HHS Public Access

Author manuscript

Multivariate Behav Res. Author manuscript; available in PMC 2020 May 01.

Published in final edited form as:

Multivariate Behav Res. 2019 ; 54(3): 382–403. doi:10.1080/00273171.2018.1532280.

On standardizing within-person effects: Potential problems of global standardization

Lijuan Wang,
University of Notre Dame

Qian Zhang,
Florida State University

Scott E. Maxwell,
University of Notre Dame

C. S. Bergeman
University of Notre Dame

Abstract

Person-mean centering has been recommended for disaggregating between-person and within-person effects when modeling time-varying predictors. Multilevel modeling textbooks recommended global standardization for standardizing fixed effects. An aim of this study is to evaluate whether and when person-mean centering followed by global standardization can accurately estimate fixed-effects within-person relations (the estimand of interest in this study) in multilevel modeling. We analytically derived that global standardization generally yields inconsistent (asymptotically biased) estimates for the estimand when between-person differences in within-person standard deviations exist and the average within-person relation is nonzero. Alternatively, a person-mean-SD standardization (P-S) approach yields consistent estimates. Our simulation results further revealed (1) how misleading the results from global standardization were under various circumstances and (2) the P-S approach had accurate estimates and satisfactory coverage rates of fixed-effects within-person relations when the number of occasions is 30 or more (in many conditions, performance was satisfactory with 10 or 20 occasions). A daily diary data example, focused on emotional complexity, was used to empirically illustrate the approaches. Researchers should choose standardization approaches based on theoretical considerations and should clearly describe the purpose and procedure of standardization in research articles.

Methodologists have emphasized the conceptual differences between between-person (BP) and within-person (WP) effects (e.g., Hamaker, Dolan, & Molenaar, 2005; Hamaker, 2012; Molenaar, 2004; Molenaar & Campbell, 2009) and the need to distinguish them and model both (e.g., Curran & Bauer, 2011; Hamaker, Kuiper, & Grasman, 2015; Wang & Maxwell, 2015). Using the effects of stress on positive affect as an example, a between-person effect refers to the extent to which people who are one unit above average on stress are above or below average on positive affect. In contrast, a within-person effect reflects the extent to

Correspondence should be addressed to Lijuan Wang, 390 Corbett Hall, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556. lwang4@nd.edu.

which an individual has a higher or lower score on positive affect when he or she has a one unit higher score on stress. That is, between-person questions concern who, and within-person questions concern when. The two types of research questions are distinct, and the answer to one can not generally be inferred from the other (e.g., Molenaar, 2004).

To statistically disaggregate within- and between-person effects of one variable on another, both variables need to be time-varying. Thus, longitudinal data on both variables are required. For modeling *within-person* effects, multilevel modeling (MLM) is often used, and person-mean centering (P-C) has been recommended in the literature (e.g., Bolger & Laurenceau, 2013; Curran & Bauer, 2011; Raudenbush & Bryk, 2002; Wang & Maxwell, 2015). Relatedly, cluster-mean centering has been recommended for disaggregating within- and between-cluster effects in the cross-sectional multilevel modeling literature (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). When raw data (without centering) are used, researchers have repeatedly shown that the resulting total effect is conflated or confounded and thus can be meaningless in both longitudinal and cross-sectional multilevel modeling (e.g., Curran & Bauer, 2011; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Therefore, raw data generally should not be used for estimating *within-person* effects. "Generally" is emphasized in the previous sentence, because when the target of inference is not on within-person effects, other centering methods can be preferred, as discussed in Enders and Tofighi (2007), Hamaker and Grasman (2015), and Kreft, de Leeuw, and Aiken (1995). Therefore, the choice of a centering approach should be based on theoretical considerations.

After applying person-mean centering, raw fixed-effects within-person coefficients, for example, can measure the number of points that the "average" individual's mood score will change, when he or she has a one unit higher score on stress. When the raw scales or units are meaningful (e.g., pounds for weight; miles for distance), raw coefficients can be informative and easy to interpret. However, raw coefficients can have limitations in some contexts. First, the interpretation of a raw within-person coefficient depends on the scales of two time-varying variables. In behavioral research, measurement scales sometimes are arbitrary and lack practical meaningfulness. Thus, interpretations of a raw within-person coefficient can be arbitrary and not very meaningful. For example, a raw coefficient of −.65 for the within-person effect of stress on positive affect is almost impossible to interpret meaningfully without reference to the items and/or scales used to measure stress and positive affect. Even after studying the items, it may remain challenging to effectively interpret the raw coefficient. Second, raw coefficients are generally not good effect size measures because they are not scale/unit invariant (Kelley & Preacher, 2012). Third, a raw coefficient usually is less useful for comparing the strengths of within-person relations across different pairs of time-varying variables (e.g., stress and positive affect vs. sleeping quality and positive affect) because different scales may be used for different variables.

To address some of the limitations, standardization can be useful.[1] When the target of estimation (i.e., *estimand*) concerns *within-person relations*, researchers have recently argued for within-person standardization. For example, Zhang and Wang (2014) discussed

[1]Standardization also has limitations, which will be discussed later in the discussion section.

the use of person-mean-SD standardized (P-S) data for aggregating intraindividual correlations across individuals with multilevel modeling. More recently, Schuurman, Ferrer, de Boer-Sonnenschein, and Hamaker (2016) have argued for using within-person standardization to compare within-person cross-lagged relations in multilevel autoregressive modeling. They explained that "we are interested in Granger-causal psychological processes, which happen within persons, at the level of the individual. It does not seem reasonable to conflate this WP variation with variation between persons, given that the person-specific Granger-causal processes are not concerned with differences in the means of these processes between individuals." (page 213). However, researchers may still rely on popular multilevel modeling textbooks that recommend global standardization, where a standardized fixed-effects coefficient is calculated by multiplying the raw coefficient with the ratio of the grand standard deviation (SD) of the predictor to the grand SD of the outcome (e.g., Hox, 2010; Snijders & Bosker, 2012). The procedure is essentially the same as the one used in regressions.

To shed light on the standardization practices used in multilevel modeling with time-varying predictors, a small literature review was conducted. Twenty four recent (published in 2011–2018) empirical papers were included (more details are described in Part A of the online supplemental materials[2]). Although global standardization was recommended originally in the context of cross-sectional multilevel modeling, researchers are using it to study within-person relations in longitudinal research (e.g., Aafjes-van Doorn et al., 2017; Armeli et al., 2014; Foshee et al., 2013). For example, Armeli et al. (2014) stated that "To aid in the evaluation of the strength of the effects, we calculated standardized coefficients as per Hox (2010)." (page 769). The review also revealed that the rationale of standardization and the standardization procedures were often not clearly described in the empirical papers (more than 41% of the studies; e.g., Freeman & Gottfredson, 2017). Specifically, it was not clear to readers why the researchers conducted standardization, whether the standardization was at the person level or at the global level, and whether the outcomes were standardized. Due to the lack of descriptions on standardization approaches yet detailed descriptions of the person-mean centering procedures, we suspected that many of these studies likely used global standardization. Moreover, the review showed that researchers have begun to use WP standardization to study within-person relations (about 17% of the studies). Specifically, Ramseyer et al. (2014) conducted idiographic dynamic modeling using WP standardized variables and obtained the average standardized coefficients by averaging the individual standardized coefficients across individuals. More recently, researchers implemented WP standardization on the variables of interest and then multilevel modeling was conducted on the WP standardized variables for studying within-person relations (e.g., Dejonckheere et al., 2017, 2018; Lydon-Staley et al., 2018). Their implemented WP standardization approach was similar to the P-S approach evaluated in Zhang and Wang (2014).

Given the use of different methods for standardizing within-person effects and the lack of clarity in the rationale and procedure of standardization in empirical longitudinal research, we conducted the current study to achieve two aims. The first aim is to further evaluate

[2]The online supplemental materials can be downloaded from https://ldhrm.nd.edu/assets/289171/supplemental_materials_2018mbr.pdf.

whether global standardization is problematic, when it is problematic, and how much of a problem it would be for estimating *within-person relations (the estimand of interest of the study)*. The second aim is to evaluate how the within-person standardization of data followed by multilevel modeling approach (in the remainder of the paper, we call it the P-S approach) performs. We decided to evaluate P-S mainly because it has recently begun to appear in empirical research (e.g., Dejonckheere et al., 2017, 2018; Lydon-Staley et al., 2018) and is easy to implement by researchers with basic knowledge of multilevel modeling. To achieve the aims, we conducted analytical derivations, Monte Carlo simulations, and a real daily diary data analysis. Our analytical derivations revealed the asymptotic estimation (i.e., consistency) performance of various standardization approaches for *estimating within-person relations* and uncovered the core factors that influence the performance. The consistency evaluation answered an unresolved question raised in Schuurman et al. (2016) regarding how different the asymptotic results (i.e., results obtained when the sample size and the number of time points approach infinity) are using global standardization vs. within-person standardization. The simulations facilitated the evaluation of inferential properties with finite samples. For example, we were interested in learning how many assessments per person are needed for making accurate inferences about *within-person relations* using the P-S approach. Furthermore, the real data example illustrated how much the results from different methods might differ in empirical applications.

The remainder of the article is organized as follows. First, we describe the population model, the estimand, various ways of obtaining standardized coefficients using multilevel modeling, and the modeling and estimation assumptions. Then, we derive the population correlations between two time-varying variables for different types of data including raw data, person-mean centered data, and P-S data under or relaxing the normality distribution assumption. With the derived population correlations, we show when global standardization can yield asymptotically unbiased estimates of fixed-effects within-person relations, under either homogeneous or heterogeneous within-person relation conditions. Next, we use results from a simulation study to evaluate the inferential properties of both global standardization and the P-S approach with finite samples. A real data analysis example is provided to demonstrate the findings and substantively study the emotional complexity issue, illustrating how conclusions regarding emotional complexity might differ depending on the method used to analyze the data. We conclude the article with recommendations, implications, and future research directions.

Multilevel models and standardization methods for modeling within-person relations

We begin this section by describing the population model and the estimand of interest in the current study when there is one time-varying predictor. The case of having more than one time-varying predictor will be considered later in the paper. Let the population/true *within-person* or *intraindividual* correlation between time-varying predictor X and time-varying outcome Y be $\rho_{w,i}$ for individual i. We have $\rho_{w,i} = \frac{E\left[\left(X_{it} - \mu_{Xi}\right)\left(Y_{it} - \mu_{Yi}\right)|i\right]}{\sigma_{Xi}\sigma_{Yi}}$, where μ_{Xi} and μ_{Yi} stand for the population within-person means in X and Y of individual i, σ_{Xi} and σ_{Yi} are the

individual *i*'s population within-person standard deviations in X and Y, and the expectation is taken over time/waves for individual *i*. Throughout the paper, we assume that individual *i*'s dynamic process is stationary over time. Note that in the derivations, we do not assume stationarity. This assumption is just for facilitating interpretations. That is, $\rho_{w,i}$ represents the dynamic relationship between two time-varying variables for individual *i*. $\rho_{w,i}$ can be used to compare the strengths of within-person linear relations between X and Y across individuals or across different sets of variables. Let $\mu_{\rho w}$ be the true *within-person* correlation for the "average" person or the average within-person correlation between X and Y in the population. In this article, $\mu_{\rho w}$ *is the focal parameter of interest (the estimand) when there is one time-varying predictor*. $\mu_{\rho w}$ is useful for (1) quantifying the average within-person correlation between two time-varying variables and (2) comparing the relative strengths of linear associations across different sets of variables at the population level.

In practice, $\rho_{w,i}$ is unknown. A consistent estimate of $\rho_{w,i}$ is $r_{w,i} = \frac{\sum_t (x_{it} - x_{i.})(y_{it} - y_{i.})}{T_i s_{xi} s_{yi}}$, where y_{it} and x_{it} are the observed scores of variables Y and X of individual i ($i = 1, 2, \ldots N$) at time t ($t = 1, 2, \ldots, T_i$). $x_{i.} = \sum_t x_{it}/T_i$ is the observed WP (within-person) mean score of X for individual i averaged across the T_i time points. $y_{i.}$ is the counterpart for Y. s_{xi} and s_{yi} are the individual *i*'s sample WP standard deviation in X and Y respectively. For example, we have $s_{xi} = \sqrt{\sum_t (x_{it} - x_{i.})^2/(T_i - 1)}$. The relations among $r_{w,i}$, $\rho_{w,i}$, and $\mu_{\rho w}$ can be expressed via the following multilevel model

$$r_{w,i} = \rho_{w,i} + e_{it} \qquad (1)$$
$$\rho_{w,i} = \mu_{\rho w} + u_{1i}.$$

The estimated fixed-effects within-person relation between X and Y, $\hat{\mu}_{\rho w}$, is a precision weighted average of the individual sample within-person correlations $r_{w,i}$ (e.g., Bryk & Raudenbush, 1987; Maxwell & Delaney, 2004). The weight is a function of the number of time points an individual has, with more time points corresponding to greater weights. Multilevel models are often estimated by maximum likelihood (ML) or generalized least square (GLS) methods (e.g., Raudenbush & Bryk, 2002). When Level-1 residuals are normally distributed, the ML and GLS estimators for fixed effects have the same form. Under regular assumptions in multilevel modeling (e.g., normality for the Level-1 and Level-2 residuals), $\hat{\mu}_{\rho w}$ from ML or GLS is a consistent estimator of $\mu_{\rho w}$. When the normality assumption is relaxed, GLS estimates are still consistent.

