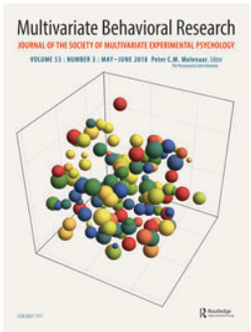

Multivariate Behavioral Research

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

The Power to Detect and Predict Individual Differences in Intra-Individual Variability Using the Mixed-Effects Location-Scale Model

Ryan W. Walters, Lesa Hoffman & Jonathan Templin

To cite this article: Ryan W. Walters, Lesa Hoffman & Jonathan Templin (2018) The Power to Detect and Predict Individual Differences in Intra-Individual Variability Using the Mixed-Effects Location-Scale Model, Multivariate Behavioral Research, 53:3, 360-374, DOI: 10.1080/00273171.2018.1449628

To link to this article: https://doi.org/10.1080/00273171.2018.1449628

MULTIVARIATE BEHAVIORAL RESEARCH
2018, VOL. 53, NO. 3, 360–374
https://doi.org/10.1080/00273171.2018.1449628

Routledge
Taylor & Francis Group

The Power to Detect and Predict Individual Differences in Intra-Individual Variability Using the Mixed-Effects Location-Scale Model

Ryan W. Walters[a], Lesa Hoffman[b], and Jonathan Templin[b]

[a]Creighton University; [b]University of Kansas

ABSTRACT

Our goal is to provide empirical scientists with practical tools and advice with which to test hypotheses related to individual differences in intra-individual variability using the mixed-effects location-scale model. To that end, we evaluate Type I error rates and power to detect and predict individual differences in intra-individual variability using this model and provide empirically-based guidelines for building scale models that include random and/or systematically-varying fixed effects. We also provide two power simulation programs that allow researchers to conduct *a priori* empirical power analyses. Our results aligned with statistical power theory, in that, greater power was observed for designs with more individuals, more repeated occasions, greater proportions of variance available to be explained, and larger effect sizes. In addition, our results indicated that Type I error rates were acceptable in situations when individual differences in intra-individual variability were not initially detectable as well as when the scale-model individual-level predictor explained all initially detectable individual differences in intra-individual variability. We conclude our paper by providing study design and model building advice for those interested in using the mixed-effects location-scale model in practice.

KEYWORDS

Mixed-effects location-scale model; intra-individual variability; individual differences; statistical power; systematically-varying effects

Hypotheses about physical and psychological processes often focus on individual differences in the mean level of a repeatedly-measured outcome. However, researchers in diverse areas have begun to examine how individual differences in *intra-individual variability* across repeated measurements—also known as residual variability, inconsistency, or instability—may reveal novel findings (see Hultsch, Strauss, Hunter, & MacDonald, 2008). For example, individual differences in intra-individual variability have been associated with cognitive decline (Dixon et al., 2007), mild dementia (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000), major depressive and bipolar disorders (Gallagher et al., 2015; Schneider et al., 2012), nicotine tolerance (Hedeker & Mermelstein, 2007; Hedeker, Mermelstein, Berbaum, & Campbell, 2009), and impending death (MacDonald, Hultsch, & Dixon, 2003).

Although multistage methods are also available to study individual differences in intra-individual variability (Hultsch et al., 2008), our present focus is the single-stage *mixed-effects location-scale model* used to concurrently quantify and predict individual differences in mean level and intra-individual variability via random effects (see Hedeker, Mermelstein, & Demirtas, 2008 who expanded on the work of Cleveland, Denby, & Liu, 2000). For example, Hedeker et al. (2008) estimated the mixed-effects location-scale model to evaluate ecological momentary assessments of positive and negative affect collected from 461 adolescents over a seven-day study period. Although boys and girls averaged similar mean levels of positive affect, girls had significantly more variable (i.e., less consistent) positive affect across occasions compared to boys. The mixed-effects location-scale model has also been applied to eye movement in schizophrenia (Lee & Noh, 2012), moderate-to-vigorous physical activity in children (Dunton et al., 2013), resilience to stressful events (Rast, Hofer, & Sparks, 2012), and sleep efficiency (Ong, Hedeker, Wyatt, & Manber, 2016).

These examples highlight the utility of the mixed-effects location-scale model to simultaneously test mean- and variability-related hypotheses, but this model is estimated infrequently in practice. A portion of the blame may fall on the lack of readily available software to estimate scale-model random effects that quantify individual differences in intra-individual variability as it is not currently possible to estimate scale-model random effects in the MIXED procedures in SAS or SPSS (although software has been developed in a frequentist framework by Hedeker & Nordgren, 2013, and in a Bayesian framework by Kupar, Li, Blood, & Hedeker, 2015, Rast et al., 2012, and Wang, Hamaker, & Bergeman, 2012).

CONTACT Ryan W. Walters ryanwalters@creighton.edu 7710 Mercy Road, Suite 502, Room 51266, Omaha 68124, Nebraska, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/HMBR.
Supplemental data for this article can be accessed on the publisher's website.

Beyond software, however, we believe that the infrequent use of the mixed-effects location-scale model may be due to insufficient instruction regarding sampling designs for testing individual differences in intra-individual variability and a lack of best practices for including predictors of intra-individual variability. The purpose of our study was to address these two perceived gaps in the literature based on a two-level sampling design in which the repeated measurement of a level-1 continuous outcome is nested within level-2 individuals. We were particularly interested in outcome data that *fluctuate* across occasions (i.e., no change or growth), as would be expected from non-intervention, observational studies (see Dunton et al., 2013; Hedeker et al., 2008; Hedeker et al., 2012; and Rast & Zimprich, 2011).

First, because little is currently known about the interplay between research design and statistical power within the context of quantifying individual differences in intra-individual variation via scale-model random effects, we sought to investigate how study design characteristics moderate the Type I error rates and the power to detect and predict individual differences in intra-individual variability. We also created two simulation programs by which researchers can conduct *a priori* empirical power analyses (available in a supplementary appendix).

Second, we sought to define best practices for including individual-level predictors of intra-individual variability in the mixed-effects location-scale model. Although Leckie, French, Charlton, and Browne (2014) have shown that erroneously omitting scale-model random effects has dire consequences for Type I error rates for individual-level predictors of intra-individual variability, the proper course of action remains unclear for including individual-level predictors when scale-model random effects are *not* initially detectable. Can fixed effects of individual-level scale-model predictors still be tested accurately via a *heterogenous variance model* without scale-model random effects (see Davidian & Giltinan, 1995; Hoffman, 2007; Raudenbush & Bryk, 2002)? The same question is relevant when a scale-model random effect was initially significant, but was reduced to near zero after being fully explained by individual-level predictors. In either case, the individual differences in intra-individual variability are *systematically varying* across individuals—not as randomly varying, but varying solely as a function of known predictors.

