Published in final edited form as:

Addict Behav. 2019 July ; 94: 156–161. doi:10.1016/j.addbeh.2018.09.031.

A Straightforward Approach for Coping with Unreliability of Person Means when Parsing Within-person and Between-person Effects in Longitudinal Studies

Nisha C. Gottfredson

University of North Carolina Gillings School of Global Public Health, 135 Dauer Drive, Chapel Hill, NC 27599

Abstract

Longitudinal studies enable researchers to distinguish within-person (i.e., time-varying) from between-person (i.e., time invariant) effects by using the person mean to model between-person effects and person-mean centering to model within-person effects using multilevel models (MLM). However, with some exceptions, the person mean tends to be based on a relatively small number of observations available for each participant in longitudinal studies. Unreliability inherent in person means generated with few observations results in downwardly biased between-person and cross-level interaction effect estimates. This manuscript considers a simple, easy-to-implement, post-hoc bias adjustment to correct for attenuation of between-person effects caused by unreliability of the person mean. This correction can be applied directly to estimates obtained from MLM. We illustrate this method using data from a panel study predicting adolescent alcohol involvement from perceived parental monitoring, where parental monitoring was disaggregated into within-person (i.e., person-mean-centered) and between-person (i.e., person-mean) components. We then describe results of a small simulation study that evaluated the performance of the post-hoc adjustment under data conditions that mirrored those of the empirical example. Results suggested that, under a condition in which parameter bias is known to be problematic (i.e., moderate ICC_X , small n , presence of a compositional effect), it is preferable to use the bias-adjusted MLM estimates over the unadjusted MLM estimates for between-person and cross-level interaction effects.

Keywords

between-person effect; longitudinal; compositional effect; post-hoc adjustment; multilevel model

gottfredson@unc.edu.

Contributors

Nisha Gottfredson was responsible for all aspects of this manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest

The author declares no conflicts of interest.

Declarations of Interest

None.

Longitudinal designs provide an advantage over cross-sectional designs in that they allow researchers to distinguish and quantify associations that are due to within-person processes from those due to between-person processes. Multilevel regression models (MLM) are commonly used for analysis of longitudinal data, but effects of covariates that are measured repeatedly represent a mixture, or *aggregate*, of within- and between-person effects unless a proper centering approach is applied (Curran & Bauer, 2011; Enders & Tofighi, 2007; Kreft, De Leeuw, & Aiken, 1995; Raudenbush & Bryk, 2002; Wang & Maxwell, 2015). To disaggregate within- and between-person effects, repeatedly measured covariates are separated into their time-stable, between-person component and their time-varying, within-person component. For example, Hawes, Crane, Henderson, and colleagues (2015) tested whether an observed association between alcohol use and psychopathic features was due to long-term, stable effects of prolonged alcohol use (a between-person effect), or whether the association was due to shorter-term effects of high levels of alcohol use (a within-person effect). The between-person component (e.g., long-term alcohol use levels) is calculated by taking the average of all n repeated measures available for each individual, i , to arrive at the *person-mean* ($\bar{X}_i$). The within-person component (e.g., recent levels of heavy alcohol use) is calculated by person-mean centering to obtain a time (t)-specific deviation from the person mean ($X_{ti} - \bar{X}_i$). These two disaggregated components of X_{ti} are included as two separate predictors in the mixed model in order to obtain unique estimates of the within-person and between-person effects.

Equation 1 displays an example of how an analyst would use this centering approach to disaggregate effects of a single repeated covariate X_{ti} in a MLM with fixed effects denoted by β and random effects denoted by u . β_W represents the within-person effect of $X_{ti} - \bar{X}_i$ and β_B represents the between-person effect of $\bar{X}_i$. In this example, the random intercept (u_{0i}), which represents inter-individual variation in $\bar{y}_i$ (e.g., individual differences in baseline psychopathology), has variance τ_{00} . The random slope (u_{1i}), which represents inter-individual variation in the within-person effect of $X_{ti} - \bar{X}_i$ (e.g., subject-specific effects of heavy alcohol consumption on psychopathic symptoms), has variance τ_{11} , and the random intercept and random slope have a covariance of τ_{10} .