In the following, we describe several multilevel modeling approaches coupled with different standardization methods for estimating the average within-person relation $\mu_{\rho w}$. Note that the following methods are used in recent empirical research based on our literature review summarized earlier in the introduction.

Person-mean centering followed by global standardization

A frequently discussed multilevel model for modeling a time-varying predictor (e.g., Curran & Bauer, 2011; Wang & Maxwell, 2015) has the following form.

$$\begin{aligned} y_{it} &= \gamma_{0i}^{C1} + \gamma_{1i}^{C1}(x_{it} - x_{i.}) + e_{it}^{C1} \quad (2)\\ \gamma_{0i}^{C1} &= \gamma_{00}^{C1} + \gamma_{01}^{C1} x_{i.} + u_{0i}^{C1},\\ \gamma_{1i}^{C1} &= \gamma_{10}^{C1} + u_{1i}^{C1} \end{aligned}$$

where superscript $C1$ denotes that the coefficients are from the P-C (person-mean centered) model in Eq (2). $x_{it}^{PC} = x_{it} - x_{i.}$ is the person-mean centered predictor score of individual i at time t. γ_{00}^{C1} is the model implied level of Y when both $x_{i.}$ and $x_{it} - x_{i.}$ are 0, and thus sometimes may lack practical meaningfulness. γ_{01}^{C1} represents the between-person effect of X on Y. When u_{1i}^{C1} is included in the model, individuals are allowed to differ in the within-person effects. In this case, γ_{1i}^{C1} describes individual i's within-person effect of X on Y and γ_{10}^{C1} describes the average within-person effect or the within-person effect for an "average" person in the population. When u_{1i}^{C1} is not included in the model, we have $\gamma_{1i}^{C1} = \gamma_{10}^{C1}$, meaning that different individuals have the same within-person effects. The simplified model (i.e., u_{1i}^{C1} is not included) has been applied about as frequently as the full model (i.e., u_{1i}^{C1} is included) in psychological research, 17 vs. 12 times, as reported in a review conducted by Baird (2016). However, the simplified model yielded inflated Type I error rates for testing γ_{10}^{C1} when the individual within-person effects are heterogeneous in reality, whereas the full model had higher nonconvergence rates when the individual within-person effects are homogeneous in reality (e.g., Baird & Maxwell, 2016). In this study, we will study both the homogeneous and heterogeneous within-person effect scenarios.

Although it may appear that person-mean centering is conducted for the predictor variable only, person-mean centering is also conducted for the outcome variable because of the inclusion of the random intercept term γ_{0i}^{C1}. Specifically, when the first equation in Eq (2) is rewritten as

$$y_{it} - \gamma_{0i}^{C1} = \gamma_{1i}^{C1}(x_{it} - x_{i.}) + e_{it}^{C1}, \quad (3)$$

it is clearer to see that person-mean centering is conducted on both variables. Note that in the model of Eq (2), observed person means are used for centering the predictor X whereas latent/true means are used for centering the outcome Y. When observed person means are used for centering both X and Y, the P-C model becomes

$$y_{it}^{PC} = \gamma_{1i}^{C2} x_{it}^{PC} + e_{it}^{C2} \quad (4)$$
$$\gamma_{1i}^{C2} = \gamma_{10}^{C2} + u_{1i}^{C2},$$

where $y_{it}^{PC} = y_{it} - y_{i.}$, $x_{it}^{PC} = x_{it} - x_{i.}$, and superscript $C2$ denotes that the coefficients are from the alternative P-C model (the one in Eq 4). For each individual, y_{it}^{PC} and x_{it}^{PC} have marginal means of 0. The P-C model in Eq (4) focuses on studying within-person effects or relations only.

Empirically, the estimated reliabilities of observed WP means (estimated squared correlations between observed and true WP means) are often higher than .92 and .96, even when the number of time points is as low as 5 or 10, respectively (Wang & Grimm, 2012). When T goes to infinity, the reliabilities at the population level (N is infinity) approach 1 (Estabrook, Grimm, & Bowles, 2012; Schmiedek, Lövdén, & Lindenberger, 2009; Wang & Grimm, 2012). When T is infinity, γ_{10}^{C1} and γ_{10}^{C2} (population parameters; i.e., N is infinity) are mathematically equivalent because (1) $x_{i.}$ and x_{it}^{PC} are uncorrelated and (2) we can equate the observed person means ($x_{i.}$) to their true values. Thus, *asymptotically* (both T and N are infinity), $\hat{\gamma}_{10}^{C1}$ and $\hat{\gamma}_{10}^{C2}$ from the same estimator (e.g., ML or GLS) should converge to the same value. Because the estimation in Eq (4) is simpler, we conducted the derivation about $\hat{\gamma}_{10}^{C2}$ under Eq (4) but the asymptotic results apply to $\hat{\gamma}_{10}^{C1}$ in Eq (2) as well. Later in the paper, we also evaluated the performance of both $\hat{\gamma}_{10}^{C1}$ and $\hat{\gamma}_{10}^{C2}$ with finite samples separately via simulations.

The GLS estimator of γ_{10}^{C2} in Eq (4) is given in Eq (8) of the online supplemental materials (see Appendix C of Part B). Under the homogeneous condition (i.e., u_{1i}^{C2} is not included in the model), the GLS estimator becomes the ordinary least square (OLS) estimator. The OLS estimator of γ_{10}^{C2} has a very simple form, which is $r_{cx,cy}\frac{s_{cy}}{s_{cx}}$, where s_{cy} is the sample grand standard deviation (SD) of the person-mean centered outcome variable Y and s_{cx} is the sample grand SD of the person-mean centered predictor variable X. Mathematically, $s_{cy} = \sqrt{\sum_i \sum_t \left(y_{it}^{PC}\right)^2 / \left(\sum_i T_i - 1\right)}$ and $s_{cx} = \sqrt{\sum_i \sum_t \left(x_{it}^{PC}\right)^2 / \left(\sum_i T_i - 1\right)}$. $r_{cx,cy}$ is the sample grand correlation between the person-mean centered variables. Note that s_{cy}, s_{cx}, and $r_{cx,cy}$ are obtained using the stacked long person-mean centered data from all the individuals and time points (y_{it}^{PC} and x_{it}^{PC}).

In multilevel modeling textbooks, a recommended standardization approach for γ_{10}^{C1} in Eq (2) or γ_{10}^{C2} in Eq (4) is $\hat{\gamma}_{10}^{C*} = \hat{\gamma}_{10}^{C}\frac{s_{predictor}}{s_{outcome}}$ (e.g., Hox, 2010; Snijders & Bosker, 2012), where $s_{predictor}$ and $s_{outcome}$ are the sample grand SDs of the predictor and outcome variables

calculated using the stacked long data, respectively. We refer to the standardization approach as global standardization. When the P-C model in Eq (2) is fitted, there are two potential global standardization options. One is to use the sample grand SD of the person-mean centered outcome variable and the other is to use the sample grand SD of the original outcome variable for the calculations, respectively. In some of the reviewed empirical articles (more than 41% of the studies), it was not clear which one was used. Thus, we evaluated both in this study. A globally standardized estimator of γ_{10}^{C1} in Eq (2) is

$$\hat{\gamma}_{10}^{G1*} = \hat{\gamma}_{10}^{C1}\frac{s_{cx}}{s_{cy}}, \quad (5)$$

where s_{cy} is the sample grand SD of the person-mean centered outcome variable. We call the standardization method in Eq (5) the M_{G1} method. And an alternative globally standardized estimator of γ_{10}^{C1} in Eq 2 is

$$\hat{\gamma}_{10}^{G2*} = \hat{\gamma}_{10}^{C1}\frac{s_{cx}}{s_{y}} \quad (6)$$

where s_y is the sample grand SD of the original outcome variable. We call the standardization method in Eq (6) the M_{G2} method.

When the model in Eq 4 is fitted, with global standardization, we have

$$\hat{\gamma}_{10}^{G3*} = \hat{\gamma}_{10}^{C2}\frac{s_{cx}}{s_{cy}}. \quad (7)$$

Eq (7) can be simplified under the homogeneous within-person effect assumption:

$$\hat{\gamma}_{10,OLS}^{G3*} = \hat{\gamma}_{10,OLS}^{C2}\frac{s_{cx}}{s_{cy}} = r_{cx,cy}. \quad (8)$$

We call the standardization method in Eqs (7) and (8) the M_{G3} method.

Note that in global standardization, regular MLM assumptions apply to the unstandardized coefficients. For example, when normal-theory-based ML is used for inference, the unstandardized random coefficients (e.g., γ_{1i}^{C}) are assumed to have a normal distribution under the heterogeneous within-person effect condition. When GLS is used for inference, however, such assumptions are not made.

Global standardization can help answer the following kind of research question: The number of grand SDs that the “average” person’s daily mood will change when his or her daily stress

increases 1 grand SD (also see Table 1). When this research question is of interest, global standardization can be used. However, we doubt that researchers are often interested in the research question. In M_{G1} and M_{G3}, the grand SD of a person-mean centered predictor or outcome variable is a linear combination of both the average of within-person SDs and variance in within-person SDs (see Table 2). For M_{G2}, it is more complex in that the grand SD of the raw outcome variable is a linear combination of the average of within-person SDs, variance in within-person SDs, and variance in within-person means (see Table 2). The derivations are shown in Part B of the online supplemental materials. In other words, between-person variances are involved in global standardization. We agree with the statement in Schuurman et al. (2016) that the person-specific Granger-causal processes are not concerned with between-person differences in the processes. Thus, the kind of research question addressed by global standardization may be less interesting to researchers. When the estimand is on within-person relations or relations at the individual level (e.g., $\mu_{\rho w}$, the average within-person correlation between X and Y), we expect that the three global standardization approaches yield inconsistent estimates of $\mu_{\rho w}$ under some circumstances. We will evaluate whether and when our conjecture is correct later in the derivation section.

Within-person standardization

An alternative multilevel modeling approach is to use the person-mean-SD standardized (P-S) data for both the time-varying predictor and outcome variables, rather than person-mean centered data. This approach has begun to be used in recent empirical research for standardizing within-person effects (e.g., Dejonckheere et al., 2017, 2018; Lydon-Staley et al., 2018). Methodologically, Zhang and Wang (2014) discussed the use of this P-S approach and compared it to two meta-analysis approaches in which within-person correlations are directly used for aggregating intraindividual correlations across individuals. They found that the multilevel modeling approach worked well under certain conditions, whereas the meta-analysis approaches yielded slightly biased estimates of the population average within-person correlation under small T conditions. In addition, it is easier to model multiple time-varying predictors using the multilevel modeling approach. Thus, we focus on the multilevel modeling approach rather than directly using within-person correlation coefficients in this study. A P-S multilevel model has the following form

$$\begin{aligned} y_{it}^{PS} &= \gamma_{1i}^{PS} x_{it}^{PS} + e_{it}^{PS} \\ \gamma_{1i}^{PS} &= \gamma_{10}^{PS}\left(+u_{1i}^{PS}\right), \end{aligned} \quad (9)$$

where $y_{it}^{PS} = \left(y_{it} - y_{i\cdot}\right)/s_{yi}$ with $s_{yi} = \sqrt{\sum_t \left(y_{it} - y_{i\cdot}\right)^2/\left(T_i - 1\right)}$ and $x_{it}^{PS} = \left(x_{it} - x_{i\cdot}\right)/s_{xi}$ with $s_{xi} = \sqrt{\sum_t \left(x_{it} - x_{i\cdot}\right)^2/\left(T_i - 1\right)}$. y_{it}^{PS} and x_{it}^{PS} are the within-person standardized data (or within-person z scores) for the outcome and predictor, respectively.[3]

[3]P-S data can not be obtained when the within-person sample SD is zero. In this case, we recommend that one can just person-mean center the data of the zero-WP-SD variable for the individual. This is because (1) the within-person sample covariance is 0 between this zero-WP-SD variable and another time-varying variable for the individual and (2) 0 times any weight is still 0.

For each individual, y_{it}^{PS} and x_{it}^{PS} have marginal means of 0 and unit marginal variances. Within-person z scores have been used in personality research of within-person correlational designs (Michela, 1990). Superscript PS is used for denoting that the coefficients are from the P-S approach to distinguish them from the P-C coefficients presented earlier in the paper. Conceptually, γ_{1i}^{PS} quantifies the within-person correlation of the two time-varying variables for individual i and thus γ_{10}^{PS} measures the within-person correlation for the "average" person. The P-S approach can help answer the following kind of research question: The number of within-person SD units that the "average" person's daily mood will change when his or her daily stress increases 1 within-person SD unit (also see Table 1). Thus, the research question is a question at the individual level. Note that in Eq (9), regular MLM assumptions apply to the standardized coefficients $\left(\gamma_{1i}^{PS}\right)$.

Readers may have noticed that a random intercept term is not included in the P-S model. This is because with P-S data, the intercept is 0 for each individual. The (standardized) GLS estimator of γ_{10}^{PS} is given in Eq (11) of the online supplemental materials (see Appendix C of Part B). The (standardized) OLS estimate (also the GLS estimate for the model with $u_{1i}^{PS} = 0$ for all i) of γ_{10}^{PS} is $r_{wx,wy}$, where $r_{wx,wy}$ is the sample grand correlation between the WP standardized variables, calculated using the stacked long data (y_{it}^{PS} and x_{it}^{PS}). We call the standardization method in Eq (9) the M_{PS} (P-S) approach. The simulations in Zhang and Wang (2014) have shown that the P-S approach can be used to describe the average intraindividual correlation under certain conditions. However, they did not (1) compare P-S to global standardization; (2) consider between-person differences in within-person standard deviations (a core factor discovered from our derivations to be described later); or (3) consider two or more time-varying predictors. In addition, they did not conduct analytical derivations to evaluate the consistency of global standardization or P-S for estimating within-person relations.