To our knowledge, the appropriateness of testing systematically-varying scale-model effects of individual-level predictors has not been examined empirically, and thus it is unknown to what extent including them is acceptable given nonsignificant scale-model random effects. Therefore, after evaluating Type I error rates for a scale-model individual-level predictor using replications in which significant scale-model random intercept variance was erroneously omitted (a replication of Leckie et al., 2014), we determined the Type I error rate for a scale-model individual-level predictor using replications in which scale-model random intercept variance was not detected initially and also using replications in which significant scale-model random intercept variance was detected initially but became nonsignificantly greater than zero after being explained by the individual-level predictor.

The remainder of our paper is organized as follows. After introducing the mixed-effects location-scale model, we describe our simulation studies and results with respect to the Type I error rates and power to detect individual differences in intra-individual variability, followed by the power to predict those differences with an individual-level predictor (after also conducting a preliminary inquiry to confirm the behavior of a pseudo-R^2 statistic for describing effect size of scale-model individual-level predictors). We then detail Type I error rates for systematically-varying scale-model fixed effects. Finally, we provide advice for planning study designs and subsequent model building when utilizing the mixed-effects location-scale model.

The mixed-effects location-scale model

Location model. For conditionally normally distributed outcomes, the *location model* is the traditional linear mixed-effects model for i individuals ($i = 1, 2, \ldots, N$ individuals) with j repeated occasions ($j = 1, 2, \ldots, n_i$ occasions) as shown in (1) below.

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \tag{1}$$

In (1), $\mathbf{Y}_i$ is an $n_i \times 1$ column vector of outcomes for the observations in individual i. $\mathbf{X}_i$ is an $n_i \times p$ matrix of p location-model occasion- and/or individual-level predictors for the observations in individual i. β is a $p \times 1$ column vector of location-model fixed effects for the intercept and the $p - 1$ predictors. In $\mathbf{X}_i$, we assume the individual-level mean for each occasion-level predictor has also been included to disaggregate its effects across levels to avoid a "smushed" effect (Hoffman, 2015). $\mathbf{Z}_i$ is an $n_i \times q$ matrix of q occasion-level predictors that have location-model random effects for the observations in individual i. $\mathbf{X}_i$ and $\mathbf{Z}_i$ have an $n_i \times 1$ column vector of ones in the first column for the location-model fixed and random intercept, respectively. $\mathbf{b}_i$ is a $q \times 1$ column vector of q location-model random effects for individual i. Finally, $\mathbf{e}_i$ is an $n_i \times 1$ column vector of residuals for observations in individual i.

This model generally assumes location-model random effects $\mathbf{b}_i$ are multivariate normally distributed with a mean vector of $\mathbf{0}$ and positive semi-definite $q \times q$ covariance matrix $\mathbf{G}$ whose random effect variances $\sigma^2_{b_q}$ and covariances σ_{b_q,b'_q} quantify inter-individual differences in the *location* of the outcome: $\mathbf{b}_i \sim \mathbf{N}_q(0, \mathbf{G})$. Further, residual values $\mathbf{e}_i$ are generally assumed multivariate normally distributed with a mean vector of $\mathbf{0}$ and positive semi-definite $n_i \times n_i$ covariance matrix $\mathbf{R}_i$ whose variances and covariances quantify intra-individual variability as $\mathbf{e}_i \sim \mathbf{N}_{n_i}(0, \mathbf{R}_i)$. Although many alternative covariance structures for $\mathbf{R}_i$ are available (e.g., autoregressive; see Littell, Pedergast, & Natarajan, 2000), the residual variance σ^2_e that quantifies intra-individual variability in $\mathbf{R}_i$ is most commonly constrained to be constant (or homogeneous) with the residual values assumed conditionally independent given the random effects: $\mathbf{R}_i = \sigma^2_e \mathbf{I}_{n_i}$, in which $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix for the observations in individual *i*.

Scale model for the residual variance. The mixed-effects location-scale model relaxes the assumption of homogeneous intra-individual variability by allowing heterogeneity of variance of the level-1 residuals using the *scale model for the residual variance* shown in (2) below, which employs the (natural) log link to ensure positive predicted residual variances (Aitkin, 1987; Foulley & Quaas, 1995; Harvey, 1976) and whose matrices share conceptual overlap with the location model shown in (1).

$$\boldsymbol{\sigma}^2_{e_i} = \exp\left(\mathbf{W}_i\tau + \mathbf{A}_i\mathbf{t}_i\right) \tag{2}$$

In (2), $\boldsymbol{\sigma}^2_{e_i}$ is an $n_i \times 1$ column vector of residual variances for the observations in individual *i*. $\mathbf{W}_i$ is an $n_i \times s$ matrix of *s* scale-model occasion- and/or individual-level predictors for the observations in individual *i*. τ is a $s \times 1$ column vector for the scale-model fixed effects for the intercept and the $s - 1$ predictors. $\mathbf{A}_i$ is an $n_i \times a$ matrix of *a* occasion-level predictors that have scale-model random effects for the observations in individual *i*. $\mathbf{t}_i$ is a $a \times 1$ column vector of *a* scale-model random effect coefficients for individual *i*. The matrices $\mathbf{W}_i$ and $\mathbf{A}_i$ include an $n_i \times 1$ column vector of ones in the first column for the scale-model fixed and random intercept, respectively. Further, scale-model occasion-level predictors in $\mathbf{W}_i$ predict differential variability across *occasions*, whereas scale-model individual-level predictors in $\mathbf{W}_i$ predict differential variability across *individuals*. In general, each unit increase in a scale-model predictor with a positive fixed effect will result in increased intra-individual variability indicating a more variable (or less consistent) occasion or individual. Note, the model does not require the same predictors to be included in both the location-model $\mathbf{X}_i$ and the scale-model equivalent $\mathbf{W}_i$.

Finally, the mixed-effects location-scale model generally assumes that scale-model random effects $\mathbf{t}_i$ are multivariate normally distributed with a mean vector of $\mathbf{0}$ whose variances $\sigma^2_{t_s}$ and covariances σ_{t_s,t'_s} are included alongside the location-model random effects variances and covariances in $\mathbf{G}$ defined above. Further, because independence of the location- and scale-model random effects is not required, these random effects can share some non-zero covariance or correlation. For example, a positive correlation between the location- and scale-model random intercepts indicates that individuals with greater than average mean levels of the outcome may also tend to average more variability in the outcome across occasions.

