Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

More Precise Estimation of Lower-Level Interaction Effects in Multilevel Models

Tom Loeys, Haeike Josephy & Marieke Dewitte

MULTIVARIATE BEHAVIORAL RESEARCH
2018, VOL. 0, NO. 0, 1–13
https://doi.org/10.1080/00273171.2018.1444975

Routledge
Taylor & Francis Group

Check for updates

More Precise Estimation of Lower-Level Interaction Effects in Multilevel Models

Tom Loeys[a], Haeike Josephy[a], and Marieke Dewitte[b]

[a]Department of Data Analysis, Ghent University, Gent, Belgium; [b]Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

ABSTRACT

In hierarchical data, the effect of a lower-level predictor on a lower-level outcome may often be confounded by an (un)measured upper-level factor. When such confounding is left unaddressed, the effect of the lower-level predictor is estimated with bias. Separating this effect into a within- and between-component removes such bias in a linear random intercept model under a specific set of assumptions for the confounder. When the effect of the lower-level predictor is additionally moderated by another lower-level predictor, an interaction between both lower-level predictors is included into the model. To address unmeasured upper-level confounding, this interaction term ought to be decomposed into a within- and between-component as well. This can be achieved by first multiplying both predictors and centering that product term next, or vice versa. We show that while both approaches, on average, yield the same estimates of the interaction effect in linear models, the former decomposition is much more precise and robust against misspecification of the effects of cross-level and upper-level terms, compared to the latter.

KEYWORDS

Centering; Confounding; Interactions; Multilevel models

Introduction

When measures are collected repeatedly over time in individuals (e.g., in daily diary studies), such data can yield much more information compared to a cross-sectional sample. For example, when studying the relationship between intimacy and positive relationship feelings in a daily diary study, a between-person effect can be disentangled from a within-person effect (Curran & Bauer, 2011; Wang & Maxwell, 2015). In this example, the between-person effect reflects the extent to which individuals with higher intimacy differ in their positive relational feelings from individuals with a lower intimacy. The within-person effect on the other hand reflects the extent to which an individual exhibits higher (or lower) positive relational feelings when (s)he had more (or less) intimacy on a particular day, as compared to other days.

During the last two decades, the behavioral science literature has increasingly focused on separating within- from between-effects in multilevel models (Curran & Bauer, 2011; Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Kreft, de Leeuw, & Aiken, 1995; Raudenbush & Bryk, 2002). Two important issues can be highlighted when disaggregating those effects within longitudinal data: centering and detrending (Curran & Bauer, 2011). The former refers to subtracting a constant from every value of a variable, while the latter refers to removing the time trend from time series. The centering issue is relevant for disaggregation, even when neither the predictor nor the outcome exhibits any trend over time, whereas the detrending issue is only relevant when at least one of those variables exhibits some trend over time (Wang & Maxwell, 2015). In this article, we assume no time effects on either the predictor or the outcome and consequently limit our focus to the centering issue.

The multilevel literature typically considers two levels: the lower-level or level 1 (e.g., the daily measurements within the individual), and the upper-level or level 2 (e.g., the individuals in a diary study). Within such two-level data structures, three types of centering can be distinguished: no centering (i.e., the raw scores are used), grand-mean centering (i.e., subtraction of the overall average across individuals and time points) and cluster-mean centering (i.e., subtraction of a person-specific mean, averaged across time points within the individual). There is a general consensus that cluster-mean centering (also referred to as "CWC," centering within clusters) is deemed most appropriate when lower-level predictors are of primary substantive interest (Enders & Tofighi, 2007). More specifically, CWC may solve potential confounding issues in estimating the effect of a predictor on an outcome. A detailed explanation on why is discussed in the next section. When unmeasured upper-level common

CONTACT Tom Loeys tom.loeys@ugent.be Department of Data Analysis, Ghent University, Henri Dunantlaan 1, 9000 Gent, Belgium.
TL would like to thank the Flemish Research Council (FWO) for financial support (Grant G019317N).

causes of the predictor–outcome relationship are present, we refer to such causes as unmeasured upper-level confounders. In the econometrics literature, this type of unmeasured confounding at the subject- or cluster-level is referred to as upper-level endogeneity (Wooldridge, 2010). Here, we argue that confounding at the upper-level is very common in many contexts. In our illustration, for example, it is not unlikely that the daily measurements of intimacy and positive relational feelings are both affected by unmeasured stable (personality) traits of the individual. We will show that under a specific set of assumptions for the unmeasured upper-level confounder, CWC allows unbiased estimation of the within-subject effect of a lower-level predictor on an outcome.

Unfortunately, discussions on the role of centering are mostly limited to the assessment of main effects in multilevel models (MLM) and ignore the centering of interactions. An issue of particular importance entails the centering of interactions in the $1 \times (1 \rightarrow 1)$ design, where the first "1" corresponds to the level at which the moderator is measured, the second "1" represents the level of the predictor, and the last "1" defines the level of the outcome (Preacher, Zhang, & Zyphur, 2016; Ryu, 2015). We will refer to such interactions as "lower-level interactions." When cluster-mean centering such interactions, the question arises whether the predictor and moderator should be centered first and multiplied next (hereafter labeled as "C1P2," center-first and product-second), or whether it should be the other way around (labeled hereafter as "P1C2"). In contrast to cluster-mean centering an interaction between an upper- and a lower-level variable, or between two upper-level variables, C1P2 and P1C2 produce different predictors when cluster-mean centering a lower-level interaction. Some scholars favored P1C2 (Josephy, Vansteelandt, Vanderhasselt, & Loeys, 2015), while others advised against it and promoted C1P2 instead (Preacher et al., 2016). In this article, we investigate how these two approaches deal with unmeasured upper-level confounding and whether they can unbiasedly estimate the moderated within-subject effect.

While Josephy et al. (2015) considered the traditional multilevel modeling (MLM) framework (also referred to as mixed modeling), Preacher et al. (2016) relied on Structural Equation Modeling (SEM). In contrast to the traditional MLM-framework, in which the within- and between-cluster decomposition of a predictor relies on the observed cluster means, latent cluster means are generally used in SEM. The latent means in a multilevel SEM-framework avoid bias due to sampling error, which is typically associated with the observed cluster means in the MLM-framework (Lüdtke et al., 2008). And although the impossibility of the MLM-framework to deal with measurement error is a serious limitation, this does not pose an issue when the interest lies with the within-cluster effects. When the lower-level variables are assumed to be measured without error (Lüdtke et al., 2008) have shown that the estimator of the within-effect is unbiased (we will not repeat their proof here). Additionally, Lüdtke et al. (2008) reported a similar performance in terms of standard errors for the estimated within-effects, when using the observed mean versus the latent mean. Unfortunately, the MLM-approach can result in substantially biased estimates of between-effects, as well as severely underestimate the associated standard errors in the presence of upper-level measurement error. However, since Nesselroade and Molenaar (2016) have recently re-emphasized the importance of studying within-subject processes in lower-level designs (Molenaar, 2004, 2009), we will primarily focus on the estimation of these effects. Given that MLM and SEM perform similarly for within-cluster effects, we limit our exposition to the MLM-framework.

In the following sections, we first introduce our illustrating example and describe cluster-mean centering within the MLM-framework for main effect models. Next, we consider MLMs with lower-level interaction effects and enumerate the various existing modeling strategies proposed for estimating such effects. We demonstrate how different estimates (and standard errors) are found for the moderating effect, when applying these strategies to the diary data on intimacy and relationship feelings. In a next step, we explore why and when those centering approaches perform differently by means of a simulation study. Finally, we discuss the interpretation of the parameters for the different modeling strategies and end with a short discussion.

Illustrating example

We consider longitudinal diary data on sexual behavior from a Flemish study in 66 heterosexual couples (Dewitte, Van Lankveld, Vandenberghe, & Loeys, 2015). Every morning during three weeks, participants were asked about their sexual and intimate behavior since the last time they had filled out their morning diary (i.e., sexual behavior over the past 24 hours). Every evening, the participants were asked to report on their individual, relational, and partner-related feelings and behavior, experienced during that day. In this article, we limit our focus to the reports of the 66 male partners. Because the diary reports were not always completed meticulously over the course of the 21 days, the number of observations per participant ranges from 5 to 21, with a median cluster size of 18. In total, we have 1127 observations clustered within 66 men, implying a missing rate of about 19%. The variables of interest are the extent (on a 7-point scale from "not at all" to "very much") to which they report that intimate acts had occurred with their partner (described as the amount of kissing, cuddling, and caressing), the men's daily reports of masturbation (defined as any sexual act that involved self-stimulation in the absence of their partner), as well as their daily evening reports on positive relationship feelings. The latter were obtained by averaging the scores (on a seven-point scale) on nine items (the extent to which they felt happy, satisfied, understood, supported, accepted, loved, in love, connected, and close). The research question we will focus on, considers the contribution of intimacy to next-day positive relationship feelings within a man, and to what extent that the occurrence of masturbation during the previous day (yes or no) changes this effect.