As discussed earlier, Schuurman et al. (2016) argued for conducting standardization of random effects at the person level to facilitate a more valid comparison of the cross-lagged relations in multilevel autoregressive modeling. They did not recommend global standardization in that context. Their approach is similar to the P-S approach in that both are within-person standardization approaches, but there are some differences. For example, Schuurman et al. (2016) first person-mean center the data and then use latent person SDs to standardize the random effects at the person level, whereas in the P-S approach (Eq 9), the data are first WP standardized at the person level and then a multilevel model is fit to the standardized data. Furthermore, in Schuurman et al. (2016), sophisticated Bayesian modeling was used, in which latent variances are modeled assuming stationarity in the time series data.

In the current study, we further evaluated the performance of the P-S approach as a within-person standardization approach for estimating and inferring within-person relations under a broader variety of conditions than Zhang and Wang (2014), using both derivations and

simulations. We focused on P-S because it has recently began to appear in the empirical literature and can be easily fitted by substantive researchers with basic knowledge of multilevel modeling using standard statistical software such as SPSS and SAS. In addition, as suggested by a reviewer, we also evaluated another frequentist within-person standardization procedure. In this method, individual empirical Bayes estimates of γ_{1i}^{C1} in Eq (2) are first obtained (e.g., Fitzmaurice et al., 2011) and then standardized at the person level using individual i's observed within-person SD.[4] The standardized estimates were averaged across individuals for estimating the average within-person relation. We call this approach the EB standardization approach (the M_{EB} method). The empirical Bayes estimates are the "best linear unbiased predictor" (or BLUP; e.g., Fitzmaurice, Laird, & Ware, 2011). For individual i, after standardizing $\hat{\gamma}_{1i}^{C1}$ using within-person SDs, asymptotically, the EB standardized coefficient is individual i's within-person correlation.

We conducted analytical derivations to evaluate the asymptotic performance of various standardization approaches for estimating the average within-person relation $\mu_{\rho w}$ when there is one time-varying predictor. That is, we evaluated whether $\hat{\gamma}_{10}^{G3*}$ and $\hat{\gamma}_{10}^{G2*}$ from global standardization, and $\hat{\gamma}_{10}^{PS}$ from P-S are *consistent* estimates of $\mu_{\rho w}$ when the within-person relations are heterogeneous or homogeneous. As discussed above, the consistency results about $\hat{\gamma}_{10}^{G3*}$ apply to $\hat{\gamma}_{10}^{G1*}$. Similarly, the consistency results about $\hat{\gamma}_{10}^{PS}$ also apply to those from the EB standardization approach.

Two or more time-varying predictors

The consistency findings from multilevel modeling with one time-varying predictor can be directly extended to multilevel modeling with two or more time-varying predictors.[5] If a bivariate relation $\mu_{\rho w}$ can not be consistently recovered using a standardization approach, the partial correlation and standardized coefficient estimates after controlling for the other time-varying predictors from the standardization approach would generally be inconsistent as well because the partial correlations and standardized coefficient estimates are functions of the relevant bivariate correlations. Therefore, the standardization approach may produce inconsistent estimates for within-person relations after controlling for the other time-varying predictors.

To evaluate the inferential properties of the standardization approaches, we conducted simulations when there are one or two time-varying predictors under both the heterogeneous and homogeneous within-person relation scenarios with finite N and T_i. For the real data analysis illustrations, models with one or two time-varying predictors were fitted.

[4]Observed within-person SD, rather than latent WP SD, are used. So this approach is different from the approach in Schuurman et al. (2016).

[5]Due to space limitations, we do not list the model forms with two or more time-varying predictors.

Derivations

When within-person effects are homogeneous in reality and thus there is no need to include random effects for the γ_{1i}s, the standardized GLS estimates of γ_{10} in Eqs (4) and (9) are the sample correlations from the corresponding stacked long data (i.e., $r_{CX,CY}$ and $r_{WX,WY}$), as discussed in the previous section. Those sample correlations asymptotically approach their population correlations correspondingly. Thus, to examine estimation consistency, in this section, we first derive the population correlations for different kinds of stacked long data including raw data, person-mean centered data, and WP standardized data. Our derivations revealed that some of the correlations involve the between-person variances in within-person means and SDs. Below, we define the relevant concepts and terms, and describe the assumptions used for deriving the conclusions. Then, we summarize the correlation derivation results. After that, we describe the asymptotic performance of the different standardization approaches for estimating $\mu_{\rho w}$. The detailed derivation process and results are included in the online supplemental materials (see Part B).

Between-person differences in person means and variabilities

Let μ_{Xi} and μ_{Yi} represent the population intraindividual/person means in X and Y of individual i, where $\mu_{Xi} = E(X_{it}|i)$ and $\mu_{Yi} = E(Y_{it}|i)$. The between-person population covariance matrix of μ_{Xi} and μ_{Yi} is

$$\begin{pmatrix} \sigma^2_{\mu X} & \sigma_{\mu X,\mu Y} \\ \sigma_{\mu X,\mu Y} & \sigma^2_{\mu Y} \end{pmatrix}, \quad (10)$$

where $\sigma^2_{\mu X}$ and $\sigma^2_{\mu Y}$ are the population variances of the person means for X and Y, quantifying between-person differences in the person means of X and Y, respectively. The correlation between the person means (μ_{Xi} and μ_{Yi}), $\rho_b = \frac{\sigma_{\mu X,\mu Y}}{\sigma_{\mu X}\sigma_{\mu Y}}$, quantifies the population *between-person* correlation between X and Y. For example, a positive ρ_b shows that a person with a higher average score in X tends to have a higher average score in Y, compared to another person with a lower average score in X.

A widely used indicator of intraindividual variability is the intraindividual or person standard deviation (e.g., Gerstorf et al., 2009; Nesselroade & Salthouse, 2004). It may be worth noting that the intraindividual SD here is different from the innovation SD of time series analysis. The former is the overall intraindividual SD of a person, whereas the latter is the residual SD after controlling for a dynamic process (e.g., a first-order autoregressive process; Jongerling, Laurenceau, & Hamaker, 2015). Let σ_{Xi} and σ_{Yi} be individual i's population intraindividual standard deviations in X and Y, where $\sigma_{Xi} = \sqrt{E\left[(X_{it} - \mu_{Xi})^2|i\right]}$ and $\sigma_{Yi} = \sqrt{E\left[(Y_{it} - \mu_{Yi})^2|i\right]}$. The population means of σ_X and σ_Y are $E(\sigma_X) = \mu_{\sigma X}$ and

$E(\sigma_Y) = \mu_{\sigma Y}$, representing the average of the WP standard deviations averaged across individuals, respectively. The covariance matrix of the person SD variables (σ_X and σ_Y) is

$$\begin{pmatrix} \sigma^2_{\sigma X} & \sigma_{\sigma X, \sigma Y} \\ \sigma_{\sigma X, \sigma Y} & \sigma^2_{\sigma Y} \end{pmatrix}, \quad (11)$$

where $\sigma^2_{\sigma X}$ and $\sigma^2_{\sigma Y}$ are the population between-person variances of σ_X and σ_Y, quantifying between-person differences in the within-person SDs of X and Y, respectively.

In the past few decades, researchers have found that substantial between-person differences in within-person standard deviations existed in psychological and behavioral variables such as mood, motor performance, and stress (e.g., Ferrer, Gonzales, & Steele, 2013; Hedeker, Mermelstein, Berbaum, & Campbell, 2009; Nesselroade & Salthouse, 2004; Wang, Hamaker, & Bergeman, 2012). Between-person differences in intraindividual variability have been found to be predictive of important outcomes (e.g., Eid & Diener, 1999; Fiske & Rice, 1955; Hedeker et al., 2009; Nesselroade, 1991; Nesselroade & Molenaar, 2010; Ram & Gerstorf, 2009). Our derivations discovered *how* between-person differences in both within-person means and/or variabilities play a role in the population correlations of the different kinds of stacked long data and thus the standardized fixed-effects coefficients from various standardization approaches.

The population correlation derivation results

We derived correlations under two different assumptions. First, we assumed that σ_X, σ_Y, X^{PS}, and Y^{PS} follow a joint multivariate normal distribution. Second, we relaxed this assumption. Under the normality assumption, the derived population correlations of stacked long raw data, P-C data, and P-S data all have simple forms (see Table 2). When the normality assumption is relaxed, the population correlation of stacked long P-S data is still the population average WP correlation; whereas more elements are involved in the derived population correlations for stacked long raw data and P-C data (see Eqs 3 and 5 in Part B of the online supplemental materials respectively).

Overall, we have the following findings from the correlation derivations. First, $\rho_{X,Y}$, the population correlation between stacked long raw data of variables X and Y, is a weighted average of the BP correlation ρ_b and the average WP correlation $\mu_{\rho w}$. Thus, $\rho_{X,Y}$ reflects neither the average WP correlation nor the BP correlation; instead, it measures a conflated and often meaningless relation.

Second, person-mean centering successfully removes the BP correlation ρ_b from the population correlation between stacked long P-C data of X^{PC} and Y^{PC} ($\rho_{CX,CY}$). However, $\rho_{CX,CY}$, is still generally not equal to the population average WP correlation $\mu_{\rho w}$ when $\mu_{\rho w} \neq 0$. When $\mu_{\rho w} = 0$ and/or there are no between-person differences in the within-person SDs ($\sigma_{\sigma X} = \sigma_{\sigma Y} = 0$), the population correlation of P-C data is the population average WP correlation ($\rho_{CX,CY} = \mu_{\rho w}$).

Third, the population correlation between WP standardized data of variables X^{PS} and Y^{PS}, $\rho_{WX,WY}$, is equal to $\mu_{\rho w}$, regardless of whether the data are normally distributed or not and whether the within-person correlations are homogeneous or heterogeneous.

In the correlation derivations, we did not assume the covariances between a person mean variable and a person SD variable to be zero and those covariances do not appear explicitly in the population correlation formulas.

Asymptotic performance of global standardization and P-S for estimating the average WP relation

With the correlation derivation results, we analytically evaluated the asymptotic performance of global standardization and the P-S approach for estimating the average within-person correlation ($\mu_{\rho w}$), under homogeneous or heterogeneous within-person relation conditions. We utilized some ideal-case scenarios to help understand how between-person differences in within-person SDs play a role in the asymptotic performance of global standardization.

Our derivation results revealed that regardless of whether within-person relations are homogeneous or heterogeneous, global standardization (M_{G1} and M_{G3}) generally yields inconsistent estimates of the average within-person correlation ($\mu_{\rho w}$) when (1) $\mu_{\rho w} \neq 0$ and (2) there are between-person differences in the WP standard deviations of one or both of the time-varying variables. For M_{G2}, even under the ideal case that there are no BP differences in the WP standard deviations of either variables, the standardized estimates are generally inconsistent for $\mu_{\rho w}$ when the population grand SD of Y is different from that of person-mean centered Y ($\sigma_{CY} \neq \sigma_Y$) and $\mu_{\rho w} \neq 0$. In contrast, P-S yields consistent estimates of $\mu_{\rho w}$.

Note that the normality assumption was not used in the derivations of asymptotic estimation performance. Under the heterogeneous within-person relations, the derivations were based on the condition in which all individuals have the same number of time points so that simple-form results can be obtained. In the next section, we evaluated the performance of the standardization methods for estimating and inferring within-person relations under both equal and unequal number of assessments conditions with a finite sample size.

The aforementioned derivation results apply to the multilevel models with only one time-varying predictor or bivariate within-person relations. With two or more predictors, the standardized coefficients are functions of the relevant bivariate correlations. When the bivariate within-person relations are inconsistently recovered by a global standardization approach, one can infer that the standardized coefficient estimates from the global standardization approach for multilevel models with two or more predictors are generally inconsistent estimates of the multivariate within-person relations.

A simulation study

In this section, we conducted a simulation study to evaluate the inferential properties of the global standardization and within-person standardization approaches for models with one or two time-varying predictors. The simulation study is helpful for evaluating the performance

of the methods with finite samples (finite N and finite T_i) in terms of both estimation accuracy (bias in the point estimates) and statistical inference (coverage rates of confidence intervals).

