to implement in standard mixed-model software and can perform adequately when focus is on within effects. Future research could directly compare both approaches.

As explained we opted not to separate within and between effects using a manifest decomposition (Zhang et al., 2009). The decomposition is not needed to eliminate upper level confounding when allowing correlated random slopes. The downside of the approach presented in this paper is that it only allows for unmeasured upper level confounding of the mediator-outcome relationship. In non-randomized studies there may also be exposure-mediator or exposure-outcome confounding. In principle one could add a third model to the stacked approach of Bauer et al. (2006), and model the exposure. By allowing the random intercepts of all three models to be correlated, one allows for all three types of upper level confounding. Obviously, such approach would further increase the dimensionality of the covariance matrix of the random effects. Alternatively, one may follow the suggestion of Bauer (2015), and separate the within-effect and between-effect of the exposure to avoid conflated effects. The advantage of the latter approach is that it does not require an additional model for the exposure.

In sum, we have shown that omitting random slopes in a $1 \rightarrow 1 \rightarrow 1$ mediation model results in biased average indirect effects, as well as underestimated standard errors, leading to invalid inference. Given the limited negative impact for unnecessarily allowing random slopes, we encourage (when in doubt) to include random slopes in a $1 \rightarrow 1 \rightarrow 1$ mediation model. This advice is in line with conclusions from other scholars (Barr et al., 2013; Bell et al., 2019; LaHuis et al., 2020; Schielzeth & Forstmeier, 2009; Schmidt-Catran & Fairbrother, 2016). At the same time it is important to acknowledge potential convergence issues for the RS approach due to the increased complexity of the model when all random slopes are included (Heisig et al., 2017; LaHuis et al., 2020; Matuschek et al., 2017).

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

Data Availability: Simulation study data is available at Bogaert et al. (2025)

Supplementary Materials

Type of supplementary materials	Availability/Access
Data	
Stimulation study data.	Bogaert et al. (2025)
Code	
R Code.	Bogaert et al. (2025)
Material	
a. Study results - pngs.	Bogaert et al. (2025)
b. ReadMe file - instructions.	Bogaert et al. (2025)
Study/Analysis preregistration	
The study was not preregistered.	—
Other	
No other material to report.	—

References

- Aroian, L. A., Taneja, V. S., & Cornwell, L. W. (1978). Mathematical forms of the distribution of the product of two normal variables: Mathematical forms of the distribution. *Communications in Statistics – Theory and Methods*, 7(2), 165–172. <https://doi.org/10.1080/03610927808827610>
- Baird, R., & Maxwell, S. E. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods*, 21(2), 175–188. <https://doi.org/10.1037/met0000070>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. <https://doi.org/10.48550/arXiv.1506.04967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauer, D. J. (2015). *I want to disaggregate the within- and between-effects of the predictors in my model. Should I use group-mean centering on X and/or M to accomplish this?* Daniel J. Bauer, Ph.D. — University of North Carolina at Chapel Hill. <https://dbauer.web.unc.edu/wp-content/uploads/sites/7494/2015/08/Centering-in-111-Mediation.pdf>
- Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11(2), 142–163. <https://doi.org/10.1037/1082-989X.11.2.142>

- Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity*, 53(2), 1051–1074. <https://doi.org/10.1007/s11135-018-0802-x>
- Berkhof, J., & Kampen, J. K. (2004). Asymptotic effect of misspecification in the random part of the multilevel model. *Journal of Educational and Behavioral Statistics*, 29(2), 201–218. <https://doi.org/10.3102/107699860290022>
- Bogaert, J., Loh, W. W., Loeyts, T., Moerkerke, B., & Rosseel, Y. (2025). *The impact of omitting random slopes in a 1-1-1 mediation model* [OSF project page containing simulation study data, results, illustrations, and code]. Open Science Framework. <https://osf.io/vu5q2/overview>
- Bryan, M. L., & Jenkins, S. P. (2016). Multilevel modelling of country effects: A cautionary tale. *European Sociological Review*, 32(1), 3–22. <https://doi.org/10.1093/esr/jcv059>
- Fitzsimmons-Craft, E. E., Bardone-Cone, A. M., Crosby, R. D., Engel, S. G., Wonderlich, S. A., & Bulik, C. M. (2016). Mediators of the relationship between thin-ideal internalization and body dissatisfaction in the natural environment. *Body Image*, 18, 113–122. <https://doi.org/10.1016/j.bodyim.2016.06.006>
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55(292), 708–713. <https://doi.org/10.1080/01621459.1960.10483369>
- Heisig, J. P., & Schaeffer, M. (2018). Why you should always include a random slope for the lower-level variable of a cross-level interaction. *European Sociological Review*, 35(2), 258–279. <https://doi.org/10.1093/esr/jcy053>
- Heisig, J. P., Schaeffer, M., & Giesecke, J. (2017). The costs of simplicity: why multilevel models may benefit from accounting for cross-cluster differences in the effects of controls. *American Sociological Review*, 82(4), 796–827. <https://doi.org/10.1177/0003122417717901>
- Hill, N. L., Bhargava, S., Bratlee-Whitaker, E., Turner, J. R., & Brown, M. J., & Mogle, J. (2021). Longitudinal relationships between subjective cognitive decline and objective memory: Depressive symptoms mediate between-person associations. *Journal of Alzheimer's Disease*, 83(4), 1623–1636. <https://doi.org/10.3233/JAD-210230>
- Hoffman, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology*, 73, 659–89.
- Hox, J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, 8(2), 157–174. https://doi.org/10.1207/S15328007SEM0802_1
- Hox, J., & McNeish, D. (2020). Small samples in multilevel modeling. In R. van de Schoot & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 215–225). Routledge.
- Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42(4), 805–820.
- Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8(2), 115–128. <https://doi.org/10.1037/1082-989X.8.2.115>

- Kim, S., & Hong, S. (2020). Comparing methods for multilevel moderated mediation: a decomposed-first strategy. *Structural Equation Modeling*, 27(5), 661–677.
<https://doi.org/10.1080/10705511.2019.1683015>
- King, G., & Roberts, M. E. (2015). How robust standard errors expose methodological problems they do not fix, and what to do about it. *Political Analysis*, 23(2), 159–179.
<https://doi.org/10.1093/pan/mpu015>
- LaHuis, D. M., Jenkins, D. R., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2020). The effects of misspecifying the random part of multilevel models. *Methodology*, 16(3), 224–40.
<https://doi.org/10.5964/meth.2799>
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86–92. <https://doi.org/10.1027/1614-1881.1.3.86>
- MacKinnon, D. P. (2012). *Introduction to statistical mediation analysis*. Routledge.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4
- Manolov, R., Tanious, R., & Onghena, P. (2022). Quantitative techniques and graphical representations for interpreting results from alternating treatment design. *Perspectives on Behavior Science*, 45(1), 259–294. <https://doi.org/10.1007/s40614-021-00289-9>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–15.
<https://doi.org/10.1016/j.jml.2017.01.001>
- McNeish, D. (2014). Modeling sparsely clustered data: design-based, model-based, and single-level methods. *Psychological Methods*, 19(4), 552–563. <https://doi.org/10.1037/met0000024>
- McNeish, D. (2017). Multilevel mediation with small samples: a cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 609–625. <https://doi.org/10.1080/10705511.2017.1280797>
- McNeish, D. (2023). A practical guide to selecting and blending approaches for clustered data: Clustered errors, multilevel models, and fixed-effect models. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000620>
- McNeish, D., & Bauer, D. J. (2022). Reducing incidence of nonpositive definite covariance matrices in mixed effect models. *Multivariate Behavioral Research*, 57(2–3), 318–340.
<https://doi.org/10.1080/00273171.2020.1830019>
- Metcalfe, J., & Greene, M. J. (2007). Metacognition of agency. *Journal of Experimental Psychology-General*, 136(2), 184–199. <https://doi.org/10.1037/0096-3445.136.2.184>
- Miyazaki, Y., & Frank, K. A. (2006). A hierarchical linear model with factor analysis structure at Level 2. *Journal of Educational and Behavioral Statistics*, 31(2), 125–156.
<https://doi.org/10.3102/10769986031002125>
- Park, J., Cardwell, R., & Yu, H-T. (2020). Specifying the random effect structure in linear mixed effect models for analyzing psycholinguistic data. *Methodology*, 16(2), 92–111.
<https://doi.org/10.5964/meth.2809>