$$\hat{y}_{ti} = \beta_0 + \beta_B \bar{X}_i + \beta_W (X_{ti} - \bar{X}_i) + u_{0i} + u_{1i} (X_{ti} - \bar{X}_i) \quad (1)$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{10} \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

With longitudinal designs, in addition to an interest in time-varying and time-invariant covariate effects that are the focus of this manuscript, time trends are often of interest as well. Time-varying and time-invariant covariates may interact with time, as in the example model displayed in Equation 2. Here, the random effect for time indicates that there are individual differences with respect to rate of change. Polynomial effects of time are also

common when the number of waves exceeds three. An interaction between time and a time-invariant covariate (TIC), shown as β_{Bt} in Equation 2, indicates that individual differences in the TIC are associated with subject-specific rate of change in the dependent variable. An interaction between time and a time-varying covariate (TVC), shown as β_{Wt} in Equation 2, indicates that within-person perturbations in the TVC have different effects on y at different times (e.g., at different points on a developmental trajectory).

$$\hat{y}_{ti} = \beta_0 + \beta_B \bar{X}_i + \beta_W (X_{ti} - \bar{X}_i) + \beta_t \text{time}_{ti} + \beta_{Bt} \bar{X}_i * \text{time}_{ti} + \beta_{Wt} (X_{ti} - \bar{X}_i) * \text{time}_{ti} \quad (2)$$

$$+ u_{0i} + u_{1i} (X_{ti} - \bar{X}_i) + u_{2i} \text{time}_{ti}$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{bmatrix} \sim N \begin{pmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{pmatrix}$$

Failure to properly disaggregate within-person and between-person covariate effects using the centering approach shown in Equations 1 and 2 can result in meaningless results (Curran & Bauer, 2011; Curran, Lee, Howard, Lane, & MacCallum, 2012). Taking the most extreme case, if β_W and β_B have opposite signs, then failing to disaggregate could result in estimating a nonsignificant aggregate effect of X_{ti} . For example, Hussong, Hicks, Levy, et al. (2001) found no association between stable, person-level sadness and alcohol consumption among college students; however, college students with low quality friendships consumed more alcohol than usual following sadness (a within-person effect). Had Hussong et al. (2001) not disaggregated sadness into its stable and time-varying component, they may not have detected a significant effect of sadness and alcohol use. Even if an association had been detected, it would not have been clear whether the association was a result of a within-person or a between-person process.

Nevertheless, the disaggregation approach just described can result in seriously biased estimates of β_B under conditions that are routinely observed in longitudinal studies (Curran et al., 2012; Lüdtke, Marsh, Robitzsch, et al., 2008). This manuscript draws from results of Lüdtke et al. (2008) to explain why and when such bias occurs, and then proposes a straightforward post-hoc correction for attenuation that can be applied directly to estimates obtained from a multilevel model. The post-hoc correction is demonstrated using data on time-varying and time-invariant effects of parental monitoring on adolescent alcohol involvement, and results of a simulation study evaluating performance of the post-hoc correction are reported to validate the correction.

Unreliability and Bias in Between-Person Effects

The reliability with which the person mean is measured depends on the intra-class correlation (ICC) of X_{it} and on the number of repeated measures, n . The ICC measures stability of a repeated measure over time, or how correlated the repeated measures are with one another: in this sense, the ICC is directly associated with the reliability with which a person mean can be estimated. Reliability of the person mean also increases with n . Assuming X_{it} is measured without measurement error, the reliability of $\bar{X}_i$ is $(n * ICC_X) / (1 + (n - 1) * ICC_X)$ (Brown, 1910; Lüdtke et al., 2008; Spearman, 1910). The top panel of Figure 1 displays the calculated reliability of $\bar{X}_i$ as a function of the n (from 3 to 15) and ICC_X within ranges typically observed (.2, .5, and .8). From this figure, we can see that n has little influence on reliability when the ICC_X is very high (indicating little variation in X_i from one time point to the next), but that there is a steep influence of n at more moderate ICC_X values; reliability is never high within the typical range of n when the ICC_X is low (indicating very little stability in X_i over time).