Simulation design

For models with one time-varying predictor (X_{it}), we generated $\rho_{w,i}$ from a normal distribution with $\mu_{\rho w}$ = 0, −.25, or −.5 (corresponding to the null, small-medium, and large correlations) and SD = 0 or .25 (corresponding to the homogeneous and heterogeneous within-person relation conditions). For each of the N = 50, 100, or 300 individuals, T = 5, 10, 20, 30, 56, or 100 data points of X^{PS} and Y^{PS} were generated from a bivariate normal distribution with means of 0, SDs of 1, and the within-person correlation $\rho_{w,i}$. Then, N person mean scores of X and N person mean scores of Y were generated from a bivariate normal distribution with means of 13 and 30 and a covariance matrix of (19, −6.5, −6.5, 70). Meanwhile, N person SD scores of X and N person SD scores of Y were generated from a bivariate gamma distribution with means of 2.7 and 5.3 and a covariance matrix of (3, 0.6, 0.6, 2.5), (0, 0, 0, 2.5), or (0, 0, 0, 0). The middle covariance matrix corresponds to the situation in which between-person differences in person SDs of X do not exist, but do for Y (an ideal scenario considered in the derivations). The zero covariance matrix is for the situation in which there are no between-person differences in the person SDs for both variables (another ideal scenario considered in the derivations). Raw scores of X and Y were then generated. The multilevel models in Eqs (2), (4), and (9) were fitted to the observed person-mean centered or WP standardized data. In the simulations, the correlations between the observed and true person SDs were not 1, but were about .62, .76, .87, .91, .95, and .97 for Y with T = 5, 10, 20, 30, 56, or 100 respectively. The corresponding values are slightly higher or similar for X. The P-S approach (M_{PS}) was implemented by fitting the model in Eq (9). The global standardization approaches (M_{G1}, M_{G2}, and M_{G3}) were implemented using Eqs (5) to (8). In addition, the EB standardization approach (M_{EB}) was also implemented. The EB standardization approach conducts standardization at the person level, so we expect reasonably accurate point estimates of $\mu_{\rho w}$. However, the EB estimates suffer a shrinkage problem and the problem becomes more severe when T is smaller (e.g., Fitzmaurice et al., 2011). Thus, we expect poor inferential properties (e.g., undercoverage) from the EB method when T is not sufficiently large. In total, when there is one time-varying predictor, we had $3 \times 6 \times 3 \times 2 \times 3 = 324$ conditions for each of the five evaluated methods.

For models with two time-varying predictors (X_{1it} and X_{2it}), we generated N = 50, 100, or 300 sets of three pairwise within-person correlations, $\rho_{w1,i}$ (X_{1it} and X_{2it}), $\rho_{w2,i}$ (X_{1it} and Y_{it}), and $\rho_{w3,i}$ (X_{2it} and Y_{it}). Specifically, each follows a normal distribution with $mean$ = .50 and SD = .20, $mean$ = −.20 and SD = .25, and $mean$ = −.40 and SD = .25, respectively. For each of the N individuals, T = 5, 10, 20, 30, 56, or 100 data points of X_1^{PS}, X_2^{PS}, and Y^{PS} were generated from a trivariate normal distribution with means of 0, SDs of 1, and the three pairwise within-person correlations. Then, N person mean scores for each of X_1, X_2, and Y were generated from a trivariate normal distribution with means of 13, 19, 30 and a covariance matrix of $\begin{pmatrix} 19 & & \\ 20 & 35 & \\ -6 & -22 & 70 \end{pmatrix}$. Meanwhile, N person SD scores for each of X_1, X_2,

and Y were generated from a trivariate gamma distribution with means of 2.7, 3.5, and 5.3 and a covariance matrix of $\begin{pmatrix} 3 & & \\ 1.2 & 1.2 & \\ 0 & 0 & 2.4 \end{pmatrix}$ or a 0 covariance matrix. The 0 covariance matrix is for the situation in which there are no between-person differences in the person SDs for any of the three variables. The other data generation and model fitting steps were the same as those in the bivariate simulation settings except that two predictors were included in the multilevel models. Theoretically, the two standardized fixed-effects within-person coefficients of X_1 and X_2 are 0 and −0.4 at the population level. Empirically, some of the generated correlations that are above 1 (or below −1) were set to be .99 (or −.99) and thus the empirical reference values were −.0135 and −.3907 when there are between-person differences in the person SDs and −.0126 and −.3914 when there are not.

Recall that under the heterogeneous within-person relation condition, the derivations were done with the data assumption that there are equal number of assessments across individuals for obtaining simple-form results. With the data assumption, we found that global standardization generally yields inconsistent estimates of the average within-person correlation. When the data assumption is violated, we expect that the same conclusion still holds for global standardization. It was less clear, however, how within-person standardization performs when individuals have different numbers of assessments. Thus, to further evaluate the performance of the methods, we also generated T_i from a Poisson distribution with $mean(T) = sd(T)$ = 5, 10, 20, 30, 56, or 100 respectively for the simulations with two time-varying predictors. The distributions of the number of time points are displayed in Part C of the online supplemental materials. In total, $3 \times 6 \times 2 \times 2 = 72$ conditions were considered when there are two time-varying predictors. For both bivariate and trivariate simulations, the non-null population values for the person mean and person SD variables were adopted based on the real data in the real data example section.

For each of the conditions in which individual differences in within-person relations exist, all of the five standardization methods were evaluated. For each of the conditions in which individual differences in within-person relations do not exist, four of the five standardization methods were evaluated because the EB standardization approach is not applicable here. The number of replications is 1000 for each condition. Bias or relative bias based on point estimates and coverage rates based on 95% confidence intervals were evaluated for each method under each condition. Relative biases more than 10% are treated as nonignorable and coverage rates outside the range of 91% – 98% are viewed as unsatisfactory (Muthén & Muthén, 2002).

Simulation results

Due to space limitations, a subset of the simulation results from the models with one predictor (bivariate) and two predictors (trivariate) are included in Tables 3–4 and Table 5, respectively. The results from the other conditions share very similar patterns and those from the conditions with $N = 50$ or $N = 300$ are displayed in Part E of the online supplemental materials.

Table 3 displays the bivariate results when $\mu_{\rho w} = 0$. When $\mu_{\rho w} = 0$, regardless of whether there are between-person differences in person SDs of X and Y, our derivation results showed that $\rho_{CX,CY} = \mu_{\rho w} = 0$ and thus all the global standardization approaches (M_{G1}, M_{G2}, and M_{G3}) and the P-S standardization approach (M_{PS}) can yield consistent estimates of $\mu_{\rho w}$. The simulation results were aligned with the derivation results. In addition, M_{G1}, M_{G2}, M_{G3}, and M_{PS} yielded satisfactory coverage rates except when M_{G3} is used for data with 5 time points under the scenario in which between-person differences in the person SDs of both X and Y exist. The empirical biases in estimating $\mu_{\rho w}$ from the EB standardization approach (M_{EB}) were also 0.00 but the coverage rates were all below 90% when $T \leq 56$. When $T = 100$, M_{EB} had more satisfactory coverage rates (90.8% to 92.0%).

Table 4 displays the bivariate results when $\mu_{\rho w} = 0.5$. When $\mu_{\rho w} \neq 0$, our derivation results showed that the M_{G2} approach generally yields inconsistent estimates of $\mu_{\rho w}$ regardless of whether there are between-person differences in person SDs of X and Y. The simulation results confirmed the derivation results. Specifically, M_{G2} yielded biased estimates of $\mu_{\rho w}$ with nonignorable relative biases and poor coverage rates under all of the conditions with $\mu_{\rho w} = 0.5$. Increasing N and/or T did not improve its performance. When $\mu_{\rho w} = 0.5$, the biases in estimating $\mu_{\rho w}$ from the EB standardization approach were ignorable when $T \geq 20$ but the coverage rates were below 91% in most of the studied conditions (M_{EB} in Table 4). As T increases, performance of the EB standardization approach became better in both point estimation accuracy and coverage rates.

As shown in our derivations, when $\mu_{\rho w} \neq 0$, estimation performance of the other two global standardization methods, M_{G1} and M_{G3}, depends on whether there are between-person differences in person SDs of X and Y. First, when between-person differences in person standard deviations of both X and Y exist, our derivation results showed that M_{G1} and M_{G3} generally yielded inconsistent estimates of $\mu_{\rho w}$ with a nonzero $\mu_{\rho w}$. Under the scenario and $\mu_{\rho w} = 0.5$, our simulation results showed that M_{G1} and M_{G3} had nonignorable relative biases in estimating $\mu_{\rho w}$ under most of the relevant simulation conditions. Increasing N and/or T did not lower the relative biases. Furthermore, there was severe undercoverage in the 95% confidence intervals from M_{G1} and M_{G3}. Second, when $\mu_{\rho w} \neq 0$ and between-person differences in person standard deviations exist in Y but not in X, our derivation results showed that M_{G1} and M_{G3} generally also yields inconsistent estimates of $\mu_{\rho w}$. Under this scenario, although it appeared that M_{G1} and M_{G3} had ignorable relative biases in estimating $\mu_{\rho w}$, increasing N and/or T did not reduce the relative biases and made the coverage rates even more unsatisfactory (the second row block of Table 4). Third, when $\mu_{\rho w} = 0.5$ and between-person differences in person standard deviations of X and Y do not exist, our derivation results showed that M_{G1} and M_{G3} can yield consistent estimates of $\mu_{\rho w}$. The simulation results were also aligned with the derivation results (Table 4). In addition, M_{G1} and M_{G3} had very similar results when T is large (e.g., $T \geq 56$), aligned with our derivations that the results from the two methods asymptotically approach the same values. When T is small (e.g., 5 or 10), M_{G1} performed better than M_{G3} in the simulations.

The P-S approach (M_{PS}) yielded ignorable biases in estimating $\mu_{\rho w}$ and the biases became lower when T increases (Table 4). This was consistent with our derivation results that the estimates from P-S are consistent for estimating $\mu_{\rho w}$. For obtaining satisfactory coverage

rates, our simulation results revealed that, in some conditions, 30 or more measurement occasions may be needed in order to use P-S (Table 4).

The results from the models with two predictors (see Table 5 for the results under both the equal T_i and unequal T_i data conditions) shared the same patterns as those from the models with one predictor, regardless of whether there are equal or unequal numbers of time points across individuals. Thus, we do not repeat the details here.

Summary of the simulation results

In summary, our simulation results were consistent with the derivations in that (1) global standardization methods including M_{G1} and M_{G3} yielded inaccurate estimates of fixed-effects within-person relations when the average relations exist (e.g., $\mu_{\rho w} \neq 0$) and between-person differences in person SDs exist; (2) global standardization M_{G2} yielded inaccurate estimates of fixed-effects within-person relations when the average relations exist (e.g., $\mu_{\rho w} \neq 0$) regardless of whether between-person differences in person SDs exist; (3) P-S (M_{PS}) yielded accurate estimates of fixed-effects within-person relations when T is not too small (e.g., $T \geq 10$); and (4) the estimates from the EB standardization approach (M_{EB}) also had ignorable biases when T is not too small.

The simulation results revealed extra information about inferential properties of the approaches. Generally, when average within-person relations do not exist (e.g., $\mu_{\rho w} = 0$), M_{EB} was the only method that yielded unsatisfactory coverage rates. When average within-person relations exist (e.g., $\mu_{\rho w} \neq 0$), global standardization M_{G2} had unsatisfactory coverage rates in all of the studied conditions. When average within-person relations exist (e.g., $\mu_{\rho w} \neq 0$) and between-person differences in person SDs exist, global standardization approaches including M_{G1} and M_{G3} had unsatisfactory coverage rates in most of the studied conditions. When average within-person relations exist (e.g., $\mu_{\rho w} \neq 0$) and between-person differences in person SDs do not exist, M_{G1} and M_{G3} had satisfactory coverage rates. P-S (M_{PS}) yielded satisfactory coverage rates when $T \geq 30$ and $N \geq 50$. In many conditions, the coverage rates were also satisfactory with a smaller T (e.g., 10 or 20 occasions).

A summary of all the evaluated methods is listed in Table 6. The consistency performance of the methods and the performance with finite samples are summarized in the last three columns of Table 6.

A real data analysis example

In this section, we use real daily diary data to illustrate similarities and differences in the results from the various standardization approaches. Substantively, we analyze the data to better understand the emotional complexity issue, or the degree of relationship between positive and negative affect. The issue of ongoing debate is whether positive and negative emotions are opposite ends of a bipolar continuum or independent dimensions in a bivariate distribution. The Affect Circumplex model suggests that one would expect positive affect (PA) and negative affect (NA) to be independent of one another (Watson, Weise, Vaidya & Tellegen, 1999). The Dynamic Model of Affect (DMA), on the other hand, stresses the importance of contextual factors in feelings and emotions. For example, using the DMA

perspective, one would predict that under stress, affect becomes bipolar (Zautra, Potter, & Reich, 1997). In this real data analysis example, daily data on PA, NA and Stress were analyzed using multilevel modeling to study the emotional complexity issue from the day-to-day intraindividual variability perspective. Specifically, we are interested in how these variables are related intraindividually over time. In this example, within-person relations among the variables are more useful than between-person relations.

A subset of data from the middle-aged and older cohorts of the Notre Dame Study of Health & Well-Being (Bergeman & Deboeck, 2014) were used. For the purpose of illustration, we chose cases that have complete data for all variables involved in the analyses ($N = 101$). The participants, ranging in age from 31 to 91, were measured daily on PA and NA using the Positive Affect and Negative Affect scale (PANAS; Watson et al., 1988) and stress using the Perceived Stress Scale (Cohen et al., 1983) for 56 onsecutive days. Therefore, for each variable, we have $101 \times 56 = 5656$ data points. In addition, the data were detrended to eliminate the influence of trend in the time series when trends were detected for some participants. The rationale of detrending was that we did not expect increasing or decreasing patterns in the variables due to the nature of the sample of typical aging adults. With the detrended data, we studied the relations after controlling for the effect of time (i.e., day). More details about how the data were detrended can be found in Zhang, Wang, and Bergeman (2018). Descriptive statistics of the variables are displayed in Table 7 and summary statistics of the person mean and SD variables are displayed in Table 8.