- Preacher, K. J. (2011). Multilevel SEM strategies for evaluating mediation in three-level data. *Multivariate Behavioral Research*, 46(4), 691–731. <https://doi.org/10.1080/00273171.2011.589280>
- Preacher, K. J., & Selig, J. P. (2010). *Monte Carlo method for assessing multilevel mediation: An interactive tool for creating confidence intervals for indirect effects in 1-1-1 multilevel models* [Computer software]. Kristopher Preacher. <http://quantpsy.org/>
- Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6(2), 77–98. <https://doi.org/10.1080/19312458.2012.679848>
- Preacher, K. J., Zhang, Z., & Zychur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: the advantages of multilevel SEM. *Structural Equation Modeling*, 18(2), 161–182. <https://doi.org/10.1080/10705511.2011.557329>
- Preacher, K. J., Zychur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209–233. <https://doi.org/10.1037/a0020141>
- Pustejovsky, J. (2023). *clubSandwich: cluster-robust (Sandwich) variance estimators with small-sample corrections* [Computer software manual]. clubSandwich <http://jepusto.github.io/clubSandwich/>.
- R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software manual]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (Vol. 1). SAGE.
- Schielezeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, 20(2), 416–420. <https://doi.org/10.1093/beheco/arn145>
- Schmidt-Catran, A. W., & Fairbrother, M. (2016). The random effects in multilevel models: Getting them wrong and getting them right. *European Sociological Review*, 32(1), 23–38. <https://doi.org/10.1093/esr/jcv090>
- Shadish, W. R., Kyse, E. N., & Rindskopf, D. M. (2013). Analyzing data from single-case designs using multilevel models: New applications and some agenda items for future research. *Psychological Methods*, 18(3), 385–405. <https://doi.org/10.1037/a0032964>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422–445. <https://doi.org/10.1037/1082-989X.7.4.422>
- Tan, Y., Fan, Z., Wei, X., & Yang, T. (2022). School belonging and reading literacy: A multilevel moderated mediation model. *Frontiers in Psychology*, 13, Article 816128. <https://doi.org/10.3389/fpsyg.2022.816128>
- Tofghi, D., West, S. G., & MacKinnon, D. P. (2013). Multilevel mediation analysis: The effects of omitted variables in the 1–1–1 model. *British Journal of Mathematical and Statistical Psychology*, 66(2), 290–307. <https://doi.org/10.1111/j.2044-8317.2012.02051.x>
- Vuurro, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, 50(5), 2125–2143. <https://doi.org/10.3758/s13428-017-0980-9>

- West, B. T., Welch, K. B., & Galecki, A. T. (2022). *Linear mixed models: A practical guide using statistical software*. CRC Press.
- Yu, H., & Hu, J. (2022). ICT self-efficacy and ICT interest mediate the gender differences in digital reading: A multilevel serial mediation analysis. *International Journal of Emerging Technologies in Learning*, 17(5), 211–225. <https://doi.org/10.3991/ijet.v17i05.25691>
- Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, 12(4), 695–719. <https://doi.org/10.1177/1094428108327450>
- Zigler, C. K., & Ye, F. (2019). A comparison of multilevel mediation modeling methods: recommendations for applied researchers. *Multivariate Behavioral Research*, 54(3), 338–359. <https://doi.org/10.1080/00273171.2018.1527676>