Centering of main effects in multilevel models

In this section, we first explain why a difference in within- and between-subject effects may result from omitted variable bias at the subject-level. Let X_{ij} denote the predictor and Y_{ij} the outcome of individual j $(j = 1, \ldots, N)$ at time i $(i = 1, \ldots, n_j)$. In our example, X_{ij} and Y_{ij} represent the daily measurements of intimacy and next day's positive relational feelings, respectively. As mentioned in the introduction, it is not unlikely that the daily measurements of intimacy and positive relational feelings are both affected by unmeasured stable (personality) traits of the individual. We referred to such unmeasured common causes of X_{ij} and Y_{ij} as an unmeasured upper-level confounder, which we will from now on denote by b_j.

Consider a simple causal model for the effect of X_{ij} on Y_{ij} that takes an unmeasured subject-level confounder b_j into account:

$$E(Y_{ij} \mid X_{ij}, b_j) = \beta_0 + \beta X_{ij} + b_j, \qquad (1)$$

where we assume that the unmeasured confounder has an additive effect on the outcome. The left panel of Figure 1 represents the corresponding data-generating process. For a given subject, the β-parameter reflects the average increase in the outcome for a one-unit increase in the predictor. As such, this parameter can be interpreted as the within-person effect of X_{ij} on Y_{ij}. Several remarks deserve some additional attention. First, in order for β to have a causal interpretation, the predictor X_{ij} should temporally precede Y_{ij}. For example, in our illustration we aim to estimate the causal effect of intimacy on next day's positive relationship feelings, implying a clear temporal ordering. Second, we assume a time-constant effect of X_{ij} on Y_{ij}; there is no reason to assume that the effect on day one is any different from the effect on day two. Third, we assume the absence of any unmeasured lower-level confounders of the X_{ij}–Y_{ij} relationship. That is, given the unmeasured personality traits for example, we do not allow for further occasion-specific unmeasured common causes of X and Y. The question that we want to address now is: how can β be unbiasedly estimated, despite the presence of the unmeasured upper-level confounder b_j?

Naively, we could consider the following multilevel model:

$$E(Y_{ij} \mid X_{ij}, b_j) = \gamma_0 + \gamma X_{ij} + u_j, \qquad (2)$$

Note that to clearly contrast estimation model (2) to data-generating process (1), we rely on different notations here, as well as in the remainder of the article. Fixed effect parameters will be denoted by γ's and random effects by u_j in estimation models, while β's and b_j will represent these effects in the true causal models. To fit model (2), we could simply rely on standard multilevel modeling approaches. Unfortunately, an important but often ignored assumption in hierarchical linear modeling requires the random effect u_j in (2) to be uncorrelated with the predictor X_{ij} (McNeish, Stapleton, & Silverman, 2016). This assumption is violated in case of upper-level endogeneity. Consequently, the naive MLM-estimator (based on maximum likelihood, restricted maximum likelihood, or feasible generalized least squares, abbreviated FGLS) that aims to estimate β in model (1), and which we will refer to as $\hat{\gamma}^{RE}$, will suffer from omitted variable bias (Raudenbush & Bryk, 2002; Castellano, Rabe-Hesketh, & Skrondal, 2014).

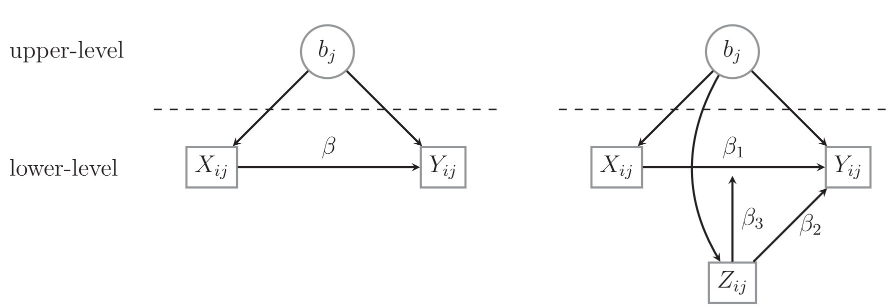

Figure 1. Left panel: unmeasured subject-level confounding of the X_{ij}–Y_{ij} relationship. Right panel: lower-level interaction model with unmeasured subject-level confounding.

In a similar vein, it is important to stress that lagged variables should not be added to model (2), i.e.:

$$E(Y_{ij} \mid X_{ij}, Y_{i-1,j}, u_j) = \gamma_0 + \gamma_1 X_{ij} + \gamma_2 Y_{i-1,j} + u_j, \quad (3)$$

Since model (3) applies to all time points, u_j has a direct effect on $Y_{i-1,j}$. However, if u_j affects $Y_{i-1,j}$, it cannot be statistically independent of $Y_{i-1,j}$ at the same time. The violation of this independence assumption in traditional hierarchical linear modeling can bias both the coefficient for the lagged dependent variable, as well as the coefficients for the other variables (Allison, 2015).

One possible way to deal with omitted variable bias in model (2) is to rely on the fixed effects approach (Mundlak, 1978), where u_j is treated as fixed rather than random. This approach is very popular within the econometrics literature (Wooldridge, 2010) and has recently resurfaced in behavioral science literature (Castellano et al., 2014). In practice, N dummy variables dn_j (i.e., one for each subject) are created in a way that $dn_j = 1$ if $n = j$, and $dn_j = 0$ when $n \neq j$ $(n = 1, \ldots, N)$. Consequently, Y_{ij} is regressed on $d1_{-}j, \ldots, dN_j$ and x_{ij}:

$$\begin{aligned} &E(Y_{ij} \mid X_{ij}, d1_j, \ldots, dN_j) \\ &= \gamma_1^* d1_j + \gamma_2^* d2_j + \cdots + \gamma_N^* dN_j + \gamma_{FE} X_{ij} \quad (4) \end{aligned}$$

Under causal data-generating model (1), the OLS-estimator for γ_{FE}, denoted by $\hat{\gamma}_{\text{FE}}$ (obtained from estimation model (4)), represents an unbiased estimator for β (Wooldridge, 2010). Intuitively, this can be understood by the fact that the predictors in (4) are allowed to be correlated (in contrast to the predictor and random intercept in model (2)). One side effect of the fixed effects approach is that it cannot be used to investigate between-subject effects, as between-subject characteristics are perfectly collinear with the dummies.

A possible alternative that can deal with omitted variable bias and additionally allows the estimation of both within- and between-effects, is to rely on group-mean centering (i.e., the CWC-approach). That is, the predictor X_{ij} is separated into a between- (i.e., $\overline{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$) and a within-subject (i.e., $X_{ij} - \overline{X}_j = X_{ij}^c$) component within the MLM-framework. As such, we consider the following model, originally proposed by Neuhaus and Kalbfleisch (1998):

$$E(Y_{ij} \mid X_{ij}, u_j) = \gamma_0 + \gamma_W X_{ij}^c + \gamma_B \overline{X}_j + u_j \quad (5)$$

with u_j assumed i.i.d. $\sim N(0, \tau^2)$ and independent of the predictors. Goetgeluk and Vansteelandt (2008) proved that the estimator $\hat{\gamma}_W$ from model (5) is consistent (i.e., asymptotically unbiased) for β in model (1), even in the presence of unmeasured upper-level confounding of X_{ij} and Y_{ij}. The rationale behind this is that by subject-mean centering the predictor, any subject-specific effects are effectively eliminated. Relying on simple OLS-estimators for γ_W and γ_B, we see that $\hat{\gamma}_W = \frac{\text{cov}(Y_{ij}, X_{ij} - \overline{X}_j)}{\text{var}(X_{ij} - \overline{X}_j)}$ and $\hat{\gamma}_B = \frac{\text{cov}(\overline{Y}_j, \overline{X}_j)}{\text{var}(\overline{X}_j)}$, which will converge to β and $\beta + \frac{\text{cov}(b_j, \overline{X}_j)}{\text{var}(\overline{X}_j)}$, respectively, under model (1). These two expressions clearly illustrate two important points. First, cluster-mean centering the predictor permits unbiased estimation of the within-person effect under upper-level endogeneity in causal model (1). Second, when b_j is a confounder of the X_{ij}–Y_{ij} relationship, this implies that $\text{cov}(b_j, \overline{X}_j) \neq 0$, and that $\hat{\gamma}_B$ will no longer converge to β. In other words, upper-level endogeneity elicits differences in the between- and within-subject effects. Only in the absence of upper-level endogeneity in model (1) (i.e., $\text{cov}(b_j, \overline{X}_j) = 0$), will $\hat{\gamma}_B$ be equal to $\hat{\gamma}_W$.