The standard deviations of the person means were 4.36 (95% bootstrap percentile confidence interval [95% BPCI]: 3.06, 5.59), 5.92 (95% BPCI: 4.74, 7.01), and 8.39 (95% BPCI: 6.95, 9.74) for NA, Stress, and PA respectively. The standard deviations of the person standard deviations were 1.72 (95% BPCI: 1.39, 1.99), 1.29 (95% BPCI: 1.10, 1.46), and 1.53 (95% BPCI: 1.35, 1.69), for NA, Stress, and PA respectively. Thus, statistically significant between-person differences existed in both the person means and person SDs of all three variables.

The sample correlations were different with different kinds of stacked long data. For example, the sample correlation between NA and Stress from raw data was .72, whereas that from person-mean centered data was .65 and from P-S standardized data was .55. We also applied our derived formulas for calculating the sample correlations. Plugging the relevant values into the formulas, as shown in Part D of the online supplemental materials, we can recover the sample correlation estimates. Thus, the calculations confirmed the accuracy of the population correlation derivations in the online supplemental materials.

We applied the five standardization approaches to various multilevel models with NA or Stress predicting PA and models with both NA and Stress predicting PA. Specifically, the single predictor model (e.g., NA predicting PA) was fitted to study the concurrent within-person relations between daily NA and daily PA, whereas the double predictor model (both NA and Stress predicting PA) was fitted to study the within-person relations between daily NA and daily PA after controlling for daily Stress. The results are presented in Table 9.

Methodologically, the real data analysis results from the models with one predictor were consistent the derivation and simulation results. For example, our analytical and simulation results revealed that global standardization generally does not accurately recover the average within-person relations but the P-S approach can. Under the heterogeneous within-person relation assumption, we obtained $\hat{\gamma}^{G1*}_{10,heter,NA} = \hat{\gamma}^{G3*}_{10,heter,NA} = -.29$ and $\hat{\gamma}^{G1*}_{10,heter,Stress} = \hat{\gamma}^{G3*}_{10,heter,Stress} = -.46$, which were substantially different from those of the P-S approach ($\hat{\gamma}^{PS}_{10,NA} = -.23$ and $\hat{\gamma}^{PS}_{10,Stress} = -.40$; 26% = .06/.23 and 15% = .06/.40 relative differences) respectively. As another instance, with $T = 56$, the standard error estimates from the EB standardization approach (M_{EB}) were smaller than those from the P-S approach (M_{PS}) due to the shrinkage problem. This is consistent with the undercoverage problem from EB when T is not large enough.

The standard error estimates from the P-S models (M_{PS}) under the homogeneous within-person relation assumption were all less than half of those from the P-S models that relaxed the assumption. For our real example, there were statistically significant different within-person relations across individuals before and after controlling for the other predictor, using various variance tests (e.g., Ke & Wang, 2015). The homogeneous within-person relation assumption is not met for our example. Therefore, u_{1i} should be included in the multilevel models.

Based on the derivation and simulation results, given that we have 56 occasions of data (> 30), significant between-person differences in person SDs, and the research interest is in evaluating within-person relations, we recommend the results from the P-S approach (M_{PS}) under the heterogeneous within-person relation assumption for this example. Substantively, using P-S, we found that on average, daily NA did not relate intraindividually to daily PA after controlling for daily Stress (.001 with its standard error estimate at .030; 95% CI: [−.058, .060]), but daily Stress significantly related intraindividually to daily PA after controlling for daily NA (−.40 with its standard error estimate at .032: 95% CI: [−.463, −.337]). Note that the results from global standardization yielded noticeably different point and interval estimates of within-person relations, although the significance of the results was the same. Schuurman et al. (2016) stated that the person-specific Granger-causal processes should not involve between-person differences in the processes. For these data, we found that the grand SDs used in global standardization involve the between-person variance in within-person SDs (M_{G1} and M_{G3}) or between-person variances in both within-person SDs and within-person means (M_{G2}), as shown in Table 2 and Part B of the online supplemental materials. Therefore, we think that researchers may be less interested in the findings from global standardization than those from within-person standardization.

Recalling the debate on the two substantive theories about emotional complexity, our results supported both. That is, at the person level, daily NA was a significant predictor of daily PA; whereas after controlling for daily stress, daily PA and daily NA were not significantly correlated. It is interesting to note that these relations are salient not only under extreme duress, which is usually tested (e.g., Zautra, Smith, Affeck & Tennen, 2001), but also hold under typical daily hassles.

Discussion

In the current study, we evaluated five standardization approaches for studying within-person relations using multilevel modeling. The global standardization approaches we evaluated are based on the sample grand standard deviations of a variable. Global standardization is currently recommended in multilevel modeling textbooks for standardizing fixed-effects coefficients (e.g., Hox, 2010; Snijders & Bosker, 2012). Our derivations showed that when there are individual differences in the within-person standard deviations and the average within-person relation is not 0, global standardization generally yields inconsistent (asymptotically biased) estimates of fixed-effects within-person relations (e.g., $\mu_{\rho w}$). Our simulation results expanded the findings to performance with finite samples. In the real example, all three time-varying variables had substantial individual differences in within-person standard deviations and thus the results from global standardization can be misleading for describing the average within-person relations.

The current article also provides both analytical and simulation results to show whether and when the person-mean-SD standardization (P-S) approach can be used for estimating and inferring about within-person relations. Specifically, our derivation results revealed that the within-person relations can be consistently estimated from P-S, regardless of whether there are individual differences in the within-person standard deviations. In addition, our simulation results revealed that the P-S approach had ignorable estimation bias when $T \geq 10$ and $N \geq 50$, and satisfactory coverage rates when $T \geq 30$ and $N \geq 50$. In many conditions, P-S had satisfactory coverage rates with a smaller T (e.g., 10 or 20 occasions) as well.

A benefit of using the P-S approach is that the within-person relation estimates satisfy Kelley and Preacher's (2012) properties of a good effect size estimate. Specifically, fixed-effects within-person relation estimates from P-S are invariant to the change of the units of the measurements. In addition, it is easy to obtain confidence interval estimates as shown in the real example. A practical implication is that the fixed-effects within-person relation estimates from P-S can be used for comparing the quantitative strengths of within-person relations from different pairs of time-varying variables.

The choice of a standardization method depends on the estimand and the research goal

When considering the type of centering to use, Enders and Tofighi (2007), Hamaker and Grasman (2015), and Kreft et al. (1995) all emphasized that one may prefer different centering methods for different estimands (the target of estimation/inference) in cross-sectional or dynamic multilevel modeling. On standardization, we echo that the choice of a standardization method also depends on the estimand. In this study, our focus was on estimating and making inferences about within-person relations (e.g., $\mu_{\rho w}$). With this estimand in mind, we argue for P-S over global standardization. Comparing P-S or WP standardized coefficients across individuals on the same pair of variables or across different pairs of variables, however, may have limited use under some circumstances. For example, when guiding decisions concerning interventions, standardization takes into account neither the relative difficulty nor importance of manipulating a variable in practice. In our real example, stress and positive affect had a statistically higher average within-person relation compared to the pair of negative affect and positive affect. Decreasing negative affect by one

standard deviation, however, may be easier than decreasing stress by one standard deviation in a real life context. As another instance, when the raw within-person effect coefficients of stress on positive affect are the same between two individuals (e.g., −.65), we can interpret the results such that when stress increases by 1 point, positive affect decreases by .65 point for both individuals. After WP standardization, the two standardized coefficients can be different (e.g., $-.65\frac{s_{x1}}{s_{y1}} = -.65\frac{0.5}{0.5} = -.65$ for the first individual and $-.65\frac{s_{x2}}{s_{y2}} = -.65\frac{0.5}{1} = -.325$ for the second individual). In this case, we interpret the standardized results such that when stress increases by 1 within-person SD, positive affect decreases by .65 within-person SD for the first individual whereas positive affect decreases by .325 within-person SD for the second individual. The standardized coefficient is larger for the first individual due to his or her smaller within-person SD in the outcome variable (i.e., positive affect). Whether the raw within-person coefficients or the WP standardized coefficients are more useful for evaluating invention effects for these two individuals is arguable and depends on the context. Thus, comparing within-person standardized coefficients may have limitations in developing new interventions.

When raw scales are compatible with within-person relations, standardization may not be needed at all. Moreover, if the time-varying predictor and/or outcome are binary, different considerations may need to be made on whether and how to standardize the variables (e.g., Hedges, 2007); future research on this issue is needed. Therefore, it is important to identify the estimand and the research goal with substantive/theoretical considerations (e.g., see a discussion in Schuurman et al., 2016), and then select an appropriate standardization approach for estimating the estimand of interest.

Future research directions

In the derivations, the within-person mean variables as well as the within-person SD variables are reliable because T is infinity. Our derivation results have shown that even under this best-case condition, global standardization may yield asymptotically-biased estimates of average within-person relations. For our real data example, the estimated reliabilities of the person mean variables were above .99 and those of the within-person SD variables were all above .92, calculated using the formulas in Wang and Grimm (2012). Thus, the observed person mean and SD variables used in the P-S multilevel models in the real data example were highly reliable. When the number of time points is relatively small, however, the within-person SD variables can be unreliable (e.g., Du & Wang, 2018; Estabrook et al., 2012; Wang & Grimm, 2012). Our simulation results did reveal that a relatively large number of occasions (e.g., 30 or more) may be needed for obtaining satisfactory coverage rates from the P-S approach. When the number of time points is not large (e.g., 5 or 10), we should use the P-S approach with caution, especially for statistical inference (e.g., confidence interval construction).

When T is relatively small, latent variables or random effects may be helpful for modeling the within-person means and SDs. For example, Schuurman et al. (2016) used sophisticated Bayesian modeling with MCMC to obtain person variances as random effects for

standardizing the within-person random-effects coefficients. Future research should investigate how to use Bayesian modeling with MCMC to model random effects for intraindividual/person means and variances (instead of using observed person means and variances) in the P-S approach. Furthermore, as shown in Table 1, standardizing the variables before multilevel modeling vs. standardizing the random coefficients during or after multilevel modeling can result in different models due to different modeling assumptions. It is not immediately clear, however, which assumption should be preferred and which estimation method (Bayesian vs. frequentist) should be preferred. Our expectation was that when the modeling assumption is met for the Bayesian approach in Schuurman et al. (2016), the approach may require fewer time points than our evaluated within-person standardization approaches (M_{PS} and M_{EB}) for making accurate inferences about within-person relations, due to their capability of directly modeling person variances as random effects. In addition, our simulation results revealed that M_{EB} required many more time points of data than M_{PS} (e.g., 100 vs. 30 for some conditions). This may be explained by the current simulation study design: The standardized coefficients were generated to follow a normal distribution whereas the unstandardized coefficients were not. We conducted a small simulation study to evaluate the performance of M_{EB} and M_{PS} when the unstandardized coefficients followed a normal distribution but the standardized coefficients did not. Our simulation results from the additional conditions revealed that M_{EB} required 20 or more time points of data whereas M_{PS} required 10 or more time points of data (simulation results are listed in Part E of the online supplemental materials). Thus the data requirement differences appeared smaller, but M_{EB} still required more time points than M_{PS}. Future research should evaluate the different modeling and estimation approaches for within-person standardization.

Another possibility is to conduct the analysis in the latent variable modeling framework. Lüdtke et al. (2008) modeled cluster means as latent intercepts for better understanding between-cluster effects. They did not, however, allow within-cluster effects to vary across clusters. Nor did they include cluster SDs or variances as latent variables in the models. Therefore, future research should be done to develop and evaluate the possibility. In addition, the location-scale model proposed by Hedeker and colleagues allows for direct modeling of heterogeneous within-person variances in only the outcome variable using the frequentist framework (e.g., Hedeker, Mermelstein, Berbaum, & Campbell, 2008; Hedeker, Mermelstein, Demirtas, 2009, 2012). Future research should be done to explore extensions of this model to allow for direct modeling of heterogeneous within-person variances in both the predictor and outcome variables.

Non-stationarity (e.g., trends or cycles) may exist in the time-varying variables. For example, when trends exist, detrending may or may not be needed, depending on the research goals (e.g., Curran & Bauer, 2011; Liu & West, 2016; Wang & Maxwell, 2015). Thus, we have the following data analysis recommendations for studying within-person relations (e.g., correlations) when trends exist in the time-varying variables. First, whether detrending is necessary should be determined. As discussed in Wang and Maxwell (2015), detrending is necessary when the research interest is on net-time (conditional) within-person relations, but is not necessary when the focus is on original (unconditional) within-person relations. Second, when detrending is needed, either (a) detrending first then using the model

in Eq 9 or (b) including time (or a function of time) as a Level-1 covariate in Eq 9 can be done to make inferences about fixed-effects net-time within-person relations. Future research is needed to compare the performance of the two methods. Third, when detrending is not needed, the multilevel model in Eq 9 can be fitted to make inferences about fixed-effects within-person relations when the number of time points is large enough. Note that when non-stationarity exist, Granger causality can not be implied (Granger, 1969) and only association relationships can be implied. Overall, future research can be done to propose and evaluate standardization methods for better understanding within-person processes when the dynamic processes are not stationary.