Unreliability of predictors attenuates regression effect estimates and reduces power to draw inferences. Drawing again from Lüdtke et al. (2008), when the between-person effect, β_B , is estimated using $\bar{X}_i$ to measure stable variance in X_{it} , bias in $\hat{\beta}_B$ is a function of ICC_X and n , as well as the magnitude of the difference between the within-person and between-person effects ($\beta_W - \beta_B$). This difference has been called the *compositional effect* in discussions of group-level contextual effects versus individual-level effects; for continuity, I use the term in the longitudinal context as well (Raudenbush & Bryk, 2002). A non-null compositional effect in a longitudinal setting might arise if the stable meaning of a repeatedly assessed predictor differs from the meaning of a fluctuation in the value of the same predictor. For instance, adolescent substance use might be lower for adolescents whose parents have high, stable levels of monitoring, but adolescent substance use might be higher during waves in which parents are being more attentive to adolescent behavior (perhaps in response to the adolescent's substance use). The formula for bias in the between-person effect is shown in Equation 3 (Lüdtke et al., 2008):

$$\text{Bias}(\hat{\beta}_B) = (\beta_W - \beta_B) * \frac{1}{n} * \frac{1 - ICC_X}{ICC_X + \frac{(1 - ICC_X)}{n}}. \quad (3)$$

The bottom panel of Figure 1 illustrates the absolute value of bias in the estimated standardized between-person effect at ICC_X values of .2, .5, and .8, as a function of n and the compositional effect. Unless n is very small and the compositional effect is very high, little bias is observed when ICC_X is high, regardless of n or the magnitude of the compositional effect. Bias is negligible at moderate ICC levels only when n is fairly large or when the compositional effect is small. The number of repeated assessments has a large influence on bias at lower ICC values, and bias is severe except when n is high and the compositional effect is low.

In practice, although a full range of ICCs has been observed across a range of studies, in general ICC_X and n tend to be slightly positively associated, meaning that researchers with more intensive longitudinal studies (e.g., ecological momentary assessment studies) will tend to have highly reliable measures of TICs and unbiased estimates of their effects, whereas researchers using panel designs tend to measure X_i with lower reliability and will have more biased estimates of TIC effects (Bauer & Sterba, 2011). Whereas a researcher with intensive longitudinal measures and a highly stable predictor need not be concerned with bias resulting from estimating the between-person effect using a person-mean, a researcher with fewer than five repeated assessments of an unstable predictor should be concerned.

Post-Hoc Bias Correction for Multilevel Models

For analysts using MLM to estimate disaggregated within-and between-person effects, I propose the use of a post-hoc correction for attenuation based on the classic formula for disattenuating correlation-based estimates. When the reliability of two correlated measures are known, the attenuated correlation estimate ($\hat{r}_{yx}$) can be disattenuated using the following formula (Spearman, 1904, 1907):

$$\tilde{r}_{yx} = \frac{\hat{r}_{yx}}{\sqrt{r_{xx}r_{yy}}}. \quad (4)$$

This formula can be applied to correct attenuation in $\hat{\beta}_B$ that is due to small n and ICC_X less than unity. Per the Spearman-Brown prediction formula, the reliability of X_{it} can be estimated empirically using $\hat{r}_{XX} = (\bar{n} * ICC_X) / (1 + (\bar{n} - 1) * ICC_X)$. The reliability of Y_{it} is assumed to be 1 (Lüdtke, Marsh, Robitzsch et al., 2011). The reliability formula uses the average within-group sample size, $\bar{n}$, to allow for unbalanced designs. The bias correction would then be implemented as follows:

$$\tilde{\beta}_B = \frac{\hat{\beta}_B}{\sqrt{\frac{\bar{n} * ICC_X}{1 + (\bar{n} - 1) * ICC_X}}}. \quad (5)$$

It is important to note that this correction is meant for unreliability due to sampling variability of repeated measures, and not for unreliability due to measurement error.