Appendices

Appendix A: Illustration of the Considered Estimation Approaches

Part 1

Illustration of the Considered Estimation Approaches

```
#####
###                                     ###
###      Illustration approaches      ###
###                                     ###
#####

### Load packages
library(MASS)
library(reshape2)
library(lme4)
library(lmerTest)
library(clubSandwich)

#### Read in example data set
dat <- read.csv("dat_illustration.csv")

### Reorganize data as in Bauer et al. (2006)
dat$id <- 1:nrow(dat)
dat_stacked <- melt(dat, id.vars = c("id", "cluster", "X", "M"),
  measure.vars = c("Y", "M"), value.name = "Z")
dat_stacked <- within(dat_stacked, {
  sY <- as.integer(variable == "Y")
  sM <- as.integer(variable == "M")
})

### 1. fit model with random slopes
mm_RS <- lmer(Z ~ 0 + # no intercept
  sM + sM:X + # model for the mediator
  sY + sY:M + sY:X + # model for outcome
  (0 + sM + sM:X + sY + sY:M + sY:X | cluster) + # random effects
  (0 + sM | id), data = dat_stacked, # Var(e_Zij) depending on sM
  control=lmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5))

# check and save summary output
(smm_RS <- summary(mm_RS))

# information criteria and LRT test statistic
AIC_RS <- AIC(mm_RS)
BIC_RS <- BIC(mm_RS)
LL_RS <- as.numeric(logLik(mm_RS))

# fixed effects
FEmatrix_RS <- coef(smm_RS)
a_RS <- FEmatrix_RS['sM:X', 'Estimate']
b_RS <- FEmatrix_RS['sY:M', 'Estimate']
cprime_RS <- FEmatrix_RS['X:sY', 'Estimate']

# random effects
REmatrix_RS <- as.data.frame(VarCorr(mm_RS))
cov_aj.bj_RS <- REMatrix_RS[which(REMatrix_RS$var1 == 'sM:X' &
```

Part 2

Illustration of the Considered Estimation Approaches

```

REmatrix_RS$var2 == 'sY:M'), 'vcov']

# asymptotic covariance matrix for the fixed effects
AVCOV_FE_RS <- vcov(mm_RS)
names_vcov_F <- c("sM", "sY", "sM:X", "sY:M", "X:sY")
rownames(AVCOV_FE_RS) <- colnames(AVCOV_FE_RS) <- names_vcov_F
var_a_RS <- AVCOV_FE_RS["sM:X", "sM:X"]
var_b_RS <- AVCOV_FE_RS["sY:M", "sY:M"]
var_c_RS <- AVCOV_FE_RS["X:sY", "X:sY"]
cov_a_b_RS <- AVCOV_FE_RS["sM:X", "sY:M"]
cov_a_c_RS <- AVCOV_FE_RS["sM:X", "X:sY"]
cov_b_c_RS <- AVCOV_FE_RS["sY:M", "X:sY"]

# asymptotic covariances for the random effects
# (Using the vcov_vc() in the bootmlm library)
AVCOV_RE_RS <- try(vcov_vc(mm_RS, sd_cor = FALSE, print_names = TRUE))
if(!inherits(AVCOV_RE_RS, "try-error")){
  var_cov_aj.bj_RS <- AVCOV_RE_RS["cov_sY:M.sM:X|cluster", "cov_sY:M.sM:X|cluster"]
} else{
  var_cov_aj.bj_RS <- NA
}

# indirect effect
IE_RS <- a_RS*b_RS + cov_aj.bj_RS
var_IE_RS <- (b_RS^2)*var_a_RS + (a_RS^2)*var_b_RS + var_a_RS*var_b_RS +
  2*a_RS*b_RS*cov_a_b_RS + cov_a_b_RS^2 + var_cov_aj.bj_RS
CI_LB_IE_RS <- IE_RS - 1.96*sqrt(var_IE_RS)
CI_UB_IE_RS <- IE_RS + 1.96*sqrt(var_IE_RS)

### 2. fit model without random slopes and regular SE
# fit model without random slopes and regular SE
mm_noRS <- lmer(Z ~ 0 + # no intercept
  sM + sM:X + # model for the mediator
  sY + sY:M + sY:X + # model for outcome
  (0 + sM + sY | cluster) + # random effects (no random slopes)
  (0 + sM | id), data = dat_stacked) # Var(e_Zij) depending on SM

# check and save summary output
(smm_noRS <- summary(mm_noRS))

# information criteria and LRT test statistic
AIC_noRS <- AIC(mm_noRS)
BIC_noRS <- BIC(mm_noRS)
LL_noRS <- as.numeric(logLik(mm_noRS))

# fixed effects
FEmatrix_noRS <- coef(smm_noRS)
a_noRS <- FEmatrix_noRS['sM:X', 'Estimate']
b_noRS <- FEmatrix_noRS['sY:M', 'Estimate']
cprime_noRS <- FEmatrix_noRS['X:sY', 'Estimate']

# asymptotic covariance matrix for the fixed effects