Concluding remarks

We analytically derived that when both (a) between-person differences in within-person standard deviations exist and (b) the average within-person relation is not 0, global standardization generally yields inconsistent (asymptotically biased) estimates for the average within-person relations (the estimand of interest in the current study). We hope that our results can help researchers further appreciate the importance of considering individual differences in intraindividual variability and understand the consequences of ignoring the differences when studying within-person relations. Alternatively, a within-person standardization approach (within-person standardize both the time-varying predictor and outcome variables and the standardized variables are used in multilevel modeling; P-S) yields consistent estimates. Our simulation results were aligned with the derivation results and further revealed that the P-S approach provided proper confidence interval coverage of fixed-effects within-person relations when the number of occasions is 30 or more (in many conditions, performance was also satisfactory with 10 or 20 occasions). To conclude the paper, we want to emphasize that different studies may have different research questions and thus estimands of interest. Therefore, we suggest that researchers should make the choice of standardization based on theoretical considerations and should clearly describe the purpose and procedure of standardization in research articles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Aafjes-van Doorn K, Lilliengren P, Cooper A, Macdonald J, & Falkenström F (2017). Patientsâ affective processes within initial experiential dynamic therapy sessions. Psychotherapy, 54(2), 175–183. doi: 10.1037/pst0000072 [PubMed: 27869472]

Armeli S, OâHara RE, Ehrenberg E, Sullivan TP, & Tennen H (2014). Episode-specific drinking-to-cope motivation, daily mood, and fatigue-related symptoms among college students. Journal of Studies on Alcohol and Drugs, 75(5), 766–774. doi: 10.15288/jsad.2014.75.766 [PubMed: 25208194]

Baird R (2016). The effect of misspecifying random-effect time-varying predictors as fixed on estimates of other parameters (doctoral dissertation) University of Notre Dame, Notre Dame, IN, USA.

Baird R, & Maxwell SE (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. Psychological Methods, 21(2), 175–188. doi: 10.1037/met0000070 [PubMed: 26950731]

Bergeman C, & Deboeck PR (2014). Trait stress resistance and dynamic stress dissipation on health and well-being: The reservoir model. Research in Human Development, 11(2), 108–125. doi: 10.1080/15427609.2014.906736 [PubMed: 29354022]

Bolger N, & Laurenceau J-P (2013). Intensive longitudinal methods: An introduction to diary and experience sampling research New York, NY, US: Guilford Press.

Bryk AS, & Raudenbush SW (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101(1), 147–158. doi: 10.1037//0033-2909.101.1.147

Cohen S, Kamarck T, & Mermelstein R (1983). A global measure of perceived stress. Journal of Health and Social Behavior, 24(4), 385–396. Retrieved from http://www.jstor.org/stable/2136404 [PubMed: 6668417]

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. doi: 10.1146/annurev.psych.093008.100356

Dejonckheere E, Bastian B, Fried EI, Murphy SC, & Kuppens P (2017). Perceiving social pressure not to feel negative predicts depressive symptoms in daily life. Depression and Anxiety, 34, 836–844. doi: 10.1002/da.22653 [PubMed: 28499066]

Dejonckheere E, Mestdagh M, Houben M, Erbas Y, Pe M, Bastian B, … Kuppens P (2018). The bipolarity of affect and depressive symptoms. Depression and Anxiety doi: 10.1037/pspp0000186

Du H, & Wang L (2018). Reliabilities of intraindividual variability indicators with autocorrelated longitudinal data: Implications for longitudinal study designs. Multivariate Behavioral Research doi: 10.1080/00273171.2018.1457939

Eid M, & Diener E (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. Journal of Personality and Social Psychology, 76 (4), 662–676. doi: 10.1037/0022-3514.76.4.662

Enders CK, & Tofighi D (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. Psychological Methods, 12, 121–138. doi: 10.1037/1082-989X.12.2.121 [PubMed: 17563168]

Estabrook R, Grimm KJ, & Bowles RP (2012). A Monte Carlo simulation study of the reliability of intraindividual variability. Psychology and Aging, 27, 560–576. doi: 10.1037/a0026669 [PubMed: 22268793]

Ferrer E, Gonzales JE, & Steele J (2013). Intra-and interindividual variability of daily affect in adult couples. GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry, 26(3), 163–172. doi: 10.1024/1662-9647/a000095

Fiske DW, & Rice L (1955). Intra–individual response variability. Psychological Bulletin, 52, 217–250. doi: 10.1037/h0045276 [PubMed: 14371891]

Fitzmaurice GM, Laird NM, & Ware JH (2011). Applied longitudinal analysis (2nd ed.). Hoboken, NJ, US: John Wiley & Sons.

Foshee VA, Benefield TS, Reyes HLM, Ennett ST, Faris R, Chang L-Y, … Suchindran CM (2013). The peer context and the development of the perpetration of adolescent dating violence. Journal of Youth and Adolescence, 42(4), 471–486. doi: 10.1007/s10964-013-9915-7 [PubMed: 23381777]

Freeman LK, & Gottfredson NC (2017). Using ecological momentary assessment to assess the temporal relationship between sleep quality and cravings in individuals recovering from substance use disorders. Addictive Behaviors doi: 10.1016/j.addbeh.2017.11.001

Gerstorf D, Siedlecki KL, Tucker-Drob EM, & Salthouse TA (2009). Within-person variability in state anxiety across adulthood: Magnitude and associations with between-person correlates. International Journal of Behavioral Development, 33, 55–64. doi: 10.1177/0165025408098013

Granger CW (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424–438.

Hamaker EL (2012). Why researchers should think “within-person”: A paradigmatic rationale. In Mehl M & Conner T (Eds.), Handbook of Methods for Studying Daily Life (pp. 43–61). New York, NY: Guilford Publications.

Hamaker EL, Dolan CV, & Molenaar P (2005). Statistical modeling of the individual: Rationale and application of multivariate time series analysis. Multivariate Behavioral Research, 40, 207–233. doi: 10.1207/s15327906mbr4002-3 [PubMed: 26760107]

Hamaker EL, & Grasman RP (2015). To center or not to center? investigating inertia with a multilevel autoregressive model. Frontiers in Psychology doi: 10.3389/fpsyg.2014.01492

Hamaker EL, Kuiper RM, & Grasman RP (2015). A critique of the cross-lagged panel model. Psychological Methods, 20(1), 102–116. doi: 10.1037/a0038889 [PubMed: 25822208]

Hedeker D, Mermelstein R, Berbaum M, & Campbell R (2009). Modeling mood variation associated with smoking: An application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. Addiction, 104, 297–307. doi: 10.1111/j.1360-0443.2008.02435.x [PubMed: 19149827]

Hedeker D, Mermelstein RJ, & Demirtas H (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics, 64, 627–634. doi: 10.1111/j.1541-0420.2007.00924.x [PubMed: 17970819]

Hedeker D, Mermelstein RJ, & Demirtas H (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. Statistics in Medicine, 31(27), 3328–3336. doi: 10.1002/sim.5338 [PubMed: 22419604]

Hedges LV (2007). Effect sizes in cluster-randomized designs. Journal of Educational and Behavioral Statistics, 32(4), 341–370. doi: 10.3102/1076998606298043

Hox JJ (2010). Multilevel analysis: Techniques and applications (second ed.). New York, NY: Routledge.

Jongerling J, Laurenceau J-P, & Hamaker EL (2015). A multilevel AR (1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. Multivariate Behavioral Research, 50(3), 334–349. doi: 10.1080/00273171.2014.1003772 [PubMed: 26610033]

Ke Z, & Wang L (2015). Detecting individual differences in change: Methods and comparisons. Structural Equation Modeling, 22(3), 382–400. doi: 0.1080/10705511.2014.936096

Kelley K, & Preacher KJ (2012). On effect size. Psychological Methods, 17(2), 137–152. doi: 10.1037/a0028086 [PubMed: 22545595]

Kreft IGG, de Leeuw J, & Aiken LS (1995). The effect of different forms of centering in hierarchical linear models. Multivariate Behavioral Research, 30, 1–21. doi: 10.1207/s15327906mbr3001-1 [PubMed: 26828342]

Liu Y, & West SG (2016). Weekly cycles in daily report data: An overlooked issue. Journal of Personality, 84(5), 560–579. doi: 10.1111/jopy.12182 [PubMed: 25973649]

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. doi: 10.1037/a0012869 [PubMed: 18778152]

Lydon-Staley DM, Xia M, Mak HW, & Fosco G (2018). Adolescent emotion network dynamics in daily life and implications for depression doi: Obtained from psyarxiv.com.

Maxwell SE, & Delaney HD (2004). Designing experiments and analyzing data: A model comparison perspective (2nd ed.). New York, NY: Taylor & Francis Group.

Michela JL (1990). Within-person correlational design and analysis. In Hendrick C & Clark MS (Eds.), Review of personality and social psychology, Vol. 11, Research methods in personality and social psychology (pp. 279–311). Thousand Oaks, CA: Sage Publications.

Molenaar PCM (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology – this time forever. Measurement: Interdisciplinary Research and Perspectives, 2, 201–218. doi: 10.1207/s15366359mea0204$_{11}$

Molenaar PCM, & Campbell CG (2009). The new person-specific paradigm in psychology. Current Directions in Psychological Science, 18(2), 112–117. doi: 10.1111/j.1467-8721.2009.01619.x

Muthén LK, & Muthén BO (2002). How to use a monte carlo study to decide on sample size and determine power. Structural Equation Modeling, 9(4), 599–620. doi: 10.1207/S15328007SEM0904-8

Nesselroade JR (1991). The warp and woof of the developmental fabric. In Downs R, Liben L, & Palermo D (Eds.), Visions of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill (pp. 213–240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nesselroade JR, & Molenaar P (2010). Emphasizing intraindividual variability in the study of development over the life span. In Lerner RM, Lamb ME, & Freund AM (Eds.), The handbook of life-span development (pp. 30–54). Hoboken, NJ: John Wiley & Sons.

Nesselroade JR, & Salthouse TA (2004). Methodological and theoretical implications of intraindividual variability in perceptual motor performance. Journals of Gerontology: Psychological Sciences, 59B, 49–55. doi: 10.1093/geronb/59.2.P49

Ram N, & Gerstorf D (2009). Time-structured and net intra-individual variability: Tools for examining the development of dynamic characteristics and processes. Psychology and Aging, 24(4), 778–791. doi: 10.1037/a0017915 [PubMed: 20025395]

Ramseyer F, Kupper Z, Caspar F, Znoj H, & Tschacher W (2014). Time-series panel analysis (tspa): Multivariate modeling of temporal associations in psychotherapy process. Journal of Consulting and Clinical Psychology, 82(5), 828–838. doi: 10.1037/a0037168 [PubMed: 24932566]

Raudenbush S, & Bryk A (2002). Hierarchical linear models (second edition). Thousand Oaks, CA, US: Sage Publications.

Schmiedek F, Lövdén M, & Lindenberger U (2009). On the relation of mean reaction time and intraindividual reaction time variability. Psychology and Aging, 24, 841–857. doi: 10.1037/a0017799 [PubMed: 20025400]

Schuurman NK, Ferrer E, de Boer-Sonnenschein M, & Hamaker EL (2016). How to compare cross-lagged associations in a multilevel autoregressive model. Psychological Methods, 21(2), 206–221. doi: 10.1037/met0000062 [PubMed: 27045851]

Snijders TAB, & Bosker RJ (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling (2nd ed.). London, UK: Sage.

Wang L, & Grimm K (2012). Investigating reliabilities of intra-individual variability indicators. Multivariate Behavior Research, 47, 1–31. doi: 10.1080/00273171.2012.715842

Wang L, Hamaker E, & Bergeman CS (2012). Investigating inter-individual differences in intra-individual variability. Psychological Methods, 17(4), 567–581. doi: 10.1037/a0029317 [PubMed: 22924600]

Wang L, & Maxwell SE (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. Psychological Methods, 20, 63–83. doi: 10.1037/met0000030 [PubMed: 25822206]

Watson D, Clark LA, & Tellegen A (1988). Development and validation of brief measure of positive and negative affect: the panas scales. Journal of Personality and Social Psychology, 54(6), 1063–1070. [PubMed: 3397865]

Watson D, Wiese D, Vaidya J, & Tellegen A (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. Journal of Personality and Social Psychology, 76(5), 820–838.

Zautra A, Smith B, Affleck G, & Tennen H (2001). Examinations of chronic pain and affect relationships: Applications of a dynamic model of affect. Journal of Consulting and Clinical Psychology, 69(5), 786–795. [PubMed: 11680555]

Zautra P, Potter A, & Reich J (1997). The independence of affects is context-dependent: An integrative model of the relationship between positive and negative affect. Annual Review of Gerontology and Geriatrics, 17, 75–103.

Zhang Q, & Wang L (2014). Aggregating and testing intra-individual correlations: Methods and comparisons. Multivariate Behavioral Research, 49, 130–148. doi: 10.1080/00273171.2013.870877 [PubMed: 26741173]

Zhang Q, Wang L, & Bergeman CS (2018). Multilevel autoregressive mediation models: Specification, estimation, and applications. Psychological Methods, 23(2), 278–297. doi: 10.1037/met0000161 [PubMed: 29172610]

Table 1:

Research questions, estimands, and standardization approaches.