Visualizing the mixed-effects location-scale model. Figure 1 provides a visual depiction of an unconditional mixed-effects location-scale model that includes location- and scale-model fixed and random intercepts for two individuals each with 12 repeated occasions.

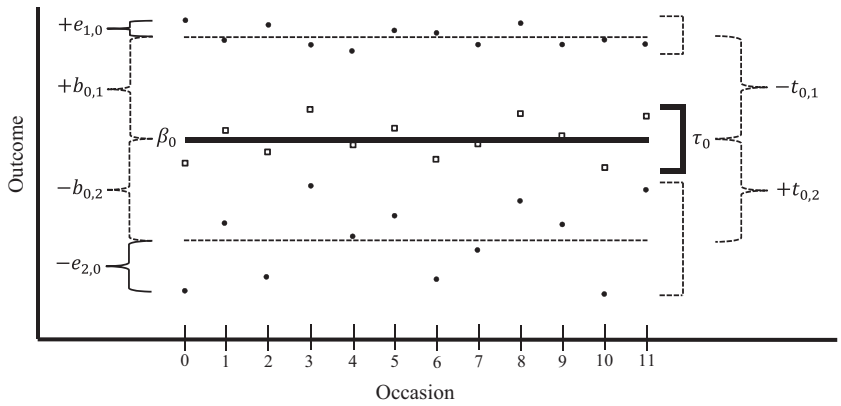

Figure 1. A visual depiction of an unconditional mixed-effects location-scale model that includes location- and scale-model fixed and random intercepts. β_0 = location-model fixed intercept. $b_{0,i}$ = location-model random intercept for individual *i*. $e_{i,j}$ = residual value for individual *i* at occasion *j*. τ_0 = scale-model fixed intercept. $t_{0,i}$ = scale-model random intercept for individual *i*.

Effects on the left side are specific to the location model, whereas effects on the right side are specific to the scale model for the residual variance. The filled circles represent the observed outcome values, the open squares represent the average of the filled circles.

Regarding the location model, the thick solid line is the location-model fixed intercept β_0 representing the sample average amount of the outcome across occasions; it fully defines the average trajectory across occasions given the absence of location-model fixed time effects. The two dashed lines represent individual-specific location-model intercepts that fully define each individual's trajectory given the absence of location-model random time effects. The deviation of each individual's trajectory from the average trajectory, indicated by the dashed curly brackets, is the location-model random intercept $b_{0,i}$ for individual i whose variance across individuals is quantified by the location-model random intercept variance $\sigma^2_{b_0}$. Further, the residual values $e_{i,j}$ represent the deviation of an observed outcome from the individual's predicted trajectory—the solid curly brackets indicate the residual value for individual i at the first occasion.

Regarding the scale-model for the residual variance, the square brackets approximate the (natural) log of the residual variances (i.e., natural log intra-individual variability), of which the thick solid square bracket indicates the scale-model fixed intercept τ_0 that represents the log of the sample average amount of residual variance, whereas the dashed square brackets represent log residual variance specific to individual i. The dashed curly brackets indicate differences in the size of the dashed square brackets relative to the thick solid square bracket and represent the scale-model random intercept $t_{0,i}$ whose variance across individuals is quantified by the scale-model random intercept variance $\sigma^2_{t_0}$.

For these data, although individual 1 (the top dashed line) has higher outcome values than average, as indicated by positive $b_{0,1}$, they have less intra-individual variability than average, as indicated by negative $t_{0,1}$. The opposite pattern is observed for individual 2 (the bottom dashed line). Together, these data indicate a negative covariance between the location- and scale-model random intercepts as individuals who have higher outcomes values than average tend to have lower intra-individual variability than average (and vice versa). Further, although no predictor fixed effects were included in the location model or scale model for the residual variance when creating this figure, in our study we will consider a scale-model fixed effect of an individual-level predictor, τ_1, that will serve to increase or decrease the intra-individual variability of the observed outcome values around an individual's dashed line depending on their predictor value.

Purpose of the current study

To summarize thus far, the mixed-effects location-scale model provides the opportunity to quantify and predict individual differences in the amount of an outcome via the location model shown in (1) as well as individual differences in intra-individual variability via the log-linked scale model for the residual variance shown in (2). To address two gaps in the literature specific to this model, we report two simulation studies. In the first study, we evaluated how sampling design and estimated model parameters affect Type I error rates and power to detect and predict random individual differences in intra-individual variability. To help researchers plan studies with adequate power for scale-model effects, we provide two simulation programs in SAS and R software (see supplementary appendix) to identify the number of individuals and repeated occasions required to detect and/or predict scale-model random intercept variance.

In the second study, we aimed to provide empirically-based guidelines for building scale models that include random effects. Specifically, we evaluated Type I error rates for systematically-varying scale-model fixed effects of individual level predictors to examine the implications of testing them in three scenarios: when significant scale-model random intercept variance is erroneously omitted, when scale-model random intercept variance is initially undetectable, and when it is no longer detectable given significant scale-model fixed effects.

Method

The data-generating mixed-effects location-scale model

In this study, we are concerned with detecting and predicting individual differences in intra-individual variability for outcomes that fluctuate across occasions. As such, we only examined scale models for the residual variance in which differential intra-individual variability is created by scale-model fixed effects of *individual-level* predictors. Given the absence of occasion-level predictors in our two-level sampling design, constant individual differences in intra-individual variability are quantified solely via scale-model random intercept variance $\sigma^2_{t_0}$ which is the scale-model analogue to quantifying constant individual differences in the location (or amount) of an outcome via location-model random intercept variance $\sigma^2_{b_0}$.

Under the conditional distribution of the observations, $\mathbf{Y}_i|\mathbf{b}_i \sim \mathbf{N}_{n_i}(\mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i = \sigma^2_{e_i}\mathbf{I}_{n_i})$, the simulation data-generating location model for the outcome of individual i at occasion j is provided in (3) below.

$$Y_{i,j} = \left(\beta_0 + b_{0,i}\right) + e_{i,j} \tag{3}$$

In (3), $Y_{i,j}$ is the outcome for individual i at occasion j, β_0 is the location-model fixed intercept, $b_{0,i}$ is the location-model random intercept for the individual-specific deviation from the fixed intercept whose variance $\sigma^2_{b_0}$ represents constant individual mean differences, and $e_{i,j}$ is the residual value for individual i at occasion j whose variance $\sigma^2_{e_i}$ is the outcome of the scale model for the residual variance described throughout the next few paragraphs.