```

Part 3

Illustration of the Considered Estimation Approaches

```

AVCOV_FE_noRS <- vcov(mm_noRS)
names_vcov_F <- c("sM", "sY", "sM:X", "sY:M", "X:sY")
rownames(AVCOV_FE_noRS) <- colnames(AVCOV_FE_noRS) <- names_vcov_F
var_a_noRS <- AVCOV_FE_noRS["sM:X", "sM:X"]
var_b_noRS <- AVCOV_FE_noRS["sY:M", "sY:M"]
var_c_noRS <- AVCOV_FE_noRS["X:sY", "X:sY"]
cov_a_b_noRS <- AVCOV_FE_noRS["sM:X", "sY:M"]
cov_a_c_noRS <- AVCOV_FE_noRS["sM:X", "X:sY"]
cov_b_c_noRS <- AVCOV_FE_noRS["sY:M", "X:sY"]

# indirect effect
IE_noRS <- a_noRS*b_noRS
var_IE_noRS <- (b_noRS^2)*var_a_noRS + (a_noRS^2)*var_b_noRS +
  var_a_noRS*var_b_noRS + 2*a_noRS*b_noRS*cov_a_b_noRS + cov_a_b_noRS^2
CI_LB_IE_noRS <- IE_noRS - 1.96*sqrt(var_IE_noRS)
CI_UB_IE_noRS <- IE_noRS + 1.96*sqrt(var_IE_noRS)

### 3. fit model without random slopes and robust SE

# check and save summary output
mm_noRS_RB <- mm_noRS
(smm_noRS_RB <- summary(mm_noRS_RB))

# information criteria and LRT test statistic
AIC_noRS_RB <- AIC(mm_noRS_RB)
BIC_noRS_RB <- BIC(mm_noRS_RB)
LL_noRS_RB <- as.numeric(logLik(mm_noRS_RB))

# fixed effects
FEmatrix_noRS_RB <- coef(smm_noRS_RB)
a_noRS_RB <- FEmatrix_noRS_RB['sM:X', 'Estimate']
b_noRS_RB <- FEmatrix_noRS_RB['sY:M', 'Estimate']
cprime_noRS_RB <- FEmatrix_noRS_RB['X:sY', 'Estimate']

# asymptotic covariance matrix for the fixed effects
AVCOV_FE_noRS_RB <- vcovCR(obj = smm_noRS_RB, type = "CRO")
rownames(AVCOV_FE_noRS_RB) <- colnames(AVCOV_FE_noRS_RB) <- names_vcov_F
var_a_noRS_RB <- AVCOV_FE_noRS_RB["sM:X", "sM:X"]
var_b_noRS_RB <- AVCOV_FE_noRS_RB["sY:M", "sY:M"]
var_c_noRS_RB <- AVCOV_FE_noRS_RB["X:sY", "X:sY"]
cov_a_b_noRS_RB <- AVCOV_FE_noRS_RB["sM:X", "sY:M"]
cov_a_c_noRS_RB <- AVCOV_FE_noRS_RB["sM:X", "X:sY"]
cov_b_c_noRS_RB <- AVCOV_FE_noRS_RB["sY:M", "X:sY"]

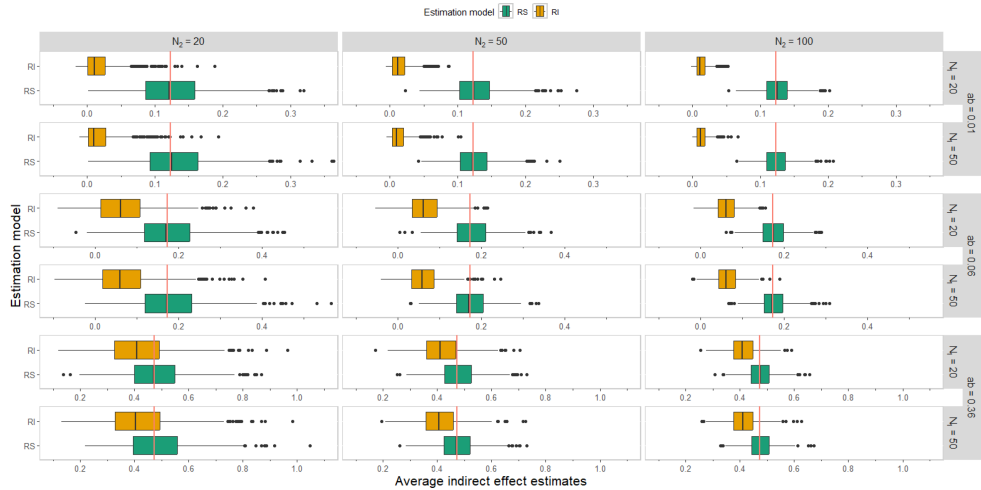
# indirect effect
IE_noRS_RB <- a_noRS_RB*b_noRS_RB
var_IE_noRS_RB <- (b_noRS_RB^2)*var_a_noRS_RB + (a_noRS_RB^2)*var_b_noRS_RB +
  var_a_noRS_RB*var_b_noRS_RB + 2*a_noRS_RB*b_noRS_RB*cov_a_b_noRS_RB +
  cov_a_b_noRS_RB^2
CI_LB_IE_noRS_RB <- IE_noRS_RB - 1.96*sqrt(var_IE_noRS_RB)
CI_UB_IE_noRS_RB <- IE_noRS_RB + 1.96*sqrt(var_IE_noRS_RB)