Note that the naive MLM-estimator $\hat{\gamma}^{RE}$ actually represents a weighted combination of $\hat{\gamma}_W$ and $\hat{\gamma}_B$ (Raudenbush & Bryk, 2002, p. 137); in balanced designs (i.e., with $n_j = n$ for all j), we see that:

$$\hat{\gamma}^{RE} = \frac{W_1\hat{\gamma}_B + W_2\hat{\gamma}_W}{W_1 + W_2}, \text{ with } W_1 = \widehat{\text{var}}(\hat{\gamma}_B)^{-1} \text{ and } W_2 = \widehat{\text{var}}(\hat{\gamma}_W)^{-1},$$

making $\hat{\gamma}^{RE}$ an uninterpretable blend of both effects. Also note that, since $\text{cov}(X_{ij} - \overline{X}_j, \overline{X}_j) = 0$, the within- and between-subject predictors are independent. As such, the cluster means can be dropped from estimation model (5) when the within-effect is the only quantity of interest:

$$E(Y_{ij} \mid X_{ij}, u_j) = \gamma_0 + \gamma_W X_{ij}^c + u_j \quad (6)$$

Furthermore, it is interesting to note that the fixed effect estimator $\hat{\gamma}^{\text{FE}}$ and the within-subject estimator $\hat{\gamma}_W$ are identical in balanced designs (Wooldridge, 2010).

Similar to Greenland (2002), Goetgeluk and Vansteelandt (2008), Brumback, Dailey, Brumback, Livingston, and He (2010), we argue that models (5) and (6) cannot be considered valid causal models. For example, in the longitudinal setting considered here, model (5) would imply that the future causes the past (i.e., future X_{ij} would cause past $Y_{i_0 j}$ for $i > i_0$, since X_{ij} is contained within $\overline{X}_j$). Also, when model (5) is interpreted as a causal model for the manipulated effect of X_{ij} for a single i, it would conflict with causal model (1) unless $\gamma_W = \gamma_B = \beta$. As mentioned before, the individual causal effect of a one-unit increase in X_{ij} is represented by β in model (1), while this is $\gamma_W(1 - 1/n_j) + 1/n_j\gamma_B$ in model (5); the latter expression only equals β when $\gamma_W = \gamma_B = \beta$. This remark does not degrade the usefulness of model (5), but it emphasizes that model (5) should be viewed as an estimation model rather than a causal one.

What are the principal implications for substantive researchers? Most importantly, that model (5) can be used as the vehicle to estimate the parameter of interest. In our example, we want to determine the effect of a one-unit increase in intimacy on next day's positive relationship feelings within a person. Unlike γ in model (2), the parameter γ_W in model (5) will target that quantity of interest, even in the presence of unmeasured time-constant subject-specific confounders. As such, we look at settings in which model (1) (graphically represented in Figure 1) rather than model (5) represents the true causal model. However, in these settings, model (5) still correctly describes the conditional association of Y_{ij} given X_{ij} and the independent subject effect u_i. In other words, while both models might be valid at the same time, model (1) constitutes the causal model, whereas model (5) represents an estimation model invoked to circumvent the issue that b_i is associated with X_{ij} (so that we can unbiasedly estimate β).

Centering of lower-level interactions in multilevel models

Researchers' interest is often not limited to assessing main effects only. In our illustrating example, researchers may want to know if the effect of intimacy on the following day's positive relational feelings differs according to whether or not the participant has masturbated during the previous day. Instead of model (1), we now assume a causal model in which an interaction effect is included:

$$E(Y_{ij} \mid X_{ij}, Z_{ij}, b_j) = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} + \beta_3 X_{ij} Z_{ij} + b_j, \tag{7}$$

with Z_{ij} being the moderator at time i in individual j. Since both X_{ij} and Z_{ij} are measured at the lower-level, we have a setting with a lower-level interaction. The right panel of Figure 1 graphically represents the assumed data-generating process. Note that an arrow-on-arrow notation was used to indicate the moderating effect of Z.

The parameter β_3 in model (7) reflects the moderating effect for a given subject, i.e., the extent to which the effect of X_{ij} on Y_{ij} varies for different values of Z_{ij}. In our example, such an effect might translate into: how does the effect of intimacy on next day's positive relationship feelings change within a participant when the man has masturbated versus when he has not? The interpretation of the main effects β_1 and β_2 in (7) on the other hand, depends on whether X_{ij} and/or Z_{ij} are grand mean centered. When X_{ij} is grand mean centered, β_2 reflects the effect of masturbation on next day's positive relationship feelings within a subject at the sample average level of intimacy. If X_{ij} were not grand mean centered, β_2 would capture the effect of masturbation at the zero-level of intimacy. This, however, would not provide a very useful interpretation, since intimacy is measured on a 1–7 scale. Similarly, when Z_{ij} is grand mean centered (i.e., in our example, the sample proportion of days with masturbation is subtracted from the raw scores), β_1 reflects the effect of a one-unit increase in intimacy on the next day's positive relationship feelings within a participant, averaged over days with and without masturbation. When both X_{ij} and Z_{ij} are grand mean centered, the intercept β_0 can be interpreted as the average positive relationship feelings over all participants and days. As such, grand-mean centering of both continuous and binary predictors in interaction models provides useful interpretations of the main effects; we will therefore assume that X and Z are grand mean centered during the remainder of this article. However, in order to avoid notational burden, we will not introduce any new notation to indicate this.

The researcher's primary interest now lies in estimating β_3. But how should β_3 be estimated? Naively, we may again consider a traditional MLM-approach:

$$E(Y_{ij} \mid X_{ij}, Z_{ij}, u_j) = \gamma_0 + \gamma_1 X_{ij} + \gamma_2 Z_{ij} + \gamma_3 X_{ij} Z_{ij} + u_j. \tag{8}$$

Note that here too, we define the parameters γ and u within the estimation model, whereas β and b are used in the causal model. Given the standard assumption of independence of the random effect and predictors in model (8), the naive MLM-estimator of the interaction effect (which we denote $\hat{\gamma}_3^{\text{RE}}$) will once again suffer from omitted variable bias. To address such unmeasured upper-level confounding, we may—similar to the main effects model—rely on separating within- from between-effects.

As was already mentioned in the introduction, two different strategies for centering lower-level interactions have been suggested. The first approach, advocated by Josephy et al. (2015), first multiplies X_{ij} with Z_{ij}, after which the cluster mean average of this product term is subtracted. As such, the "P1C2" estimation model amounts to:

$$E(Y_{ij} \mid X_{ij}, Z_{ij}, u_j) = \gamma_0 + \gamma_1 X_{ij}^c + \gamma_2 Z_{ij}^c + \gamma_3 (XZ)_{ij}^c + u_j \tag{9}$$

with $X_{ij}^c = X_{ij} - \overline{X}_j$, $Z_{ij}^c = Z_{ij} - \overline{Z}_j$ and $(XZ)_{ij}^c = X_{ij}Z_{ij} - \overline{XZ}_j$ (where $\overline{XZ}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}Z_{ij}$). Under data-generating model (7), $\hat{\gamma}_0$, $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$ of the P1C2-approach consistently (i.e., asymptotically unbiased) estimate β_0, β_1, β_2, and β_3 (Goetgeluk & Vansteelandt, 2008; Josephy et al., 2015). As such, the estimators $\hat{\gamma}_0$, $\hat{\gamma}_1$,

$\hat{\gamma}_2$, and $\hat{\gamma}_3$ share the same interpretation as β_0, β_1, β_2, and β_3 (see below model (7)).

It is also possible to add all corresponding between-effects to estimation model (9), i.e.:

$$E(Y_{ij} \mid X_{ij}, Z_{ij}, u_j) = \gamma_0 + \gamma_1 X_{ij}^c + \gamma_2 Z_{ij}^c + \gamma_3 (XZ)_{ij}^c + \gamma_4 \overline{X}_j + \gamma_5 \overline{Z}_j + \gamma_6 \overline{XZ}_j + u_j \quad (10)$$

We will refer to estimation model (10) as the "P1C2+" approach. Interestingly, since the within-predictors are independent of the between-predictors in this model, the estimated within-effects $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$ from models (9) and (10) are identical in balanced designs.