Type I error rates for detecting non-existent scale-model random intercept variance were evaluated using the data-generating scale model for the residual variance shown in (4) below.

$$\sigma^2_e = \exp\left(\tau_0\right) \tag{4}$$

In (4), σ^2_e is the residual variance that quantifies a *constant* amount of intra-individual variability across individuals and τ_0 is the scale-model fixed intercept representing the log of the residual variance. Because the scale-model random intercept is absent, this scale model reduces to that of a traditional linear mixed-effects model.

The power to detect scale-model random intercept variance was evaluated using the data-generating scale model for the residual variance in (5) below.

$$\sigma^2_{e_i} = \exp\left(\tau_0 + t_{0,i}\right) \tag{5}$$

In (5), $\sigma^2_{e_i}$ is now specific to individual i and $t_{0,i}$ is the scale-model random intercept representing the individual-specific deviation from the scale-model fixed intercept whose variance $\sigma^2_{t_0}$ quantifies constant individual differences in intra-individual variability.

The power of an individual-level predictor of the scale-model random intercept variance was evaluated using the data-generating scale-model for the residual variance in (6) below.

$$\sigma^2_{e_i} = \exp\left(\left(\tau_0 + t_{0,i}\right) + \tau_1\left(W_i\right)\right) \tag{6}$$

In (6), τ_0 now represents log residual variance when individual-level predictor W_i equals 0 and τ_1 is the scale-model fixed effect that represents the difference in log intra-individual variability per one-unit increase in W_i.

Sampling distributions for study design characteristics and model parameters

Walters (2015) conducted an initial simulation study whose 8,000 replications indicated that the power to detect and predict scale-model random intercept variance was near unity based on the model parameters from Hedeker et al. (2008), Rast et al. (2012), and Rast and Zimprich (2011). As such, the results reported in this study are based on 20,000 additional replications simulated for each of the scale models in (4), (5) and (6), respectively; all replications used the location model in (3). Table 1 provides the distributions for each sampling dimension as well as for each fixed and random effect. For all replications, we constrained the correlation between the location- and scale-model random intercepts to be zero to ensure that the likelihood ratio test for the significance of the scale-model random intercept variance was not influenced by the correlation between the location- and scale-model random intercepts. However, the empirical power programs in the supplementary appendix allow specification of any correlation (including zero) when

Table 1. Sampling distributions for all effects in the location and scale models.

	Notation	Distribution	
Sampling Dimensions			
Number of Individuals	N	U(25, 200)	
Repeated Occasions with an Individual	n_i	U(5, 50)	
Scale-Model Individual-Level Predictor	W_i	N(0, 1)	
	Notation	Natural Log Scale	Outcome Scale
Location Model			
Fixed Intercept	β_0	—	0
Random Intercept Variance	$\sigma^2_{b_0}$	—	U(1, 10)
Scale Model for the Residual Variance			
Fixed Intercept	τ_0	U(–1, 3)	U(0.4, 20)
Random Intercept Variance	$\sigma^2_{t_0}$	U(0.01, 0.15)	—
Pseudo-R^2 for Individual-Level Predictor	—	See Note	—
Fixed Effect of Individual-Level Predictor	τ_1	See Note	—
Correlation between Location- and Scale-Model Random Intercepts	$\rho_{(b_{0,i};t_{0,i})}$	—	0

Note. The distinction between the natural log and outcome scale is necessary given that all effects included in the scale model for the residual variance are estimated by a generalized linear mixed-effects model using the natural log link. Values on the outcome scale were determined by exponentiating the log-scale values. Pseudo-R^2 was sampled from U(0.01, 0.25) after which the fixed effect of the individual-level predictor was determined using equation (8); approximately 25% of scale-model fixed effect values τ_1 were sampled to be 0.

estimating the power to detect and predict scale-model random intercept variance. These programs also adjust accordingly the critical mixture χ^2 for the likelihood ratio test (as required to prevent the boundary condition problem that occurs when testing the null hypothesis that the scale-model random intercept variance is equal to zero; Stram & Lee, 1994).

Estimated model sequence

Table 2 shows the model sequence used to detect and predict individual differences in intra-individual variability as quantified by scale-model random intercept variance; the location model was held constant for all models. As shown in the top of Table 2, two models were compared to determine the Type I error rate and power to detect scale-model random intercept variance: model 1 was a traditional mixed-effects model that omitted the scale-model random intercept variance, whereas model 2 was a mixed-effects location-scale model that estimated the scale-model random intercept variance. As shown in the bottom of Table 2, four models were compared to assess the power to predict scale-model random intercept variance. Misspecified model 3 was a traditional mixed-effects model that assumed homogeneous residual variance with no individual differences in intra-individual variability: it omitted both the fixed effect of individual-level predictor W_i and the scale-model random intercept. Misspecified model 4 created heterogeneous residual variances via a systematically-varying scale-model fixed effect of W_i, but allowed no further individual differences in intra-individual variability by omitting the scale-model random intercept. Misspecified model 5 was a mixed-effects location-scale model that included individual differences in intra-individual variability through the scale-model random intercept, but it omitted the scale-model fixed effect of W_i. Finally, true model 6 was a mixed-effects location-scale model that included both the fixed effect of individual-level predictor W_i and the scale-model random intercept.

Model estimation

All models were estimated using MixRegLS, free software developed specifically to estimate mixed-effects location-scale models with a location-model random intercept and a scale-model random intercept (Hedeker & Nordgren, 2013). The program assumes outcomes are conditionally normally distributed and uses marginal maximum likelihood estimation through the Newton-Raphson algorithm. It uses numeric quadrature to integrate over the random effects; our models were estimated using 15 quadrature points per random effect dimension.

Results

Type I errors in detecting random individual differences in intra-individual variability

Type I error rates for detecting individual differences in intra-individual variability—the scale-model random intercept variance—were based on the location model in (3) and the scale model for the residual variance in (4) that constrained the true population scale-model random intercept variance to zero. We excluded 26 replications in which the mixed-effects location-scale model did not converge; therefore, the results below are based on 19,974 of the 20,000 estimated models. The statistical significance of the scale-model random intercept variance was evaluated using the likelihood ratio test (i.e., the –2 log-likelihood difference) between true model 1 and misspecified model 2 based on a critical value $\chi^2 = 2.71$

Table 2. Model Sequence of Estimated Location- and Scale-Model Fixed and Random Effects.