Implementation of this simple post-hoc correction is demonstrated using longitudinal data on parental monitoring and drinking behavior in a sample of adolescents. Validity of the method is assessed with a small-scale simulation study that was designed to mirror the data conditions in the empirical example.

Empirical Demonstration

Method

Sample and Procedure.—Data are from a longitudinal study of $N=6,998$ adolescents in grades 6 through 12 for all middle and high schools within three North Carolina counties. The study followed an accelerated cohort-sequential design and surveys were administered during school with assent from adolescents and their parents. Surveys were administered every semester for six waves, and the seventh wave was administered one year after the sixth wave. The average number of waves completed was 4.25.

Study protocols were approved by the UNC IRB. 50% of the students were male, 52% of the students identified as White, 37% identified as Black, 4% identified as Hispanic/Latino, and 7% identified as another race or ethnicity. Ten percent of students lived in homes headed by a single parent, 36% of students lived in a household in which the highest level of education attained by either parent was high school or less, and 39% of students lived in a household in which at least one parent had obtained a bachelor's degree.

Measures.—*Perceived parental monitoring* was the average of three items from the Authoritative Parenting Index (Jackson, Henriksen, & Foshee, 1998) that were assessed about both parents at every wave. Items were averaged for mothers and fathers in order to obtain commensurate measures of adolescents with a single parent. Items were: “(S)he has rules I must follow,” “(S)he tells me when to come home,” and “(S)he makes sure I don't stay up too late.” Response options ranged from “Not like him/her” (0) to “Just like him/her” (3). The alpha reliability for this scale was $\alpha=.85$.

The dependent variable, *alcohol involvement* is a factor score estimate generated from 12 items that were assessed at each wave. Five items measured alcohol-related consequences (e.g. getting in trouble with parents) and seven items measured use and hazardous use (e.g., frequency of use in the past 3 months and becoming hungover). Measurement bias due to differential item functioning related to demographic variables, including age, was ameliorated by using a moderated nonlinear factor analysis model (Bauer, 2017; Gottfredson et al, 2018)

Data Analysis.—Rather than analyzing data by wave, time was incremented in semesters from the spring of grade 6 (Biesanz, Deeb-Sossa, Papadakis et al., 2004). Data were analyzed using the MIXED procedure in SAS version 7.4 using the multilevel model shown in Equation 6. Restricted Maximum Likelihood estimation was used because it has been shown to produce less biased variance component estimates than Full Information Maximum Likelihood (Singer & Willett, 2003). Parental monitoring was disaggregated into its within-person and between-person components using the centering procedure described above. We considered a cross-level interaction between grand mean centered, person-mean parental monitoring and person-mean-centered parental monitoring to allow for the possibility that the effect of time-varying perturbations in parental monitoring depend upon overall levels of monitoring. Interactions between time and within- and between-person parental monitoring were also considered, but these interactions were dropped from the model because they were

nonsignificant. Random effects were permitted for all within-person effects using an unstructured covariance matrix.

$$alc_inv_{ti} = B_0 + B_1 time_{ti} + B_2 par_mon_{ti} + B_3 par_mon_{ti} + B_4 par_mon_{ti} * par_mon_{ti} \quad (6)$$

$$+ u_{0i} + u_{1i} time_{ti} + u_{2i} par_mon_{ti} + r_{ti}$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} \right)$$

$$r_{ti} \sim N(0, \sigma^2)$$

To account for downward bias in the between-person effect of person-mean parental monitoring due to the relatively small number of within-person observations ($\bar{n} = 4.25$), we implemented the post-hoc adjustment from Equation 5 to arrive at a presumably less biased between-person effect estimate. Even though the n observed across adolescents varied between 1 and 7, it is appropriate to use $\bar{n}$ because the post-hoc formula adjusts a sample-level estimate based on average reliability of person means; we are not applying the formula to individuals in the sample.