The second approach is suggested by Preacher et al. (2016), who are very explicit on their centering convictions in multilevel SEM (MSEM) models. If we ignore the distinction between centering at the observed versus the latent cluster means (Lüdtke et al., 2008), Preacher et al. (2016) distinctly argued that $X_{ij}Z_{ij}$ should not be separated into a within-part $X_{ij}Z_{ij} - \overline{XZ}_j$ and a between-part $\overline{XZ}_j$. These authors reason that "using these as predictors does not lead to interpretable effects, because researchers are not interested in the effects of product terms" (p. 191). When solely focusing on within-effects, the multilevel model proposed by (Preacher et al., 2016) (with observed rather than latent cluster means), can be written as:

$$E(Y_{ij} \mid X_{ij}, Z_{ij}, u_j) = \gamma_0 + \gamma_1 X_{ij}^c + \gamma_2 Z_{ij}^c + \gamma_3 X_{ij}^c Z_{ij}^c + u_j \quad (11)$$

We refer to estimation model (11) as the "C1P2"-approach. In their paper, Preacher et al. (2016) also described a more complete model that additionally includes cross- and between-level effects:

$$\begin{aligned} E(Y_{ij} \mid X_{ij}, Z_{ij}, u_j) &= \gamma_0 + \gamma_1 X_{ij}^c + \gamma_2 Z_{ij}^c + \gamma_3 X_{ij}^c Z_{ij}^c + \gamma_4 \overline{X}_j + \gamma_5 \overline{Z}_j \\ &\quad + \gamma_6 \overline{X}_j \overline{Z}_j + \gamma_7 \overline{X}_j Z_{ij}^c + \gamma_8 \overline{Z}_j X_{ij}^c + u_j \quad (12) \end{aligned}$$

which we will refer to as the "C1P2++" approach. Model (12) contained four different interaction effects: a within- subject interaction (captured by the parameter γ_3), a between-subject interaction (captured by γ_6) and two cross-level interactions (captured by γ_7 and γ_8). Note that since $\mathrm{Cov}(X_{ij}^c Z_{ij}^c, \overline{X}_j \overline{Z}_j)$ is not necessarily zero, the estimated parameters of the within-effects in the C1P2 and C1P2++ approaches are no longer identical in balanced designs (unlike in P1C2 and P1C2+).

Ryu (2015) also considered MSEM for estimating lower-level interactions in multilevel data, but in contrast to Preacher et al. (2016), Ryu relied on an earlier MSEM approach (Muthén, 1990). The latter decomposes the observed data into between- and pooled within-covariances, while fitting separate within- and between-models using the multi-group techniques of SEM. This multi-group approach does not allow for missing data or unbalanced cluster sizes, but more importantly, it cannot account for cross-level interactions. Ryu (2015) considered three types of centering: no centering (UN), grand-mean centering (CGM), and centering within clusters (CWC). First of all, MSEM with uncentered lower-level variables (UN) employs latent cluster means to define the various upper and lower-level variables. This UN approach therefore corresponds to model (10), where the observed means are replaced by their latent counterparts (denoted with a tilde):

$$\begin{aligned} &E(Y_{ij} \mid X_{ij}, Z_{ij}, u_j) \\ &\quad = \gamma_0 + \gamma_1 (X_{ij} - \tilde{X}_j) + \gamma_2 (Z_{ij} - \tilde{Z}_j) + \gamma_3 (X_{ij}Z_{ij} - \widetilde{XZ}_j) \\ &\quad\quad + \gamma_4 \tilde{X}_j + \gamma_5 \tilde{Z}_j + \gamma_6 \widetilde{XZ}_j + u_j \quad (13) \end{aligned}$$

Here, X_{ij} and Z_{ij} are not grand mean centered. Second, Ryu (2015)'s CGM approach only differs from the UN approach in that X_{ij} and Z_{ij} are first grand mean centered. Third, the CWC-approach described by Ryu (2015) uses the observed cluster means as level 2 covariates. As such, the corresponding estimation model can be written as:

$$\begin{aligned} E(Y_{ij} \mid X_{ij}, Z_{ij}, u_j) &= \gamma_0 + \gamma_1 X_{ij}^c + \gamma_2 Z_{ij}^c + \gamma_3 X_{ij}^c Z_{ij}^c \\ &\quad + \gamma_4 \overline{X}_j + \gamma_5 \overline{Z}_j + \gamma_6 \overline{X}_j \overline{Z}_j + u_j \end{aligned} \quad (14)$$

and will be referred to as "C1P2+." We employ this labeling, since the predictors are centered first and only then multiplied as in Preacher et al. (2016), but unlike model (12) it does not include any cross-level interactions.

Let us now illustrate how the P1C2- and C1P2-approaches may lead to different estimates of the moderation effect, by means of our example data. To estimate the moderating effect of masturbation on the effect of intimacy on next day's positive relationship feelings, we consider the five different estimation models (9), (10),(11), (12), and (14). In these models, X_{ij}, Z_{ij}, and Y_{ij} denote the grand mean centered intimacy, the grand mean centered masturbation, and next day's positive relationship feelings, respectively. Estimated parameters with associated standard errors, test statistics, and p-values of all within-cluster effects are summarized in Table 1, together with estimated random intercepts and residual variances.

From these results, we can deduce several things. First, as already stated in the previous section, P1C2 and P1C2+ yield identical results for all within-effects in balanced designs. In our example, the data are not perfectly balanced due to a small amount of missingness,

Table 1. The parameter estimates, standard errors (s.e.) with associated *t*-statistics, degrees of freedom, and *p*-values, for the intercept, the within-subject main effects of intimacy and masturbation, as well as their within-subject interaction effect on next day's positive relationship feelings. Additionally, the estimated random intercept variances and residual error variances are provided. Five different estimation approaches are considered: P1C2, P1C2+,C1P2, C1P2+ and C1P2++.

	Intercept		*Intimacy*		*Masturbation*		*Interaction*		*Variances*	
Parameter	Estimate (s.e.) t-value (df)	*p*-value	Estimate (s.e.) t-value (df)	*p*-value	Estimate (s.e.) t-value (df)	*p*-value	Estimate (s.e.)	*p*-value	Random intercept	Residual
P1C2	5.183 (0.102) 50.933 (65)	<.001	0.079 (0.015) 5.443 (1015)	<.001	− 0.151 (0.079) − 1.907 (1015)	.057	− 0.075 (0.039) − 1.930 (1015)	.054	0.6373	0.5576
P1C2+	5.180 (0.080) 64.667 (62)	<.001	0.079 (0.015) 5.444 (1015)	<.001	− 0.150 (0.079) − 1.890 (1015)	.059	− 0.075 (0.039) − 1.928 (1015)	.054	0.3806	0.5576
C1P2	5.179 (0.102) 50.902 (65)	<.001	0.080 (0.015) 5.523 (1015)	<.001	− 0.163 (0.080) − 2.041 (1016)	.042	− 0.102 (0.050) − 2.041 (1030)	.042	0.6366	0.5574
C1P2+	5.197 (0.084) 62.232 (61)	<.001	0.080 (0.015) 5.521 (1015)	<.001	− 0.160 (0.080) − 2.012 (1016)	.045	− 0.098 (0.045) − 1.968 (1038)	.049	0.3774	0.5574
C1P2++	5.197 (0.084) 62.236 (61)	<.001	0.080 (0.015) 5.520 (1013)	<.001	− 0.167 (0.080) − 2.091 (1014)	.037	− 0.096 (0.050) − 1.917 (1036)	.056	0.3773	0.5573

and as a consequence the estimates, standard errors and *p*-values of P1C2 and P1C2+ differ slightly. The estimated within-effects in the different C1P2-approaches, on the other hand, are much more discrepant. Second, the estimated moderating effect of masturbation is more pronounced in the C1P2-approaches compared to the estimates from P1C2. Even though all approaches point in the same direction (the positive effect of intimacy on next day's positive relationship is diluted if the man masturbated), the moderating effect is inflated by about 25% in the C1P2-approaches compared to P1C2. Third, the standard errors of the estimated interaction effect in the C1P2-approaches are about 25% larger than in P1C2. To gain further insights into the performance of the different estimation models, as well as into the precise quantities the different within-effect estimators are targeting, a simulation study is presented in the next section.

Simulation study

We consider five different simulation settings under causal model (7), where we assess the (relative) bias of the estimators and standard errors of the within-effects for the five different estimation models ((9), (10), (11), (12), and (14)), as well as their coverage and power. The bias is evaluated by contrasting the sample mean of the estimates from the 1000 simulated datasets to the true parameter value, through the use of a Wald-test. We report the relative bias of the parameter estimates, which is defined as the averaged difference of the estimated (e.g., $\hat{\beta}$) and true parameter value (e.g., β), divided by the latter. Equivalently, the relative bias of the standard errors is defined as the difference between the mean of the estimated standard errors and the empirical standard error, divided by the latter. A negative relative bias thus implies an underestimation of the true variability. The coverage is defined by the proportion of the 95%-confidence intervals that encompass their true parameter value, while the power is determined by the proportion of the 95%-confidence intervals that do not encompass zero.