	Location Model		Scale Model for the Residual Variance					
	Fixed Effect	Random Effect	Fixed Effects		Random Effect	Level-2 Random Effect Variances and Correlation		
	β_0	$b_{0,i}$	τ_0	τ_1	$t_{0,i}$	$\sigma^2_{b_0}$	$\sigma^2_{t_0}$	$\rho_{b_0 \cdot t_0}$
Type I Error Rate and Power to Detect Scale-Model Random Intercept Variance								
Model 1	•	•	•			•		
Model 2	•	•	•		•		•	
Power to Predict Scale-Model Random Intercept Variance								
Model 3	•	•	•			•		
Model 4	•	•	•	•		•		
Model 5	•	•	•		•	•	•	
Model 6	•	•	•	•	•	•	•	

Note. A • in a given column indicates that the effect is included in the model. β_0 = location-model fixed intercept. $b_{0,i}$ = location-model random intercept for individual *i* whose variance across individuals is the location-model random intercept variance $\sigma^2_{b_0}$. τ_0= scale-model fixed intercept for the log of the residual variance. τ_1 = scale-model fixed effect of level-2 predictor W_i. $t_{0,i}$ = scale-model random intercept for individual *i* whose variance across individuals is the scale-model random intercept variance $\sigma^2_{t_0}$. $\rho_{b_0 \cdot t_0}$ = the correlation between the location- and scale-model random intercepts.

from an equal mixture of chi-square distributions with $\alpha = .05$ and degrees of freedom (df) = 0 and 1, respectively, resulting from the addition of the scale-model random intercept variance.

Results indicated a marginal Type I error rate for detecting non-existent scale-model random intercept variance of 4.2%, 95% CI [3.9%, 4.5%]. After evaluating the functional form of all sampling design and model parameters using locally weighted scatterplot smoothing methods (aka, LOESS or LOWESS; Cleveland, Devlin, & Grosse, 1988), we estimated a series of logistic regression models to examine which parameters were associated with the probability of erroneously detecting non-existent scale-model random intercept variance. Results indicated that Type I error rates increased marginally as a function of increasing only the number of individuals; however, each additional individual increased the odds of committing a Type I error by only 0.3%, 95% CI [0.2%, 0.4%]. Overall, these results indicate no problems with Type I error rate when using the typical practice of likelihood ratio testing for the presence of scale-model random intercept variance.

The power to detect random individual differences in intra-individual variability

The power to detect individual differences in intra-individual variability were based on the location model in (3) and the scale model for the residual variance in (5). Results are based on 19,994 of the 20,000 estimated models; six replications failed to converge. The significance of the scale-model random intercept variance was evaluated using a likelihood ratio test between misspecified model 1 and true model 2 with a critical value $\chi^2 = 2.71$ from an equally-weighted mixture of chi-square distributions with $\alpha = .05$ and df = 0 and 1, respectively.

We observed an 83.2% marginal empirical power rate to detect the scale-model random intercept variance across all replications, 95% CI [82.7%, 83.7%]. After evaluating the functional form of all sampling design and model parameters using LOESS methods (Cleveland et al., 1988), we estimated a series of logistic regression models to examine which parameters were associated with the power to detect scale-model random intercept variance. Results of separate single-predictor logistic regression models indicated that power increased marginally as a function of increasing individuals, repeated occasions, the scale-model fixed intercept representing the log of the average amount of intra-individual variability, and the amount of scale-model random intercept variance representing individual differences in intra-individual variability. The location-model fixed intercept and location-model random intercept variance were not associated with power to detect the scale-model random intercept variance. However, given that a logistic regression model with multiple predictors indicated that the scale-model fixed intercept did not predict power after also including the amount of scale-model random intercept variance, the scale-model fixed intercept was removed as a predictor. In our final model, we observed a three-way over-additive interaction in which the power to detect the scale-model random intercept variance increased multiplicatively as a function of increasing the number of individuals, the number of repeated occasions, and/or the amount of scale-model random intercept variance available to be detected (see Table S1 in supplementary appendix). Representative power curves are presented in Figure 2, in which power was estimated from 500 replications per every five-person interval at specific combinations of the number of occasions and scale-model random intercept variance; the location-model random intercept variance and scale-model fixed intercept were held constant at one.

In general, these results map closely onto existing statistical power theory as increased power is obtained with more data (i.e., individuals or occasions) and more available variance to be detected. However, given the three-way over-additive interaction, the power increases associated with increases in the number of individuals, occasions, and scale-model random intercept variance varied as a function of the interacting variable(s). Although larger scale-model random intercept variance showed consistently large increases in power, this variance component is generally viewed as a fixed quantity (particularly in non-experimental studies). Therefore, considering sampling design characteristics that researchers can generally control, our results indicated that increases in the number of repeated occasions resulted in larger power increases relative to increases in the number of individuals—a result observed over the entire range scale-model random intercept variance considered in our simulation study.

Pseudo-R^2 as an effect size metric for scale-model individual-level predictors

Predictor effect sizes are an important consideration when conducting power analyses and reporting study findings. A well-described effect size for fixed effects in traditional mixed-effects models is *pseudo-R^2* that quantifies the proportion reduction of a given variance component (Snijders & Bosker, 1994). However, pseudo-R^2 values are rarely, if ever, discussed for scale-model fixed effects when estimating the mixed-effects location-scale model. Thus, it remains an open question whether we can reliably quantify the proportion of scale-model random intercept variance explained by scale-model