Results

The unconditional ICC for the aggregate, time-varying measure of parental monitoring was .50. In the spring of grade 6, the alcohol involvement factor score was estimated to be $-.79$ ($SE=.02$; $p<.001$) with random variation of $\tau_{00}^2 = .25$ ($SE = .02$; $p < .001$). The fixed effect of semester was $B=.17$ ($SE=.17$; $p<.001$), with significant random variation of $\tau_{11}^2 = .01$ ($SE = .00$; $p < .001$), suggesting that alcohol involvement increases with time in school but that there is individual variation in rate of change. The fixed effect of person mean-centered parental monitoring was $B=-.07$ ($SE=.01$; $p<.001$) with significant random variation of $\tau_{22}^2 = .10$ ($SE = .00$; $p < .001$). This means that adolescents have lower alcohol involvement scores during occasions when they perceive higher-than-average levels of parental monitoring. Person-mean parental involvement was also significantly associated with alcohol involvement: $B=-.18$ ($SE=.01$; $p<.001$), indicating that adolescents who perceive consistently higher levels of parental monitoring tend to exhibit less alcohol involvement. The cross-level interaction between person-mean parental monitoring and person-mean-centered parental monitoring was $B=.08$ ($SE=.02$; $p<.001$), indicating that

positive perturbations in parental monitoring are less protective against alcohol involvement for adolescents who report consistently high levels of monitoring. Covariances between the random intercept and the random effect of semester and between the random intercept and the random effect of person mean centered parental monitoring were statistically significant, but the covariance between the random effects of semester and person mean centered parental monitoring was not significant.

The post-hoc adjustment formula from Equation 5 suggests that the disattenuated between-person effect of perceived parental monitoring is $B = -.20$ (compared with $B = -.18$), and the disattenuated cross-level interaction between person-mean parental monitoring and person mean-centered parental monitoring is $B = .09$ (compared with $B = .08$). These estimates represent 11.1% and 12.5% increases in magnitude over the attenuated estimates, respectively.

Simulation Study

Method

To evaluate whether the post-hoc adjustment actually produces less biased estimates, I conducted a small simulation study using parameter estimates drawn from the empirical demonstration to generate 1000 replicated datasets, each of size $N=1000$ and $n=5$. Data were generated and analyzed using R software. The lme4 package was used to analyze each of the 1000 datasets with the multilevel model shown in Equation 6 with REML estimation.

The following population generating model was used:

$$y_{ti} = -.78 - .30\mu_{Xi} - .09(X_{ti} - \mu_{Xi}) + .35time_{ti} + .08\mu_{Xi} * (X_{ti} - \mu_{Xi}) \quad (7)$$

$$+ u_{0i} + u_{1i}time_{ti} + u_{2i}(X_{ti} - \mu_{Xi}) + r_{ti}$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} .300 & & \\ 0 & .006 & \\ 0 & 0 & .097 \end{bmatrix} \right)$$

$$r_{ti} \sim N(0, .34)$$

This model results in an unconditional ICC_x of .47 in the population generating model. This ICC, combined with $n=5$, produces a reliability of approximately .80 (see Figure 1) for $\bar{X}_i$. Bias in between-person effect estimates depends also on the compositional effect, which is moderately small in this simulation.

We calculated $\bar{X}_i$ to approximate μ_{X_i} for each replication and we subtracted $\bar{X}_i$ from the observed X_i to mirror the centering procedure used for disaggregating within- and between-person effects. ICC_x was estimated for each replicated dataset and the post-hoc adjustment was applied to the between-person effect estimate and to the cross-level interaction within each replication. We compared mean parameter estimates obtained with and without the post-hoc correction and calculated the mean percent bias.

Results

The mean estimated between-person effect of $\bar{X}_i$ was $B = -.26$ ($SD = .03$), and the mean estimated cross-level interaction effect was $B = .07$ ($SD = .03$). Thus, the bias in the between-person effect was .04, or 13.3%, and the bias in the cross-level interaction involving $\bar{X}_i$ was .01, or 12.5%. The mean disattenuated between-person effect estimate was $B = -.28$ ($SD = .03$) and the mean disattenuated cross-level interaction estimate was $B = .08$ ($SD = .03$). The bias in these estimates was 6.7% and 0%, respectively. Thus, the post-hoc correction produced less biased parameter estimates than the unadjusted estimates under the realistic conditions simulated in this study.