Mimicking the two-level structure of our illustrating data, we simulated 1000 data sets which contain 66 clusters and 21 observations within each cluster, for five different settings. The true data-generating models for Z_{ij} and Y_{ij} are:

$$Z_{ij} = \alpha_0 + \alpha_1 X_{ij} + \upsilon_j^Z + \epsilon_{ij}^Z \quad (15)$$

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} + \beta_3 X_{ij} Z_{ij} + \upsilon_j^Y + \epsilon_{ij}^Y \quad (16)$$

In these models, we generate independent lower-level residuals, ϵ_{ij}^Z and ϵ_{ij}^Y, from standard normal distributions. The upper-level confounders, υ_j^Z and υ_j^Y, follow a multivariate standard normal distribution with a correlation equal to 0.5. As such, we induce unmeasured confounding of the Z_{ij}–Y_{ij} relationship with an additive effect on the outcome. Additionally, we fix $\alpha_0 = 0$, $\beta_0 = 0$, $\beta_1 = 0.1$, $\beta_2 = 0.15$, and $\beta_3 = -0.1$ in all settings, since these values approximately correspond to those seen in our illustrating example (Table 1). As will become apparent later this section, some of the estimation models will show bias in the interaction effect estimator. Since this bias depends on the distribution of X and Z, we will therefore vary the distribution of X (see Table 2). Also, as Preacher et al. (2016) showed that the within- and between-components of the product of two lower-level predictors depends on the covariance of the predictors that form the product, we will additionally vary the value of α_1 (see Table 2). Consequently, in the scenarios where $\alpha_1 \neq 0$, we see that $\mathrm{Cov}(X_{ij}, Z_{ij}) \neq 0$. Note that when $\alpha_1 \neq 0$, Z linearly depends on X and may be viewed as a mediator in the relation between X and Y.

For the first simulation setting, we generate a standard normally distributed X, whereas X is sampled from a

Table 2. A summary of the five different simulation settings. Each setting considers a different combination of an α_1-value and a distribution for X (e.g., $B(1, \Phi(\upsilon_j^X)) - 0.5$ reflects a mean centered Bernoulli variable with success probability $\Phi(\upsilon_j^X)$, with *Phi* representing the cumulative normal distribution an υ_j^X a standard normally distributed random effect). When $\alpha_1 \neq 0$, or when a random intercept for X is introduced that is correlated with the random intercepts for Z and Y (as in in Sim 5), the covariance between X_{ij} and Z_{ij}, $\mathrm{Cov}(X_{ij}, Z_{ij})$, will consequently differ from zero.

Simulation	α_1	Distribution of X	$\mathrm{Cov}(X_{ij}, Z_{ij})$
Sim 1	0.000	$N(0.00, 1.00)$	0.000
Sim 2	0.000	$B(1, 0.500) - 0.500$	0.000
Sim 3	− 0.200	$B(1, 0.500) - 0.500$	− 0.050
Sim 4	− 1.500	$B(1, 0.500) - 0.500$	− 0.375
Sim 5	− 1.500	$B(1, \Phi(\upsilon_j^X)) - 0.500$	− 0.235

zero-centered Bernoulli distribution with success probability .5 in settings 2–4. In simulation setting five, the true data-generating model for X_{ij} is:

$$\mathrm{Probit}(X_{ij} = 1) = \upsilon_j^X$$

with υ_j^X following a standard normal distribution. Furthermore, υ_j^X is correlated with υ_j^Z and υ_j^Y, with a correlation equal to 0.5. The latter implies the existence of an unmeasured upper-level confounder of X, Y, and Z, inducing an additional covariation between X_{ij} and Z_{ij}.

In the first two settings $\alpha_1 = 0$, while $\alpha_1 = -0.2$ in the third, and $\alpha_1 = -1.5$ in the fourth and fifth setting. Although setting $\alpha_0 = 0$ and $\beta_0 = 0$ implies that both X and Z already exhibit mean zero at the population level in all settings, we additionally grand mean center all variables in the samples prior to analysis. Additionally, all estimation models were fitted using the *lmer*-function from the *lme4* R-package. The R-code used to generate the simulated data is available in the supplementary material.

The means of the 1000 parameter estimates (with the relative bias), the mean of the standard errors (with the relative bias), the coverage and power of the estimators are summarized in Table 3. Estimators that show significant bias are displayed in boldface. Note that we only displayed the results for three of the five approaches, since P1C2+ and C1P2+ yield results identical to P1C2 and C1P2, respectively (for our balanced simulation data). Before we summarize the results, we re-iterate that the fixed effects approach results in the exact same estimates as obtained by the P1C2-approach.

Let us first focus on the bias. For the first and second simulation setting, where X and Z are independent, we do not observe bias (or more precisely, the relative bias is smaller than 5%, and not significant) for all within-effects under all approaches. In the third–fifth setting, Z depends on X in a linear fashion; when the absolute value of the

Table 3. Simulation results in the balanced setting. The means of the estimates (*Est.*) over the 1000 simulations (with the relative bias, *rel.bias*), the mean standard errors (*se*) over the 1000 simulations (with the relative bias), the coverage, and power for each parameter. Not all approaches are displayed since P1C2+ and C1P2+ provide the exact same results as P1C2 and C1P2, respectively. The true values of β_1, β_2, and β_3 are 0.1, 0.15, and −0.1; when there is significant bias, the mean of the estimates is depicted in boldface.

		$\hat{\gamma}_1$				$\hat{\gamma}_2$				$\hat{\gamma}_3$			
	Estimator	Est. (*rel.bias*)	*se* (*rel.bias*)	Coverage	Power	Est. (*rel.bias*)	*se* (*rel.bias*)	Coverage	Power	Est. (*rel.bias*)	*se* (*rel.bias*)	Coverage	Power
Sim 1	P1C2	0.101 (0.006)	0.028 (−0.118)	0.92	0.93	0.150 (0.001)	0.028 (0.011)	0.96	1.00	− 0.101 (0.012)	0.020 (0.026)	0.95	1.00
	C1P2	0.101 (0.006)	0.028 (−0.126)	0.92	0.92	0.150 (0.000)	0.028 (0.011)	0.96	1.00	− 0.101 (0.008)	0.029 (0.010)	0.95	0.94
	C1P2++	0.101 (0.007)	0.028 (−0.121)	0.93	0.93	0.150 (0.000)	0.028 (0.007)	0.96	1.00	− 0.101 (0.009)	0.029 (0.014)	0.95	0.94
Sim 2	P1C2	0.103 (0.030)	0.055 (−0.009)	0.95	0.46	0.150 (−0.001)	0.028 (0.026)	0.96	1.00	− 0.101 (0.006)	0.039 (−0.015)	0.95	0.72
	C1P2	0.103 (0.029)	0.055 (−0.007)	0.95	0.47	0.150 (−0.001)	0.028 (0.018)	0.96	1.00	− 0.099 (−0.011)	0.058 (−0.003)	0.95	0.41
	C1P2++	0.103 (0.029)	0.055 (−0.009)	0.95	0.46	0.150 (−0.001)	0.028 (0.022)	0.96	1.00	− 0.099 (−0.011)	0.058 (−0.003)	0.95	0.41
Sim 3	P1C2	0.103 (0.030)	0.055 (−0.007)	0.95	0.46	0.150 (−0.001)	0.028 (0.026)	0.96	1.00	− 0.101 (0.006)	0.039 (−0.015)	0.95	0.72
	C1P2	0.103 (0.029)	0.055 (−0.007)	0.95	0.46	0.150 (−0.001)	0.028 (0.018)	0.96	1.00	− 0.099 (−0.013)	0.058 (−0.002)	0.95	0.41
	C1P2++	0.103 (0.029)	0.055 (−0.007)	0.95	0.46	0.150 (−0.001)	0.028 (0.022)	0.96	1.00	− 0.099 (−0.011)	0.058 (−0.002)	0.95	0.41
Sim 4	P1C2	0.103 (0.028)	0.069 (0.007)	0.95	0.33	0.150 (−0.001)	0.028 (0.026)	0.96	1.00	− 0.101 (0.006)	0.039 (−0.015)	0.95	0.72
	C1P2	0.103 (0.027)	0.069 (0.009)	0.95	0.33	0.150 (−0.001)	0.028 (0.018)	0.96	1.00	**− 0.090** (−0.104)	0.055 (0.009)	0.95	0.37
	C1P2++	0.103 (0.027)	0.069 (0.009)	0.96	0.33	0.150 (−0.001)	0.028 (0.022)	0.96	1.00	− 0.099 (−0.011)	0.056 (0.011)	0.96	0.41
Sim 5	P1C2	0.102 (0.020)	0.079 (0.006)	0.95	0.25	0.150 (−0.001)	0.028 (0.000)	0.95	1.00	− 0.101 (0.009)	0.044 (0.007)	0.95	0.63
	C1P2	0.102 (0.022)	0.079 (0.005)	0.95	0.25	0.150 (−0.001)	0.028 (−0.004)	0.95	1.00	**− 0.083** (−0.166)	0.058 (0.002)	0.95	0.30
	C1P2++	0.102 (0.021)	0.079 (0.005)	0.95	0.24	0.150 (−0.001)	0.028 (0.000)	0.95	1.00	− 0.101 (−0.011)	0.060 (0.050)	0.95	0.39

effect of X on Z is increased (i.e., comparing simulation 4 to simulation 3), we observe bias (i.e., the relative bias is larger than 10% in absolute value, and found to be significant) in the estimator for the interaction effect in C1P2 and C1P2+. When X and Z are zero-mean centered symmetric distributions and Z is linear in X (as is the case in the third, fourth, and fifth setting), we see for the OLS-estimator of γ_3 under C1P2:

$$E(\hat{\gamma}_3) = \beta_3 \frac{\text{cov}[X_{ij}Z_{ij}, X^c_{ij}Z^c_{ij}]}{\text{var}[X^c_{ij}Z^c_{ij}]}$$