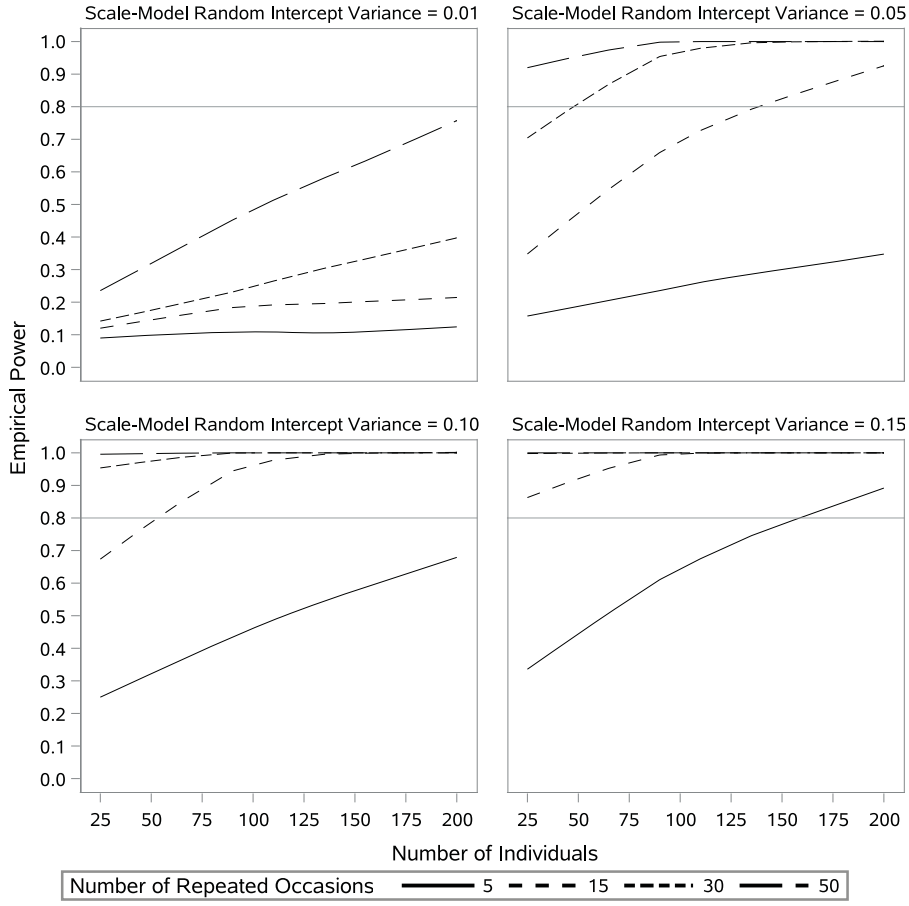

Figure 2. Power curves for detecting scale-model random intercept variance.

individual-level predictors. We answered this question via a separate simulation study to support our subsequent use of pseudo-R^2 when estimating the power to predict individual differences in intra-individual variability.

Specifically, we evaluated the recovery of population pseudo-R^2 values of 5%, 15%, and 25% based on combinations of study design and model parameters shown in Table 3; we simulated 1,000 replications within each scenario. Holding constant the location model shown in (3), we calculated the pseudo-R^2 estimate for each replication as shown in (7) below.

$$\hat{R}^2_{\text{pseudo}} = \frac{\hat{\sigma}^2_{t_0,(\text{exclude } W_i)} - \hat{\sigma}^2_{t_0,(\text{include } W_i)}}{\hat{\sigma}^2_{t_0,(\text{exclude } W_i)}}, \qquad (7)$$

In (7), $\hat{\sigma}^2_{t_0,(\text{exclude } W_i)}$ and $\hat{\sigma}^2_{t_0,(\text{include } W_i)}$ are the estimated scale-model random intercept variance from the unconditional and conditional scale models shown in (5) and (6), respectively.

Further, the unstandardized fixed effect for individual-level predictor W_i, τ_1, required to achieve the population pseudo-R^2 was calculated as shown in (8) below.

$$\tau_1 = \sqrt{\frac{1}{V(W_i)}\left(\sigma^2_{t_0,(\text{exclude } W_i)}\right)\left(R^2_{\text{population}}\right)} \qquad (8)$$

In (8), $\sigma^2_{t_0,(\text{exclude } W_i)}$ is the population unconditional scale-model random intercept variance from the model without the individual-level predictor W_i shown in (5), $R^2_{\text{population}}$ is the population pseudo-R^2 value, and $V(W_i)$ is the variance of W_i which was assumed normally distributed with a mean of zero for all replications.

Given that the distribution of pseudo-R^2 is skewed as it approaches 0 or 1, recovery of population pseudo-R^2 was indicated when the interquartile range of estimated pseudo-R^2 values across replications contained the true population pseudo-R^2. Further, to quantify the accuracy of pseudo-R^2 within each scenario, we also calculated mean signed bias (MSB), as shown in (9), to quantify the under- or over-estimation of pseudo-R^2.

$$\text{MSB} = \frac{1}{N_{\text{reps}}}\sum_{r=1}^{N_{reps}}\left(\hat{R}^2_{\text{pseudo},r} - R^2_{\text{population}}\right) \qquad (9)$$

In (9), $\hat{R}^2_{\text{pseudo}}$ is the estimated pseudo-R^2 for a given replication r, $R^2_{\text{population}}$ is the population pseudo-R^2 value, and N_{reps} is the number of within-scenario replications in which the mixed-effects location-scale model converged. A larger absolute value of MSB indicates that estimated pseudo-R^2 was further from the true population value (i.e., increased bias), with positive and negative values

indicating that pseudo-R^2 was over- and under-estimated, respectively.

We also calculated the root mean squared error (RMSE) of pseudo-R^2 as shown in (10), in which all values were defined in (9); larger values of RMSE indicate decreased accuracy.

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{reps}}}\sum_{r=1}^{N_{reps}}\left(\hat{R}^2_{\text{pseudo},r} - R^2_{\text{population}}\right)^2} \quad (10)$$

As shown in Table 3, population pseudo-R^2 was recovered in all scenarios; however, the estimate of pseudo-R^2 became more reliable with larger sampling designs as increases in the number of individuals and/or repeated occasions reduced MSB and RMSE. These results suggest that pseudo-R^2 can be used to quantify the proportion of scale-model random intercept variance explained by scale-model individual-level fixed effects.

The power to predict individual differences in intra-individual variability

We then examined the power to detect individual differences in intra-individual variability based on the location model in (3) and the scale model for the residual variance in (6). Results are based on 19,985 of the 20,000 estimated models; 15 replications were excluded due to convergence failures. Based on the likelihood ratio test between models 3 and 5 (using an equal mixture $\chi^2 = 2.71$ with $\alpha = .05$ and df = 0 and 1), we found an 83.0% marginal empirical power rate to detect scale-model random intercept variance, 95% CI [82.4%, 83.5%]. We used these 83.0% of replications to evaluate the power to detect the fixed effect of the individual-level predictor in the scale model for the residual variance. Statistical significance of the scale-model fixed effect was indicated when its 95% Wald confidence interval excluded zero.

Results indicated a 52.0% marginal empirical power to detect this scale-model predictor, 95% CI [51.2%, 52.7%]. After evaluating the functional form of all sampling design and model parameters using LOESS methods (Cleveland et al., 1988), we estimated a series of logistic regression models to examine which parameters were associated with the power to detect scale-model individual-level fixed effect. Results of separate single-predictor logistic regression models indicated that power increased marginally as a function of increasing individuals, repeated occasions, the amount of scale-model random intercept variance available to be explained, and pseudo-R^2; the location- and scale-model fixed intercept and the location-model random intercept variance were not associated with power and were excluded from further analysis. Results of a multiple logistic regression model indicated a four-way over-additive interaction which mapped closely onto existing statistical power theory as power increased multiplicatively as a function of increasing the number of individuals, the number of repeated occasions, the amount of scale-model random intercept variance available to be detected, and/or pseudo-R^2 (see Table S2 in supplementary appendix). Given this four-way interaction effect, the power increases associated with increases in the number of individuals, occasions, scale-model random intercept variance, and pseudo-R^2 varied as a function of the interacting variable(s). Although both larger pseudo-R^2 values and scale-model random intercept variances were consistently associated with substantial increases in power, both are likely to be viewed as fixed quantities, particularly in nonexperimental studies. As such, considering design characteristics that researchers can generally control, our results indicated that

Table 3. Recovery of pseudo-R^2 under twelve specific scenarios.