Discussion

In general, a longitudinal researcher who wishes to distinguish within-person from between-person effects presumably expects a non-null compositional effect (e.g., the effect of stable levels of parental monitoring on adolescent alcohol use is expected to differ from the effect of situational increases on parental monitoring); thus, bias in the between-person effect estimate should be a concern. This is particularly true when the time-varying covariate in question is unstable across repeated measures and when the number of repeated measures is relatively small. All else being equal, a construct that is measured more frequently with shorter intervals between measurements will result in a more reliable and less biased between-person effect estimate. Thus, researchers with panel data are at a much higher risk of making inferential errors about between-person effects than a researcher with intensive longitudinal data measured multiple times per day over weeks.

Simulation study results suggest that the simple post-hoc correction derived in this manuscript produces less biased, disattenuated estimates of between-person effects and cross-level interaction effects. We recommend use of this correction when multilevel models are used to estimate between-person effects based on averages of time-varying measurements, particularly when the stability of the repeated measures is not high (e.g., $ICC_x = .80$), when the average number of repeated measures is not high (e.g., $\bar{n} < 10$), and when a non-negligible compositional effect is anticipated.

Although the post-hoc correction presented here addresses one problem associated with estimating between-person effects from a relatively small number of repeated measures assessments (i.e., downwardly biased estimates), it does nothing to address the problem that these estimates will not be highly reliable; standard errors will still be large for designs that rely on a small n . Regardless of whether this post-hoc correction is applied, analysts should take care to avoid overconfidence in model results when the method for sampling repeated

measures is inadequate for generating representative estimates of person means. Having a very small $\bar{n}$, a strong correlation between n and the value of the person mean, or non-random sampling of repeated measures (e.g., event-contingent sampling designs) are situations in which disaggregation of within-person and between-person effects may be problematic.

Limitations and Future Directions

The simulation study presented in this manuscript relied on parameter estimates obtained from a single empirical example. Although the ICCs and $\bar{n}$ used in this example were in the middle range of values served in a survey of published studies, it is not clear whether the quality of disattenuation is consistent across a wider range of data conditions.

This manuscript did not consider or discuss the role of measurement error in X_{it} . The unreliability discussed in this manuscript results purely from imprecision due to sampling of X_{it} . This issue is discussed at length by Lüdtke et al. (2008) and Lüdtke et al. (2011). In cases in which a researcher has access to multiple indicators of a latent, time-varying construct, it is preferable to use structural equation modeling in order to eliminate consequences of measurement error.

Finally, we assume no residual autocorrelation amongst repeated measures. This assumption is reasonable given the bi-annual survey design used in the empirical example, but it would be a less reasonable assumption for closely spaced repeated measures. Fortunately, these intensive measurement designs typically have a large number of repeated measures, so bias is not a large concern. In the event of a systematic time trend in X , the unconditional ICC_x estimate will represent the average ICC.

Future studies should evaluate confidence interval coverage obtained when confidence intervals are constructed using the standard error estimates associated with the unadjusted point estimates. We hypothesize that these standard errors will be slightly conservative when used with disattenuated point estimates due to increased precision.

Conclusion

Longitudinal study designs permit analysts to disentangle between-person effects from within-person effects if proper centering techniques are employed. Unfortunately, this centering approach tends to produce unreliable person means based on relatively few repeated measures. In turn, estimated between-person effects tend to be underestimated. Because the post-hoc correction presented in this manuscript can be implemented using basic information obtained from default multilevel modeling output, its use can be readily adopted by longitudinal researchers wishing to disattenuate between-person and cross-level interaction effect estimates.

Acknowledgements

Thanks to Dr. Susan Ennett for providing access to the data used in the empirical illustration.