As pointed out by Croissant and Millo (2008) the OLS-estimators are equivalent to the maximum likelihood estimators (as obtained through the *lmer* function) in our simulations, since we are assuming normality, homoscedasticity, and no serial correlation of the errors. The derivation of the above expression can be found in the Appendix. Notably, the bias depends on the distribution of X, as well as on the absolute value of α_1. We can see that $\text{cov}[X_{ij}Z_{ij}, X^c_{ij}Z^c_{ij}]$ can be written as the sum of $\text{var}[X^c_{ij}Z^c_{ij}]$ and some other terms that depend on $\alpha_1^2\text{cov}[\overline{X}_j, (X_{ij} - \overline{X}_j)^2]$ and $\alpha_1^2\text{cov}[\overline{X}_j^2, (X_{ij} - \overline{X}_j)^2]$. While the latter two covariances are zero when the distribution of X_{ij} is Gaussian, these covariances no longer equal zero when the distribution of X_{ij} becomes Bernoulli (Dodge & Rousson, 2012). Interestingly, when all cross-level interactions are included in C1P2++, this bias for the interaction effect in C1P2 and C1P2+ disappears. In sum, we find that the estimators of the P1C2, P1C2+, and C1P2++ approaches target the exact same population parameters under the assumed data-generating model.

Next, we take a look at the precision and power. The mean standard error for the estimator of the interaction effect is substantially lower in the P1C2 approaches, compared to C1P2++ in all simulation settings. The mean standard errors of the main effect estimators, on the other hand, are similar across all approaches. Furthermore, from the relative bias of the estimated standard errors we can see that the empirical standard deviation and the mean of the estimated standard error closely correspond under all approaches, for the main and interaction effects (except for the main effect of X under the first simulation setting). As a consequence, we also observe appropriate coverages for these estimators. We also ran simulations with zero values for all lower-level effects (i.e., all β's equal to zero), and found appropriate type-I errors for all methods (results not shown), in line with the coverages reported. Importantly, given the higher precision of the estimated interaction effect under P1C2, we also observe the highest power for detecting the interaction under this approach. However, it should be noted that the simulation results describe average performances and, in practice, data may be encountered where the P1C2 approach yields a larger p-value for the interaction effect, compared to C1P2++.

So far, our simulation study only considered balanced data. Since our diary study was not always complete over the course of the 21 days, we repeated the above five simulation settings with a missingness pattern similar to the example data. More specifically, we introduced varying cluster sizes by sampling them as rounded values from a shifted beta-distribution, such that cluster sizes varied between 1–21 (with its mode around 18). The substantive findings from this unbalanced setting are essentially the same as in the balanced case (see Table 4). Note, however, that due to the unbalanced nature of the simulations, the estimators of the P1C2 and P1C2+ approaches, and of the C1P2 and C1P2+ approaches, are no longer identical.

We limit the results of our simulation studies to the settings presented here for two reasons. First, the specific settings we considered allow us to derive analytical expressions for the observed biases. Second, further simulation studies with different choices (e.g., non-symmetric distributions for X and Z, nonlinear associations between X and Z, ...) lead to similar conclusions: (1) both P1C2 and C1P2++ yield unbiased estimators for the interaction effect, (2) both exhibit an appropriate coverage of their 95% confidence intervals, but (2) P1C2 is always more precise. This conclusion can also be drawn from our illustrating example: we observe more precise estimators for the interaction effect in P1C2, compared to the C1P2 approaches.

What are the practical implications of these findings in terms of interpretation? First of all, we found that the parameters γ_1, γ_2, and γ_3 in estimation models (9) and (12) (i.e., the P1C2 and C1P2++ approaches) target the exact same population parameters and can hence be given the same interpretation. Considering the estimates of the P1C2 approach (and C1P2++, respectively) in our illustrating example, we see that masturbation dilutes the positive effect of intimacy on next day's positive relationship feelings with on average 0.075 units (0.096, respectively), for every unit increase in intimacy (Table 1). Averaging over days with and without masturbation, a one-unit increase in intimacy within a male individual will result in an 0.079 (0.080, respectively) increase in his next day's positive relationship feelings (Table 1). Equivalently, at average levels of intimacy, masturbation reduces next day's positive relationship feelings with on average 0.151 points (0.167 respectively) (Table 1).

Table 4. Simulation results in the unbalanced setting. The means of the estimates (*Est.*) (with relative bias), the mean standard errors (*se*) (with relative bias) the coverage, and power for each parameter and simulation setting. All approaches are displayed, as in unbalanced designs P1C2+ and C1P2+ no longer provide the same results as P1C2 and C1P2, respectively. The true values of β_1, β_2 and β_3 are 0.1, 0.15 and −0.1; when there is significant bias, the mean of the estimates is depicted in boldface.