						Pseudo-R^2			
N	n_i	$\sigma^2_{b_0}$	τ_0	$\sigma^2_{t_0}$	$V(W_i)$	Population	Estimated[a]	MSB	RMSE
25	10	2	2	0.05	1	0.05	0.14 [0.04–0.36]	0.19	0.32
25	30	2	2	0.05	1	0.15	0.19 [0.07–0.40]	0.12	0.27
25	50	2	2	0.05	1	0.25	0.30 [0.14–0.51]	0.08	0.25
50	10	1	−1	0.10	1	0.05	0.09 [0.03–0.22]	0.11	0.21
50	30	1	−1	0.10	1	0.15	0.17 [0.09–0.28]	0.04	0.14
50	50	1	−1	0.10	1	0.25	0.25 [0.18–0.36]	0.02	0.14
100	10	2	2	0.05	4	0.05	0.08 [0.03–0.21]	0.10	0.21
100	30	2	2	0.05	4	0.15	0.16 [0.08–0.25]	0.03	0.13
100	50	2	2	0.05	4	0.25	0.26 [0.18–0.35]	0.02	0.12
200	10	1	−1	0.10	4	0.05	0.05 [0.02–0.11]	0.03	0.08
200	30	1	−1	0.10	4	0.15	0.16 [0.11–0.20]	0.01	0.07
200	50	1	−1	0.10	4	0.25	0.26 [0.21–0.30]	0.01	0.07

Note. N = the number of individuals. n_i = number of repeated occasions within an individual. $\sigma^2_{b_0}$ = location-model random intercept variance. τ_0 = scale-model fixed intercept for the log of the residual variance. $\sigma^2_{t_0}$ = scale-model random intercept variance. $V(W_i)$ = variance of scale-model individual-level predictor W_i. MSB = mean signed bias. RMSE = root mean square error.

[a]Observed pseudo-R^2 presented as median [inter-quartile range].

increasing the number of individuals resulted in larger power increases to detect the effect of an individual-level predictor relative to increasing the number of occasions.

Finally, representative power curves are presented in Figure 3, in which power was estimated based on 500 replications per every five-person interval at specific combinations of the number of occasions, scale-model random intercept variance, and pseudo-R^2; both the location-model random intercept variance and scale-model fixed intercept were held constant at one.

Systematically-varying scale-model effects

We next examined the importance of including random individual differences in intra-individual variability when testing scale-model individual-level fixed effects under three scenarios: (1) when scale-model random intercept variance was significantly greater than zero but erroneously omitted, (2) when it is approximately zero because it did not exist initially, and (3) when it is approximately zero because it was fully explained by scale-model individual-level fixed effects. In all scenarios, these fixed effects would be described as *systematically varying*—varying as a function of known predictors, but not randomly otherwise. The validity of such systematically-varying effects in scale models is currently an open question.

To address scenario (1), we calculated Type I error rates for the fixed effect of the individual-level predictor from misspecified model 4 as the proportion of the 23.1% of replications in which the scale-model random intercept variance was significant (based on the likelihood ratio test between models 3 and 5 using an equal mixture $\chi^2 = 2.71$ with $\alpha = .05$ and df = 0 and 1), but the true value of the fixed effect was sampled to be zero. Consistent with Leckie et al. (2014), we found that erroneously omitting significant scale-model random intercept variance resulted in a marginal Type I error rate of 20.5% for scale-model individual-level fixed effects, 95% CI [19.2%, 21.9%]. A multiple logistic regression model indicated that the odds of a Type I error increased by 9.5% for every 0.01-point increase in the erroneously omitted scale-model random intercept variance, 95% CI [7.6%, 11.4%], and increased by 2.4% for each additional repeated occasion, 95% CI [1.7%, 3.1%]; these two effects did not interact.

These results strongly suggest that scale-model random intercept variance should be tested first before testing the significance of scale-model individual-level

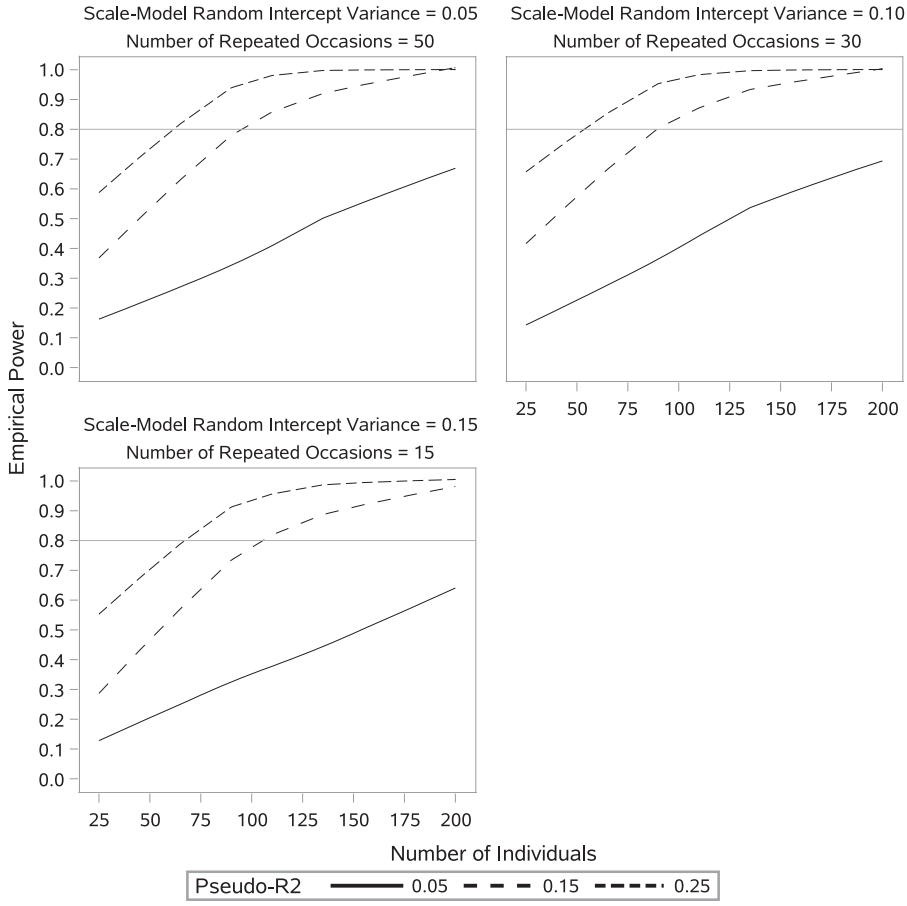

Figure 3. Power curves for predicting scale-model random intercept variance using an individual-level predictor.