```

Appendix B: Figure Containing the Boxplots of the Average Indirect Effect Estimates

Figure B1

Boxplots of the Average Indirect Effect Estimates



Note. These boxplots of the average indirect effect estimates are for when $\sigma_{a_j}^2 = \sigma_{b_j}^2 = 0.15$, balanced ICC values ($ICC_X = 0.2, ICC_M = 0.2, ICC_Y = 0.2$), and where $\sigma_{a_j b_j} = 0.113$. The true indirect effect is represented by a red vertical line.

Appendix C: ANOVA Results

Table C1

Table Containing the Results of the ANOVAs

Effect	Df	CR	MB _{RE}	MB _{SE}	MAE	CP	MSR
N_1	1	399.84***	0.00	4.17*	5.00*	23.34>***	131.40***
N_2	2	446.87***	2.74	289.73***	303.50***	1.61	116.89***
ab	2	2.17	131.31***	424.78***	107.27***	46.91***	0.31
σ_{rs}^2	3	395.66***	52.55***	428.64***	973.03***	609.90***	153.07***
ICC	2	2.32	0.01	0.50	0.00	0.14	0.37
Model	2	1161.59***	420.76***	2335.75***	255.42***	7057.10***	–
$N_1: N_2$	2	95.95***	0.69	0.91	0.39	0.14	98.09***
$N_1: ab$	2	0.84	0.01	4.90**	0.11	8.48***	0.13
$N_1: \sigma_{rs}^2$	3	65.37***	0.42	0.10	0.52	0.76	131.35***
$N_1: ICC$	2	0.61	0.01	0.75	0.01	0.05	0.34
$N_1: Model$	2	434.15***	0.06	23.92***	0.23	58.17***	–

Effect	Df	CR	MB _{in}	MB _{se}	MAE	CP	MSR
N_2 : ab	4	1.24	1.23	19.42***	16.73***	3.14*	0.19
N_2 : σ_{rs}^2	6	66.12***	0.72	41.23***	28.17***	0.61	116.84***
N_2 : ICC	4	2.08	0.00	0.06	0.05	0.05	0.14
N_2 : Model	4	478.86***	0.10	120.02***	17.38***	17.33***	–
ab : σ_{rs}^2	6	0.07	14.66***	63.14***	5.60***	2.49*	0.31
ab : ICC	4	0.00	0.03	0.01	0.01	0.02	0.00
ab : Model	4	3.23*	123.30***	371.07***	25.40***	31.41***	–
σ_{rs}^2 : ICC	6	0.09	0.01	0.49	0.03	0.03	0.37
σ_{rs}^2 : Model	6	414.75***	44.82***	335.27***	47.20***	634.89***	–
ICC: Model	4	2.10	0.01	0.23	0.00	0.04	–
Residuals	1872						

Note. Df: degrees of freedom, CR: convergence rate, MB_{in}: median bias of indirect effect, MB_{se}: median bias of standard error, MAE: median absolute error, CP: coverage probability, MSR: model selection rate for RS approach based on AIC.

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.



Methodology (METH) is the official journal of the European Association of Methodology (EAM).



PsychOpen GOLD is a publishing service provided by the Leibniz Institute for Psychology (ZPID), Germany.