		$\hat{\gamma}_1$				$\hat{\gamma}_2$				$\hat{\gamma}_3$			
Estimator		Est. (*rel.bias*)	*se* (*rel.bias*)	Coverage	Power	Est. (*rel.bias*)	*se* (*rel.bias*)	Coverage	Power	Est. (*rel.bias*)	*se* (*rel.bias*)	Coverage	Power
Sim 1	P1C2	0.101 (0.013)	0.032 (−0.061)	0.93	0.87	0.152 (0.011)	0.032 (0.009)	0.95	1.00	− 0.101 (0.013)	0.023 (−0.017)	0.96	0.99
	P1C2+	0.101 (0.013)	0.032 (−0.061)	0.93	0.87	0.152 (0.011)	0.032 (0.009)	0.95	1.00	− 0.101 (0.013)	0.023 (−0.017)	0.96	0.99
	C1P2	0.101 (0.012)	0.032 (−0.061)	0.94	0.86	0.152 (0.011)	0.032 (0.016)	0.95	1.00	− 0.101 (0.010)	0.035 (0.000)	0.94	0.83
	C1P2+	0.101 (0.012)	0.032 (−0.061)	0.94	0.86	0.152 (0.011)	0.032 (0.016)	0.95	1.00	− 0.101 (0.011)	0.035 (0.000)	0.94	0.83
	C1P2++	0.101 (0.012)	0.032 (−0.061)	0.93	0.86	0.152 (0.011)	0.032 (0.006)	0.95	1.00	− 0.101 (0.010)	0.035 (0.003)	0.94	0.84
Sim 2	P1C2	0.100 (0.004)	0.064 (0.019)	0.96	0.35	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.101 (0.008)	0.046 (0.011)	0.95	0.59
	P1C2+	0.100 (0.004)	0.064 (0.019)	0.96	0.35	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.101 (0.008)	0.046 (0.011)	0.95	0.59
	C1P2	0.101 (0.005)	0.064 (0.019)	0.95	0.35	0.149 (−0.008)	0.032 (0.006)	0.95	1.00	− 0.100 (0.001)	0.068 (−0.010)	0.95	0.31
	C1P2+	0.101 (0.005)	0.064 (0.019)	0.95	0.35	0.149 (−0.008)	0.032 (0.006)	0.95	1.00	− 0.100 (−0.001)	0.068 (−0.007)	0.95	0.31
	C1P2++	0.100 (0.004)	0.064 (0.017)	0.96	0.35	0.149 (−0.008)	0.032 (0.009)	0.95	1.00	− 0.100 (0.000)	0.068 (−0.009)	0.95	0.31
Sim 3	P1C2	0.100 (0.002)	0.065 (0.019)	0.95	0.34	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.101 (0.008)	0.046 (0.011)	0.95	0.59
	P1C2+	0.100 (0.002)	0.065 (0.019)	0.95	0.34	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.101 (0.008)	0.046 (0.011)	0.95	0.59
	C1P2	0.100 (0.003)	0.065 (0.017)	0.95	0.35	0.149 (−0.008)	0.032 (0.006)	0.95	1.00	− 0.100 (−0.001)	0.069 (−0.010)	0.95	0.31
	C1P2+	0.100 (0.003)	0.065 (0.017)	0.95	0.35	0.149 (−0.008)	0.032 (0.006)	0.95	1.00	− 0.100 (−0.003)	0.069 (−0.007)	0.95	0.31
	C1P2++	0.100 (0.001)	0.065 (0.017)	0.95	0.35	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.100 (0.000)	0.069 (−0.009)	0.95	0.31
Sim 4	P1C2	0.099 (−0.015)	0.080 (0.016)	0.96	0.25	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.101 (0.008)	0.046 (0.011)	0.95	0.59
	P1C2	0.099 (−0.015)	0.080 (0.016)	0.96	0.25	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.101 (0.008)	0.046 (0.011)	0.95	0.59
	C1P2	0.099 (−0.013)	0.081 (0.013)	0.96	0.25	0.149 (−0.008)	0.032 (0.006)	0.95	1.00	**− 0.088** (−0.119)	0.065 (−0.008)	0.95	0.28
	C1P2+	0.099 (−0.013)	0.081 (0.013)	0.96	0.25	0.149 (−0.008)	0.032 (0.006)	0.95	1.00	**− 0.088** (−0.120)	0.065 (−0.006)	0.95	0.29
	C1P2++	0.098 (−0.016)	0.081 (0.014)	0.96	0.25	0.149 (−0.009)	0.032 (0.009)	0.95	1.00	− 0.100 (−0.002)	0.066 (−0.009)−	0.95	0.34
Sim 5	P1C2	0.099 (−0.014)	0.093 (−0.019)	0.94	0.20	0.148 (−0.012)	0.032 (−0.019)	0.95	1.00	− 0.102 (0.016)	0.052 (0.002)	0.96	0.50
	P1C2	0.099 (−0.014)	0.093 (−0.019)	0.94	0.20	0.148 (−0.012)	0.032 (−0.019)	0.95	1.00	− 0.102 (0.016)	0.052 (0.002)	0.96	0.50
	C1P2	0.099 (−0.010)	0.093 (−0.021)	0.94	0.20	0.148 (−0.012)	0.032 (0.003)	0.95	1.00	**− 0.079** (−0.209)	0.069 (−0.024)	0.94	0.22
	C1P2+	0.099 (−0.010)	0.093 (−0.021)	0.94	0.20	0.148 (−0.011)	0.032 (0.003)	0.95	1.00	**− 0.079** (−0.211)	0.069 (−0.025)	0.94	0.23
	C1P2++	0.099 (−0.015)	0.093 (−0.019)	0.94	0.19	0.148 (−0.012)	0.032 (0.000)	0.95	1.00	− 0.099 (−0.015)	0.071 (−0.019)−	0.95	0.30

Discussion

This article compared two alternative approaches for the centering of lower-level interactions. In our simulation study, the P1C2-approaches outperformed the C1P2-approaches in estimating such interactions: (1) P1C2 results in more precise estimates of the interaction effect, compared to the three C1P2-approaches; (2) P1C2 is not affected by misspecification or omission of upper-level effects, in contrast to C1P2 (unless all cross-level interactions are included).

It can be argued that the data-generating models considered here are somewhat restrictive. However, it is important to note that the performance of the two prevailing approaches for centering interactions was explored in settings, where CWC is usually considered a good remedy. That is, we studied settings with additive effects for unmeasured upper-level confounders, because such effects can be effectively eliminated by CWC.

A first important assumption underlying data-generating model (7) constitutes homogeneous effects amongst subjects. In the presence of heterogeneous subject-effects, random slopes for X, Z, as well as for their interaction can be added to the estimation models. Fortunately, relying on estimation through a simple random intercept model such as (9) (which ignores any heterogeneity) will not introduce bias in the effect estimates, provided that the random slopes are independent of the predictors (Baird & Maxwell, 2016). In contrast, if the random slopes were to be correlated with the predictors, CWC would no longer effectively eliminate unmeasured upper-level heterogeneity; alternative approaches such as fixed-effect estimation or per-cluster analysis would then be required (Bates et al., 2014).

A second important assumption underlying data-generating model (7) entails the absence of unmeasured lower-level confounding. If, for example, daily intimacy, masturbation, and positive relational feelings were associated with an (unmeasured) daily positive mood (given unmeasured subject-specific confounders), this assumption would be violated. Since CWC only eliminates time-invariant confounding, we would expect biased effect estimators under unmeasured lower-level confounding. However, as recently pointed out by Loeys et al. (2016), the assessment of interaction effects in linear models often requires weaker "no-unmeasured-confounding" assumptions, compared to main effects. Hence, unbiased effect estimators for the interaction may still be found under relatively lenient assumptions.

Third, we limited our discussion to linear settings. As shown by Goetgeluk and Vansteelandt (2008), separating a within- from a between-effect in a random intercept model only yields a consistent estimator of the within-effect in the presence of upper-level confounding when the model is linear. For nonlinear models, it is possible to encounter an inconsistent estimator, though in practice this bias will often be small.

To summarize, when dealing with multilevel data, we recommend that careful consideration be given to the assumptions under which separating within- from between-effects yield valid results. When those assumptions are deemed plausible, CWC can be applied to unbiasedly estimate within-cluster effects. For the estimation of interaction effects, we advocate the P1C2-approach rather than the C1P2-approach, as the former is much more efficient. If researchers want to use the C1P2-approach (e.g., because of implementations in software packages for SEM), we recommend not to drop any cross-level or upper-level terms, even when they are not of interest.

Article Information

Conflict of Interest: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the article for publication.

Acknowledgments: The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

References

Allison, P. D. (2015). Don't Put Lagged Dependent Variables in Mixed Models. *retrieved from https://statisticalhorizons.com/lagged-dependent-variables*.

Baird, M. & Maxwell, S. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods*, *21*(2), 175–188.

Bates, M. D., Castellano, K. E., Rabe-Hesketh, S., & Skrondal, A. (2014). Handling correlations between covariates and random slopes in multilevel models. *Journal of Educational and Behavioral Statistics*, *39*(6), 524–549.

Brumback, B. a., Dailey, A. B., Brumback, L. C., Livingston, M. D., & He, Z. (2010). Adjusting for confounding by cluster using generalized linear mixed models. *Statistics and Probability Letters*, *80*(21-22), 1650–1654.

Castellano, K. E., Rabe-Hesketh, S., & Skrondal, a. (2014). Composition, context, and endogeneity in school and teacher comparisons. *Journal of Educational and Behavioral Statistics*, *39*(5), 333–367.

Croissant, Y. & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, *27*(2), 1–43.

Curran, P. J. & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619.

Dewitte, M., Van Lankveld, J., Vandenberghe, S., & Loeys, T. (2015). Sex in its daily relational context. *Journal of Sexual Medicine*, *12*(12), 2436–2450.

Dodge, Y. & Rousson, V. (2012). The complications of the fourth central moment. *The American Statistician*, *1305*(May), 2–5.

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138.

Goetgeluk, S. & Vansteelandt, S. (2008). Conditional generalized estimating equations for the analysis of clustered and longitudinal data. *Biometrics*, *64*(3), 772–780.

Greenland, S. (2002). A review of multilevel theory for ecologic analyses. *Statistics in Medicine*, *21*, 389–395.

Hofmann, D. a. & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*(5), 623–641.

Josephy, H., Vansteelandt, S., Vanderhasselt, M.-A., & Loeys, T. (2015). Within-subject mediation analysis in AB/BA crossover designs. *The International Journal of Biostatistics*, *11*(1), 1–22.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*(30), 1–21.

Loeys, T., Talloen, W., Goubert, L., Moerkerke, B., & Vansteelandt, S. (2016). Assessing moderated mediation in linear models requires fewer confounding assumptions than assessing mediation. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 352–374.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological methods*, *13*(3), 203–229.

McNeish, D., Stapleton, L., & Silverman, R. (2016). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, page Advance online publication.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology—this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218.

Molenaar, P. C. M. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*, 112–117.

Mundlak, Y. (1978). Pooling of time series and cross-section data. *Econometrica*, *46*, 69–86.

Muthén, B. O. (1990). *Mean and Covariance Structure Analysis of Hierarchical Data*, volume 62. Los Angeles, CA, ucla stati edition.

Nesselroade, J. R. & Molenaar, P. C. M. (2016). Some behaviorial science measurement concerns and proposals. *Multivariate Behavioral Research*, *51*(2–3), 396–412.

Neuhaus, J. & Kalbfleisch, J. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, *54*(2), 638–645.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*, *21*(2), 189–205.