Role of Funding Sources

Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health through grant funding awarded to Dr. Gottfredson (K01 DA035153). The content of this manuscript is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

References

- Bauer DJ (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological methods*, 22, 507–526. [PubMed: 27266798]
- Bauer DJ, & Sterba SK (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological methods*, 16(4), 373. [PubMed: 22040372]
- Biesanz JC, Deeb-Sossa N, Papadakis AA, Bollen KA, & Curran PJ (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological methods*, 9(1), 30–52. doi: 10.1037/1082-989X.9.1.30 [PubMed: 15053718]
- Brown W (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296–322.
- Curran PJ, & Bauer DJ (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology*, 62, 583–619.
- Curran PJ, Lee TH, Howard AL, Lane ST, & MacCallum RC (2012). Disaggregating within-person and between-person effects in multilevel structural equation growth models. In Hancock JHG (Ed.), *Advances in Longitudinal Methods in the Social and Behavioral Sciences* (pp. 217–253). Charlotte, NC: Information Age Publishing.
- Enders CK, & Tofighi D (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological methods*, 12(2), 121. [PubMed: 17563168]
- Gottfredson NC, Cole VT, Giordano M, Bauer DJ, Hussong AM, & Ennett ST (2018). Simplifying the implementation of modern scale scoring methods with an automated R package: Automated moderated nonlinear factor analysis (aMNLFA). under review
- Hawes SW, Crane CA, Henderson CE, Mulvey EP, Schubert CA, & Pardini DA (2015). Codevelopment of psychopathic features and alcohol use during emerging adulthood: Disaggregating between- and within-person change. *Journal of Abnormal Psychology*, 124, 729–739. [PubMed: 26375828]
- Hussong AM, Hicks RE, Levy SA, & Curran PJ (2001). Specifying the relations between affect and heavy alcohol use among young adults. *Journal of Abnormal Psychology*, 110, 449–461. [PubMed: 11502088]
- Jackson C, Henriksen L, & Foshee VA (1998). The Authoritative Parenting Index: predicting health risk behaviors among children and adolescents. *Health Education & Behavior*, 25(3), 319–337. [PubMed: 9615242]
- Kreft IG, De Leeuw J, & Aiken LS (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30(1), 1–21. [PubMed: 26828342]
- Lüdtke O, Marsh HW, Robitzsch A, & Trautwein U (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological methods*, 16(4), 444. [PubMed: 21787083]
- Lüdtke O, Marsh HW, 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. [PubMed: 18778152]
- Raudenbush SW, & Bryk AS (2002). *Hierarchical linear models: Applications and data analysis methods*: Second edition. Thousand Oaks, CA: Sage Publications.
- Singer JD & Willett JB (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* New York: University Oxford Press.
- Spearman C (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 17–101.
- Spearman C (1907). Demonstration of formulae for true measurement correlation. *American Journal of Psychology*, 18, 161–169.

- Spearman C (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.
- Wang LP, & Maxwell SE (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological methods*, 20(1), 63. [PubMed: 25822206]

- When repeated measures are averaged within an individual in order to calculate the stable, between-person effect of a covariate (as distinguished from the temporal, within-person effect of the same covariate), the resulting between person effect estimate can suffer from bias due to unreliability of the person mean due to having a small number of repeated measures with which to calculate the person mean.
- A simple-to-use correction for attenuation is provided. This correction, which is based upon the classic correction for attenuated correlation coefficients, can be applied to results obtained after estimating a multilevel model.
- A small simulation study shows that the correction produces less biased between person effect estimates when compared to unadjusted results.

Figure 1. Reliability of $\bar{X}_i$ (top) and the absolute value of bias of the estimated between effect, $\hat{\beta}_B$, (bottom) at $ICC_X = .2$ (left), $.5$ (middle), and $.8$ (right), as a function of n . Bias also varies as a function of the compositional effect, $\beta_W - \beta_B$. The bottom, solid line represents equivalent β_B and β_W , followed by a standardized difference of 0, .5, 1, 1.5, and 2.