Example research questions	Estimand (Target of estimation)	Standardization approaches	Modeling assumptions	Strengths	Limitations
The number of points that the "average" person's daily mood score will change when he/she has a 1 unit higher score on daily stress.	Average WP effect; Unstandardized WP coefficient.	Standardization not done; Person-mean centering is useful for disaggregating WP and BP effects.	Regular MLM assumptions apply to unstandardized coefficients.	Raw scales and units are kept. Easy to obtain estimates.	Raw coefficients may not be easy to interpret or be suitable for comparing the strengths of the effects.
The number of WP SDs that the "average" person's daily mood will change when his or her daily stress increases 1 WP SD.	Average WP relation or association; A standardized coefficient.	WP Standardization; 1. WP standardize variables before MLM (M_{PS}); 2. WP standardize random-effects estimates after MLM (M_{EB}); 3. WP standardize random-effects estimates during MLM.	1. Regular MLM assumptions apply to standardized coefficients. 2 & 3. Regular MLM assumptions apply to Unstandardized coefficients.	May ease interpretation and aid evaluating the strengths of WP relations. 1 &2 are easy to implement. 3 can be implemented with Bayesian methods (e.g., Mplus 8, BUGs).	Raw scales or units are not kept but the standardized coefficients can be useful and meaningful for comparing the strengths of the WP relations across different sets of variables.
The number of grand SDs that the "average" person's daily mood will change when his or her daily stress increases 1 grand SD.	Generally not the average WP relation *; A standardized coefficient.	Global Standardization; Globally standardize WP fixed-effects estimates after MLM (M_{G1}, M_{G2}, or M_{G3}).	Regular MLM assumptions apply to unstandardized coefficients.	May ease interpretation. Easy to obtain estimates.	Raw scales or units are not kept but the standardized coefficients may be useful and meaningful. But grand SD is a conflated combination of the average WP SD and BP variation in WP SDs. The research question may be less interesting to researchers.

Note: WP: within-person; BP: between-person. Regular multilevel modeling (MLM) assumptions include, for example, the random coefficients (standardized or unstandardized, depending on the modeling approach) are normally distributed when a normal-theory-based estimation approach is used. M_{G1}: the person-mean centering (P-C) model in Eq (2) followed by global standardization in Eq (5); M_{G2}: the P-C model in Eq (2) followed by global standardization in Eq (6); M_{G3}: the P-C model in Eq (4) followed by global standardization in Eq (7); M_{PS}: the P-S approach in Eq (9); M_{EB}: the P-C model in Eq (2) with EB standardization.

* Our statistical consistency derivation results (summarized in the derivation section) revealed this.

Table 2:

Derived population correlations of different kinds of stacked long data when σ_X, σ_Y, X^{PS}, and Y^{PS} follow a joint multivariate normal distribution.

Data Type	Derived population correlation	Involved parameters
Raw data (x_{it} and y_{it})	$\rho_{X,Y} = \frac{\left(\mu_{\sigma X}\mu_{\sigma Y} + \sigma_{\sigma X,\sigma Y}\right)\mu_{\rho w} + \sigma_{\mu X}\sigma_{\mu Y}\rho_b}{\sqrt{\left(\mu^2_{\sigma X} + \sigma^2_{\sigma X} + \sigma^2_{\mu X}\right)\left(\mu^2_{\sigma Y} + \sigma^2_{\sigma Y} + \sigma^2_{\mu Y}\right)}}$	Average within-person correlation ($\mu_{\rho w}$) Between-person correlation (ρ_b) Mean and SD of the person SDs ($\mu_{\sigma X}$, $\mu_{\sigma Y}$, $\sigma_{\sigma X}$, $\sigma_{\sigma Y}$) Covariance of the person SDs ($\sigma_{\sigma X,\sigma Y}$) SD of the person means ($\sigma_{\mu X}$, $\sigma_{\mu Y}$)
P-C data (x^{PC}_{it} and y^{PC}_{it})	$\rho_{CX,CY} = \frac{\left(\mu_{\sigma X}\mu_{\sigma Y} + \sigma_{\sigma X,\sigma Y}\right)\mu_{\rho w}}{\sqrt{\left(\mu^2_{\sigma X} + \sigma^2_{\sigma X}\right)\left(\mu^2_{\sigma Y} + \sigma^2_{\sigma Y}\right)}}$	Average within-person correlation ($\mu_{\rho w}$) Mean and SD of the person SDs ($\mu_{\sigma X}$, $\mu_{\sigma Y}$, $\sigma_{\sigma X}$, $\sigma_{\sigma Y}$) Covariance of the person SDs ($\sigma_{\sigma X,\sigma Y}$)
P-S data (x^{PS}_{it} and y^{PS}_{it})	$\rho_{WX,WY} = \mu_{\rho w}$	Average within-person correlation ($\mu_{\rho w}$)

Table 3:

Simulation results: Biases and coverage rates of $\mu_{\rho W}$ (average within-person relation) estimates when $\mu_{\rho W} = 0$, $N = 100$, and one predictor is included.

	Individual differences in within-person relations exist						No individual differences in within-person relations				
T	**Bias**	**Coverage rates (%)**					**Bias**	**Coverage rates (%)**			
	All 5 methods	M_{G1}	M_{G2}	M_{G3}	M_{PS}	M_{EB}	**All 4 methods**	M_{G1}	M_{G2}	M_{G3}	M_{PS}
Between-person differences in within-person standard deviations of X and Y exist											
5	0.00	94.5	94.5	94.9	95.4	***34.2***	0.00	91.1	91.1	***87.7***	91.8
10	0.00	94.1	94.1	93.9	94.4	***44.6***	0.00	93.9	93.9	92.2	94.0
20	0.00	94.0	94.0	93.7	94.1	***63.4***	0.00	93.2	93.2	92.4	94.6
30	0.00	93.7	93.7	93.9	93.8	***75.8***	0.00	94.6	94.6	94.3	95.9
56	0.00	95.3	95.3	95.6	94.7	***85.7***	0.00	93.4	93.4	93.3	94.9
100	0.00	95.3	95.3	95.4	94.6	91.8	0.00	93.3	93.3	93.3	96.2
Between-person differences in within-person standard deviations exist for Y but not for X											
5	0.00	93.2	93.2	93.9	93.2	***31.0***	0.00	95.5	95.5	93.7	92.0
10	0.00	95.1	95.1	95.2	95.9	***48.0***	0.00	95.8	95.8	94.8	94.6
20	0.00	94.4	94.4	94.4	94.3	***67.9***	0.00	94.9	94.9	94.6	94.1
30	0.00	94.9	94.9	95.0	94.3	***74.8***	0.00	94.9	94.9	94.6	95.9
56	0.00	95.5	95.5	95.5	94.6	***87.1***	0.00	94.3	94.3	94.3	93.2
100	0.00	95.2	95.2	95.2	95.3	92.0	0.00	95.7	95.7	95.6	95.5
Between-person differences in within-person standard deviations of X and Y do not exist											
5	0.00	94.0	94.0	94.1	95.3	***32.9***	0.00	95.7	95.7	92.8	92.5
10	0.00	94.5	94.5	94.5	94.4	***53.6***	0.00	95.6	95.6	94.5	94.7
20	0.00	94.3	94.3	94.2	93.7	***70.7***	0.00	95.9	95.9	95.4	96.0
30	0.00	95.5	95.5	95.6	95.4	***83.0***	0.00	95.1	95.1	94.5	94.6
56	0.00	93.7	93.7	93.7	94.1	***86.1***	0.00	95.4	95.4	95.2	94.9
100	0.00	94.3	94.3	94.3	94.2	***90.8***	0.00	95.6	95.6	95.6	96.1

Note: M_{G1}: the person-mean centering (P-C) model in Eq (2) followed by global standardization in Eq (5); M_{G2}: the P-C model in Eq (2) followed by global standardization in Eq (6); M_{G3}: the P-C model in Eq (4) followed by global standardization in Eq (7); M_{PS}: the P-S standardization approach in Eq (9); M_{EB}: the P-C model in Eq (2) with EB standardization.

Table 4:

Simulation results: Relative biases and coverage rates of $\mu_{\rho w}$ (average within-person relation) estimates when $\mu_{\rho w} = -0.5$, $N = 100$, and one predictor is included.

	Individual differences in within-person relations exist										No individual differences in within-person relations							
T	Relative bias (%)					Coverage rates (%)					Relative bias (%)				Coverage rates (%)			
	M_{G1}	M_{G2}	M_{G3}	M_{PS}	M_{EB}	M_{G1}	M_{G2}	M_{G3}	M_{PS}	M_{EB}	M_{G1}	M_{G2}	M_{G3}	M_{PS}	M_{G1}	M_{G2}	M_{G3}	M_{PS}
Between-person differences in within-person standard deviations of X and Y exist																		
5	2.3	***−49.3***	9.6	−7.7	−8.8	93.0	***0.1***	91.2	***90.1***	***64.1***	***−15.6***	***−58.1***	***−15.6***	−9.5	***60.6***	***0.0***	***51.3***	***76.3***
10	***13.2***	***−40.3***	***16.5***	−3.4	***−11.3***	***79.3***	***0.4***	***73.3***	92.6	***40.7***	***−15.2***	***−55.3***	***−15.2***	−3.7	***28.3***	***0.0***	***23.9***	91.9
20	***28.0***	***−30.7***	***29.5***	−1.5	−8.7	***32.1***	***3.9***	***28.9***	92.6	***52.9***	***−15.4***	***−54.1***	***−15.4***	−1.5	***3.8***	***0.0***	***3.4***	94.6
30	***37.2***	***−25.1***	***38.1***	−1.0	−6.7	***12.1***	***11.6***	***11.1***	94.0	***65.6***	***−15.4***	***−53.8***	***−15.4***	−0.8	***1.0***	***0.0***	***0.8***	95.9
56	***50.6***	***−17.1***	***50.9***	−0.8	−4.5	***1.1***	***38.0***	***1.0***	94.7	***81.7***	***−15.2***	***−53.3***	***−15.2***	−0.2	***0.0***	***0.0***	***0.0***	96.6
100	***61.1***	***−11.2***	***61.2***	−0.2	−2.4	***0.2***	***59.2***	***0.2***	94.9	***90.1***	***−15.3***	***−53.4***	***−15.3***	0.1	***0.0***	***0.0***	***0.0***	97.5
Between-person differences in within-person standard deviations exist for Y but not for X																		
5	−4.4	***−52.5***	−4.3	−7.7	***13.4***	95.5	***0.0***	95.7	***88.3***	***46.9***	−3.8	***−52.1***	−3.8	−9.6	96.6	***0.0***	94.4	***77.7***
10	−4.0	***−49.4***	−3.9	−3.1	5.5	94.4	***0.0***	94.4	***90.9***	***62.1***	−3.7	***−49.2***	−3.7	−3.8	93.2	***0.0***	91.4	***89.1***
20	−4.1	***−48.0***	−4.1	−1.5	2.7	93.3	***0.0***	93.5	94.5	***79.8***	−3.8	***−47.9***	−3.8	−1.6	***87.2***	***0.0***	***85.9***	94.8
30	−4.4	***−47.8***	−4.4	−1.2	1.5	***90.9***	***0.0***	***90.7***	93.9	***85.4***	−3.8	***−47.4***	−3.8	−0.9	***78.3***	***0.0***	***77.4***	96.0
56	−4.3	***−47.5***	−4.3	−0.7	0.8	***90.4***	***0.0***	***90.4***	93.9	***90.6***	−3.6	***−46.9***	−3.6	−0.1	***61.4***	***0.0***	***60.7***	97.1
100	−4.0	***−46.9***	−4.0	−0.2	0.7	***90.4***	***0.0***	***90.4***	93.8	92.4	−3.6	***−46.9***	−3.6	0.2	***37.0***	***0.0***	***36.3***	98.0
Between-person differences in within-person standard deviations of X and Y do not exist																		
5	−0.1	***−51.7***	−0.1	−7.6	7.1	96.4	***0.0***	96.4	***86.5***	***56.1***	0.8	***−51.4***	0.8	−8.9	98.0	***0.0***	96.0	***79.4***
10	0.2	***−48.7***	0.2	−3.2	1.1	95.6	***0.0***	95.9	92.7	***62.5***	0.4	***−48.6***	0.4	−3.8	97.5	***0.0***	96.8	***90.1***
20	0.2	***−47.4***	0.1	−1.4	−0.2	96.1	***0.0***	96.1	95.3	***80.5***	0.5	***−47.1***	0.5	−1.5	97.1	***0.0***	96.9	94.0
30	−0.2	***−47.2***	−0.2	−1.3	−0.7	95.1	***0.0***	95.0	92.9	***82.2***	0.6	***−46.7***	0.6	−0.7	97.4	***0.0***	97.3	94.9
56	−0.2	***−46.7***	−0.2	−0.7	−0.6	96.1	***0.0***	96.1	95.2	***88.5***	0.4	***−46.4***	0.4	−0.3	98.2	***0.0***	97.9	96.7
100	−0.1	***−46.4***	−0.1	−0.3	−0.3	92.3	***0.0***	92.3	92.3	***90.0***	0.5	***−46.0***	0.5	0.1	97.9	***0.0***	97.9	97.4

Note: M_{G1}: the person-mean centering (P-C) model in Eq (2) followed by global standardization in Eq (5); M_{G2}: the P-C model in Eq (2) followed by global standardization in Eq (6); M_{G3}: the P-C model in Eq (4) followed by global standardization in Eq (7); M_{PS}: the P-S standardization approach in Eq (9); M_{EB}: the P-C model in Eq (2) with EB standardization.

Table 5:

Simulation results from the models with two predictors included ($N = 100$).