Raudenbush, S. & Bryk, A. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods*. (Second ed.). Thousand Oaks, CA: Sage.

Ryu, E. (2015). The role of centering for interaction of level 1 variables in multilevel structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(4), 617–630.

Wang, L. P. & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*(1), 63–83.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press: Cambridge, MA.

Appendix: Bias of the interaction effect estimator under the P1C2 approach

Assume that the true models for Z and Y are:

$$Z_{ij} = \alpha_0 + \alpha_1 X_{ij} + \upsilon_j^Z + \epsilon_{ij}^Z \tag{A.1}$$

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} + \beta_3 X_{ij} Z_{ij} + \upsilon_j^Y + \epsilon_{ij}^Y \tag{A.2}$$

with ϵ_{ij}^Z and ϵ_{ij}^Y i.i.d. with mean zero and variance σ_Z^2 and σ_Y^2, respectively.

Consider the estimation model:

$$E[Y_{ij} \mid X_{ij}, Z_{ij}, u_j] = \gamma_0 + \gamma_1 X_{ij}^c + \gamma_2 Z_{ij}^c + \gamma_3 X_{ij}^c Z_{ij}^c + u_j, \tag{A.3}$$

where $X_{ij}^c = (X_{ij} - \overline{X}_j)$ and $Z_{ij}^c = (Z_{ij} - \overline{Z}_j)$.

The OLS-estimators for the parameters of model (A.3), under models (A.1) and (A.2) are given by $\Sigma^{-1}\Sigma_{VY}$ with $V_{ij} = (1 \ X_{ij}^c \ Z_{ij}^c \ X_{ij}^c Z_{ij}^c)'$, $\Sigma = E[V_{ij}V_{ij}']$ and $\Sigma_{VY} = \left(E[Y_{ij}] \ E[X_{ij}^c Y_{ij}] \ E[Z_{ij}^c Y_{ij}] \ E[X_{ij}^c Z_{ij}^c Y_{ij}]\right)'$.

Now, we have that:

$$V_{ij}V_{ij}' = \begin{pmatrix} 1 & X_{ij}^c & Z_{ij}^c & X_{ij}^c Z_{ij}^c \\ X_{ij}^c & X_{ij}^{c\,2} & X_{ij}^c Z_{ij}^c & X_{ij}^{c\,2} Z_{ij}^c \\ Z_{ij}^c & X_{ij}^c Z_{ij}^c & Z_{ij}^{c\,2} & X_{ij}^c Z_{ij}^{c\,2} \\ X_{ij}^c Z_{ij}^c & X_{ij}^{c\,2} Z_{ij}^c & X_{ij}^c Z_{ij}^{c\,2} & X_{ij}^{c\,2} Z_{ij}^{c\,2} \end{pmatrix}$$

Assuming that Z is linear in X, $E(X_{ij}) = E(Z_{ij}) = 0$, while also assuming a symmetric distribution for X, the expectation of $V_{ij}V'_{ij}$ simplifies to

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & \alpha_1 \text{var}[X^c_{ij}] \\ 0 & \text{var}[X^c_{ij}] & \alpha_1 \text{var}[X^c_{ij}] & 0 \\ 0 & \alpha_1 \text{var}[X^c_{ij}] & \alpha_1^2 E[X^{c\,2}_{ij}] + \text{var}[\epsilon^{Zc}_{ij}] & 0 \\ \alpha_1 \text{var}[X^c_{ij}] & 0 & 0 & \alpha_1^2 E[X^{c\,4}_{ij}] + \text{var}[\epsilon^{Zc}_{ij}]\text{var}[X^c_{ij}] \end{pmatrix},$$

with $\epsilon^{Zc}_{ij} = (\epsilon^Z_{ij} - \overline{\epsilon^Z}_j)$. In order to obtain the elements c_{kl} of Σ^{-1}, we need its determinant. After some tedious calculations, we find that $|\Sigma| = \text{var}[\epsilon^{Zc}_{ij}]\text{var}[X^c_{ij}]$ $(\alpha_1^2 \text{var}[X^{c\,2}_{ij}] + \text{var}[\epsilon^{Zc}_{ij}]\text{var}[X^c_{ij}])$, and $c_{21} = c_{24} = c_{31} = c_{34} = c_{42} = c_{43} = 0$, $c_{22} = \frac{\alpha_1^2}{\text{var}[\epsilon^{Zc}_{ij}]} + \frac{1}{\text{var}[X^c_{ij}]}$, $c_{23} = c_{32} = -\frac{\alpha_1}{\text{var}[\epsilon^{Zc}_{ij}]}$, $c_{33} = \frac{1}{\text{var}[\epsilon^{Zc}_{ij}]}$, $c_{41} = -\frac{\alpha_1 \text{var}[X^c_{ij}]}{\alpha_1^2 \text{var}[X^{c\,2}_{ij}] + \text{var}[\epsilon^{Zc}_{ij}]\text{var}[X^c_{ij}]}$, and $c_{44} = \frac{1}{\alpha_1^2 \text{var}[X^{c\,2}_{ij}] + \text{var}[\epsilon^{Zc}_{ij}]\text{var}[X^c_{ij}]}$.

As such, we can show that there is no bias in the OLS-estimator $\hat{\gamma}_1$ for β_1:

$$\begin{aligned} E[\hat{\gamma}_1] &= c_{22}E[X^c_{ij}Y_{ij}] + c_{23}E[Z^c_{ij}Y_{ij}] \\ &= (c_{22} + \alpha_1 c_{23})(\beta_1 + \beta_2\alpha_1)\text{var}[X^c_{ij}] + c_{23}\beta_2 \text{var}[\epsilon^{Zc}_{ij}] \\ &= \beta_1 \end{aligned}$$

Similarly, we find no bias in the OLS-estimator $\hat{\gamma}_2$ for β_2:

$$\begin{aligned} E(\hat{\gamma}_2) &= c_{32}E[X^c_{ij}Y_{ij}] + c_{33}E[Z^c_{ij}Y_{ij}] \\ &= c_{32}(\beta_1 + \beta_2\alpha_1)\text{var}[X^c_{ij}] \\ &\quad + c_{33}\alpha_1(\beta_1 + \beta_2\alpha_1)\text{var}[X^c_{ij}] + c_{33}\beta_2 \text{var}[\epsilon^{Zc}_{ij}] \\ &= \beta_2 \end{aligned}$$

And finally, we find for the OLS-estimator $\hat{\gamma}_3$ for β_3 that:

$$\begin{aligned} E(\hat{\gamma}_3) &= c_{41}E[Y_{ij}] + c_{44}E[X^c_{ij}Z^c_{ij}Y_{ij}] \\ &= \beta_3 \frac{\alpha_1^2 \text{cov}[X^2_{ij}, X^{c\,2}_{ij}]}{\alpha_1^2 \text{var}[X^{c\,2}_{ij}] + \text{var}[\epsilon^{Zc}_{ij}]\text{var}[X^c_{ij}]} \end{aligned}$$

with $E[X^{c\,2}_{ij}X_{ij}] = E[X^{c\,2}_{ij}(X^c_{ij} + \overline{X}_j)] = E[X^{c\,3}_{ij}] + \text{cov}[X^{c\,2}_{ij}, \overline{X}_j] = 0$ for a symmetric X.

As $\text{var}[X^c_{ij}Z^c_{ij}]$ can be re-expressed as:

$$\begin{aligned} \text{var}[X^c_{ij}Z^c_{ij}] &= \text{var}[X^c_{ij}(\alpha_1 X^c_{ij} + \epsilon^{Zc}_{ij}] \\ &= \alpha_1^2 \text{var}[X^{c\,2}_{ij}] + \text{var}[X^c_{ij}\epsilon^{Zc}_{ij}] \end{aligned}$$

and $\text{cov}[X_{ij}Z_{ij}, X^c_{ij}Z^c_{ij}]$ as

$$\begin{aligned} \text{cov}[X_{ij}Z_{ij}, X^c_{ij}Z^c_{ij}] &= \text{cov}\big[X_{ij}(\alpha_1 X_{ij} + \epsilon^Z_{ij}), X^c_{ij}(\alpha_1 X^c_{ij} + \epsilon^{Zc}_{ij})\big] \\ &= \alpha_1^2 \text{cov}[X^2_{ij}, X^{c\,2}_{ij}] \end{aligned}$$

we find that the bias factor for $\hat{\gamma}_3$ can be rewritten as:

$$\frac{\text{cov}[X_{ij}Z_{ij}, X^c_{ij}Z^c_{ij}]}{\text{var}[X^c_{ij}Z^c_{ij}]},$$

which will equal one when the distribution of X is normal, but will be smaller than one when X is Bernoulli distributed.