T	X_1 on Y (True value = 0)					X_2 on Y (True value = −0.4)									
	Coverage rates (%)					Relative bias (%)					Coverage rates (%)				
	M_{G1}	M_{G2}	M_{G3}	M_{PS}	M_{EB}	M_{G1}	M_{G2}	M_{G3}	M_{PS}	M_{EB}	M_{G1}	M_{G2}	M_{G3}	M_{PS}	M_{EB}
Between-person differences in within-person standard deviations of X and Y exist															
Equal number of assessments across individuals															
5	94.4	93.9	95.4	95.2	***42.2***	−7.5	***−54.1***	−2.9	***−13.0***	3.2	91.6	***0.0***	94.5	***87.5***	***81.2***
10	93.3	93.2	94.4	94.4	***59.8***	−2.5	***−48.8***	−0.6	−7.2	−4.3	94.6	***0.0***	95.8	93.3	***76.7***
20	93.8	93.9	94.1	95.2	***80.6***	4.5	***−43.5***	5.1	−3.0	−3.9	94.3	***0.1***	94.0	93.3	***84.0***
30	95.2	95.0	95.2	95.7	***87.4***	8.0	***−40.9***	8.3	−2.3	−3.3	93.2	***0.4***	93.0	95.0	***89.8***
56	94.7	95.0	94.7	95.1	91.1	***12.9***	***−38.0***	***13.0***	−0.6	−1.4	***87.5***	***0.8***	***87.4***	94.7	94.5
100	95.2	95.7	95.1	94.1	92.2	***14.7***	***−36.6***	***14.8***	−0.8	−1.3	***85.5***	***3.5***	***85.5***	94.2	95.7
Unequal number of assessments across individuals															
5	92.6	92.6	93.1	94.1	***47.4***	−6.9	***−53.9***	−2.3	***−13.1***	***79.4***	92.6	***0.1***	94.4	***86.6***	***91.2***
10	93.4	93.4	93.5	94.3	***57.3***	−1.5	***−48.2***	0.2	−6.6	0.1	94.0	***0.0***	94.0	91.9	***80.8***
20	94.2	94.2	94.2	94.2	***79.3***	4.2	***−43.7***	4.8	−3.8	−4.2	94.9	***0.3***	94.4	92.3	***83.2***
30	93.7	93.7	93.9	95.0	***87.8***	8.2	***−41.0***	8.6	−2.1	−3.1	*91.8*	***0.7***	91.6	94.4	***89.6***
56	95.0	95.0	94.9	94.7	91.7	***13.4***	***−37.7***	***13.5***	−0.5	−1.4	***86.9***	***1.6***	***86.8***	95.1	95.2
100	93.5	93.5	93.5	93.9	92.7	***14.7***	***−36.6***	***14.8***	−0.5	−1.1	***84.7***	***2.8***	***84.6***	94.3	94.2
Between-person differences in within-person standard deviations of X and Y do not exist															
Equal number of assessments across individuals															
5	93.8	93.7	94.4	93.7	***45.1***	−4.4	***−53.9***	−3.6	***−13.2***	4.2	92.9	***0.0***	93.7	***87.5***	***82.3***
10	94.6	95.2	94.7	94.1	***66.9***	−3.3	***−50.4***	−3.0	−7.0	−2.0	95.5	***0.0***	95.5	92.8	***81.0***
20	95.0	93.6	94.7	94.4	***79.5***	−1.7	***−48.4***	−1.6	−3.3	−1.8	94.9	***0.0***	94.9	94.4	***88.0***
30	93.4	93.0	93.5	93.6	***85.5***	−1.8	***−47.9***	−1.7	−2.8	−2.1	94.7	***0.0***	94.7	93.6	92.0
56	93.2	93.7	93.2	93.3	***89.9***	−0.9	***−47.0***	−0.8	−1.3	−1.1	94.4	***0.0***	94.4	94.9	94.2
100	94.2	94.0	94.2	94.0	92.0	−0.1	***−46.5***	−0.1	−0.4	−0.3	94.2	***0.0***	94.2	94.1	95.6
Unequal number of assessments across individuals															
5	94.2	94.6	94.6	94.3	***45.1***	−3.81	***−53.57***	−3.14	***−12.83***	***47.45***	95.0	***0.0***	95.2	***89.0***	**92.3**
10	94.0	94.3	93.9	94.0	***64.6***	−3.21	***−50.36***	−3.02	−7.23	−0.43	94.6	***0.0***	94.6	***90.6***	***82.1***
20	93.7	94.5	93.8	93.5	***81.6***	−2.11	***−48.48***	−2.03	−3.78	−2.15	94.3	***0.0***	94.4	93.8	***86.7***
30	95.2	95.2	95.3	95.2	***87.1***	−1.46	***−47.74***	−1.42	−2.42	−1.68	95.8	***0.0***	96.0	95.4	91.6
56	93.5	92.9	93.5	94.0	***88.1***	−1.37	***−47.28***	−1.36	−1.91	−1.69	94.7	***0.0***	94.8	94.8	94.3
100	93.7	94.3	93.7	94.4	92.5	−0.08	***−46.41***	−0.07	−0.32	−0.25	94.3	***0.0***	94.3	94.4	95.8

Note: M_{G1}: the person-mean centering (P-C) model in Eq (2) followed by global standardization in Eq (5); M_{G2}: the P-C model in Eq (2) followed by global standardization in Eq (6); M_{G3}: the P-C model in Eq (4) followed by global standardization in Eq (7); M_{PS}: the P-S approach in Eq (9); M_{EB}: the P-C model in Eq (2) with EB standardization. The empirical biases from the 5 methods were all between −.02 and .00 when the true value is 0 and thus are not listed to save space.

Table 6:

A summary of the evaluated standardization approaches in estimating and inferring fixed-effects within-person relations (e.g., $\mu_{\rho w}$).

	Model	Data used	Standardization method	Consistency	Performance with finite samples	
					Estimation accuracy	Coverage rates
M_{G1}	Eq 2	P-C	$\hat{\gamma}_{10}^{G1*} = \hat{\gamma}_{10}^{C1}\frac{s_{cx}}{s_{cy}}$ Global-1 standardization	Generally not consistent when both the average WP relation is not 0 and BP differences in person SDs exist	Generally not accurate when both the average WP relation is not 0 and BP differences in person SDs exist	Generally not satisfactory when both the average WP relation is not 0 and BP differences in person SDs exist
M_{G2}	Eq 2	P-C	$\hat{\gamma}_{10}^{G2*} = \hat{\gamma}_{10}^{C1}\frac{s_{cx}}{s_{y}}$ Global-2 standardization	Generally not consistent when the average WP relation is not 0	Generally not accurate when the average WP relation is not 0	Generally not satisfactory when the average WP relation is not 0
M_{G3}	Eq 4	P-C	$\hat{\gamma}_{10}^{G3*} = \hat{\gamma}_{10}^{C2}\frac{s_{cx}}{s_{cy}}$ Global-3 standardization	Generally not consistent when both the average WP relation is not 0 and BP differences in person SDs exist	Generally not accurate when both the average WP relation is not 0 and BP differences in person SDs exist	Generally not satisfactory when both the average WP relation is not 0 and BP differences in person SDs exist
M_{PS}	Eq 9	P-S	$\hat{\gamma}_{10}^{PS*} = \hat{\gamma}_{10}^{PS}$ Person-level P-S standardization	Consistent	Generally ignorable biases when $T \geq 10$ and $N \geq 50$ Larger T and N, smaller biases	Generally satisfactory when $T \geq 30$ and $N \geq 50$
M_{EB}	Eq 2	P-C	$\hat{\gamma}_{10}^{EB*} = \frac{1}{N}\sum \hat{\gamma}_{1i}^{C1}\frac{s_{cx,i}}{s_{cy,i}}\hat{\gamma}_{1i}^{C1}$: Empirical Bayes estimate; Person-level EB standardization	Consistent	Generally ignorable biases Larger T and N, smaller biases	Satisfactory or approaching satisfactory when T and N are large enough

Note: P-C: person-mean centered; P-S: person-mean-SD standardized.

Table 7:

Descriptive statistics of the stacked long data.

	Mean	SD	Skewness	Kurtosis	Correlation Matrix								
					PA	NA	Stress	PC-PA	PC-NA	PC-Stress	PS-PA	PS-NA	PS-Stress
PA	29.46	9.96	0.07	0.54	1.00								
NA	12.91	5.36	2.21	6.96	**−0.20**	1.00							
Stress	18.72	6.92	0.73	1.72	**−0.45**	**0.72**	1.00						
PC-PA	0	5.41	0.01	0.77	0.54	−0.15	−0.23	1.00					
PC-NA	0	3.13	1.94	10.35	−0.14	0.59	0.34	**−0.26**	1.00				
PC-Stress	0	3.62	0.71	2.44	−0.24	0.38	0.53	**−0.44**	**0.65**	1.00			
PS-PA	0	.99	−0.04	0.42	0.52	−0.14	−0.21	0.96	−0.24	−0.41	1.00		
PS-NA	0	.99	2.12	7.10	−0.13	0.49	0.30	−0.23	0.84	0.57	**−0.23**	1.00	
PS-Stress	0	.99	0.59	1.29	−0.23	0.32	0.49	−0.42	0.55	0.94	**−0.40**	**0.55**	1.00

Note: PA: positive affect; NA: negative affect; PC-PA: person-mean centered positive affect; PC-NA: person mean centered negative affect; PC-Stress: person mean centered Stress; PS-PA: person-mean-SD (P-S) standardized positive affect; PS-NA: P-S standardized negative affect; PS-stress: P-S standardized Stress.

Table 8:

Summary statistics of the person mean and standard deviation variables.

	Mean	SD	Skewness	Kurtosis	Covariance (Correlation) Matrix					
					MPA	MNA	MStress	SPA	SNA	SStress
MPA	29.46	8.39	0.26	0.74	70.39					
MNA	12.91	4.36	2.26	6.70	−6.43 (−.18)	18.99				
MStress	18.72	5.92	0.81	2.15	−22.28 (−.45)	19.62 (.76)	35.00			
SPA	5.26	1.53	−0.13	−0.54	−2.70 (−.21)	0.86 (.13)	0.78 (.09)	2.34		
SNA	2.68	1.72	1.11	1.22	−0.64 (−.04)	4.35 (.58)	3.74 (.37)	0.57 (.22)	2.96	
SStress	3.44	1.29	0.78	0.27	0.48 (.04)	2.37 (.42)	1.86 (.24)	0.64 (.33)	1.71 (.77)	1.67

Note: MPA: person means of positive affect; MNA: person means of negative affect; MStress: person means of stress; SPA: person standard deviations of positive affect; SNA: person standard deviations of negative affect; SStress: person standard deviations of stress.

Table 9:

Estimates of the fixed-effects within-person relations from various multilevel models.

	M_{G1} Eqs 2 & 5 P-C with global-1	M_{G2} Eqs 2 & 6 P-C with global-2	M_{EB} Eq 2 & EB P-C with EB	M_{G3} Eqs 4 & 7 P-C with global-3	M_{PS} Eq 9 P-S
Predictor: Negative affect; Outcome: Positive affect					
Under the homogeneous within-person relation assumption					
$\hat{\gamma}^*_{10}$	−.26 (.013)	−.14 (.007)	na	−.26 (.013)	−.23 (.013)
Under the heterogeneous within-person relation assumption					
$\hat{\gamma}^*_{10}$	−.29 (.034)	−.16 (.018)	−.24 (.022)	−.29 (.034)	−.23 (.028)
$\hat{Var}(\hat{\gamma}^*_{1i})$	na	na	.048 (na)	na	.062 (.011)
Predictor: Stress; Outcome: Positive affect					
Under the homogeneous within-person relation assumption					
$\hat{\gamma}^*_{10}$	−.44 (.012)	−.24 (.007)	na	−.44 (.012)	−.40 (.012)
Under the heterogeneous within-person relation assumption					
$\hat{\gamma}^*_{10}$	−.46 (.035)	−.25 (.019)	−.41 (.022)	−.46 (.035)	−.40 (.027)
$\hat{Var}(\gamma^*_{1i})$	na	na	.052 (na)	na	.063 (.011)
Predictors: Negative affect and Stress; Outcome: Positive affect					
Under the homogeneous within-person relation assumption					
$\hat{\gamma}^*_{1,NA}$	.04 (.016)	.02 (.009)	na	.04 (.016)	−.007 (.015)
$\hat{\gamma}^*_{1,Stress}$	−.46 (.016)	−.25 (.009)	na	−.46 (.016)	−.40 (.015)
the heterogeneous within-person relation assumption					
$\hat{\gamma}^*_{1,NA}$	.05 (.039)	.03 (.021)	.006 (.025)	.05 (.039)	.001 (.030)
$\hat{\gamma}^*_{1,Stress}$	−.47 (.042)	−.25 (.023)	−.42 (.026)	−.47 (.042)	−.40 (.032)
$\hat{Var}(\gamma^*_{1i,NA})$	na	na	.062 (na)	na	.077 (.014)
$\hat{Var}(\gamma^*_{2i,Stress})$	na	na	.069 (na)	na	.067 (.013)

Note: The values outside the parentheses are point estimates and those inside the parentheses are standard error estimates. The variance estimates were from the restricted maximum likelihood method. P-C: person-mean centering; Global-1, global-2 and global-3: s_{CY}, s_Y, and s_{CY} are used for global standardization respectively; P-C with EB: person-mean centering with Empirical Bayes person-level standardization; P-S: person-mean-SD standardization.