predictors. But what if the scale-model random intercept variance is *not* initially significant? Our simulation also allowed us to answer this question from scenario (2) as well: of the 17.0% of replications in which the scale-model random intercept variance was *not* significant initially (based on the likelihood ratio test between models 3 and 5 using an equal mixture $\chi^2 = 2.71$ with $\alpha = .05$ and df = 0 and 1), the true value of the scale-model individual-level fixed effect was zero for 26.2% of these replications. Based on these replications, a Type I error rate of 5.8% was observed from true model 6 which retained the nonsignificant scale-model random intercept variance, 95% CI [4.4%, 7.6%]; the Type I error rate was also acceptable at 6.5% for model 4 which omitted the nonsignificant scale-model random intercept variance, 95% CI [5.0%, 8.3%]. Further, for replications in which the true value of the scale-model individual-level fixed effect was greater than zero, its empirical power for true model 6 was 26.8%, 95% CI [25.1%, 28.6%], compared to 29.2% for model 4, 95% CI [27.4%, 31.0%]. Therefore, it appears that tests of systematically-varying scale-model individual-level fixed effects in the absence of detectable scale-model random intercept variance can proceed with minimal concern of inflated Type I error rates. However, these fixed effects will likely be difficult to detect given the abysmal statistical power.

We then addressed scenario (3): of the 83.0% of replications in which scale-model random intercept variance was significant initially, the inclusion of a fixed effect for the scale-model individual-level predictor explained the majority of the initially significant scale-model random intercept variance for 20.4% of these replications, such that a nonsignificant amount of unexplained variance remained (as indicated by the likelihood ratio test between models 4 and 6 using an equal mixture $\chi^2 = 2.71$ with $\alpha = .05$ and df = 0 and 1, indicating that removing scale-model random intercept variance did not result in significantly *worse* model fit). Of these replications, 22.8% had a true population value of the fixed effect of the individual-level predictor equal to zero. In these replications, the predictor's fixed effect had a Type I error rate of 6.3%, 95% CI [4.9%, 8.1%], when estimated by true model 6.

Using the power simulation programs to conduct a priori power analyses

To help researchers conduct *a priori* power analyses, we created two simulation programs with which to quantify the empirical power to detect and predict individual differences in intra-individual variability. Specifically, the researcher can use either SAS or R software to simulate any number of replicated data sets based on their sampling design, expected variance components, and pseudo-R^2, call MixRegLS to estimate the mixed-effects location-scale model, and report the empirical power and parameter recovery statistics. In the supplementary appendix, we provide these simulation programs, Excel spreadsheets to calculate the required values to input into the simulation programs, as well as a set of user's guides.

What follows below is an example scenario detailing how our simulation programs can be used to estimate the number of individuals and repeated occasions to ensure sufficient power to detect and predict individual differences in intra-individual variability. Although the power estimates in this scenario are based on results from a small pilot study, we acknowledge that pilot data may not be available or produce results representative of population effects. As such, we encourage researchers to power their studies based on the smallest substantively interesting values of scale-model random intercept variance and pseudo-R^2. With that in mind, we move on to the example.

Consider a researcher interested in quantifying individual differences in the intra-individual variability of daily physical activity in overweight and obese adolescents (under the assumption of non-zero observed physical activity in each epoch). She analyzed pilot data from 10 adolescents across 21 days using a traditional mixed-effects model, in which the location-model random intercept variance was estimated to be 125 and residual variance was estimated to be 375. A subsequent mixed-effects location-scale model estimated the scale-model random intercept variance to be 0.05, the scale-model fixed intercept to be 5.90, and a correlation between the location- and scale-model random intercepts of 0.25; the addition of scale-model random intercept variance did not provide a significant improvement to model fit as determined by a likelihood ratio test using alpha = 0.05 and a critical mixture $\chi^2 = 5.14$. The post-hoc power estimate for the scale-model random intercept variance from her pilot data was ~20%. Based on the obtained variances and the correlation between the location- and scale-model random intercepts, as well as alpha = 0.05, she could achieve ~80% power using her current 21-day study period and monitoring 80 adolescents; as an alternative, ~80% power could be achieved by monitoring 55 adolescents and extending the study period to 30 days.

Besides adequate power to *detect* individual differences in intra-individual variability, she must now assess the adequacy of her sampling design to *predict* those individual differences. Her pilot data showed a between-person measure of perceived social support ($M = 10.0$, $SD = 1.0$) to be a significant predictor of the amount of physical activity, and she expects it will explain 15% of individual differences in intra-individual variability in

physical activity. Given this effect size and the aforementioned estimates, if she were to monitor 55 adolescents for 30 days, she would have only ~50% power to detect the effect of social support. However, retaining the 30-day study period, ~80% power would be achieved by monitoring 110 adolescents.

In this example scenario, sufficient power to detect individual differences in intra-individual variability did not guarantee adequate power to predict those individual differences, this will not always be the case (such as when available scale-model random intercept variance or effect sizes are expected to be larger).

Discussion

The mixed-effects location-scale model allows quantification and prediction of individual differences in mean level and in intra-individual variability within a single model. We believe this model is infrequently estimated in practice due to limited empirical guidance regarding sampling designs for quantifying and predicting individual differences in intra-individual variability. To help address these limitations, our study had two aims. First, we investigated via simulation the interplay between research design, Type I error rates, and *a priori* expectations of statistical power in detecting and predicting individual differences in intra-individual variability. Although Type I error rates in detecting individual differences in intra-individual variability increased with larger level-2 sample sizes, the extent of this increase was inconsequential. Further, results of our simulation studies aligned with statistical power theory, as greater power to detect individual differences in intra-individual variability was observed with more data (i.e., more individuals and/or repeated occasions) and larger scale-model random intercept variances available to be detected. Similar results were found regarding the power to predict these individual differences, although with the addition that individual-level scale-model predictors with larger effect sizes (as indicated by pseudo-R^2 for the scale-model random intercept variance) will require fewer individuals and/or repeated occasions to achieve sufficient power.

Second, in addition to examining the power to predict *existing* individual differences in intra-individual variability, our simulation also sought to evaluate whether individual-level scale-model predictors can be included appropriately when individual differences in intra-individual variability are non-detectable. We did so under three scenarios: when significant scale-model random intercept variance was erroneously omitted, when scale-model random intercept variance was initially nonsignificant, and when it was initially detected but rendered nonsignificant (i.e., fully explained) by the inclusion of a scale-model individual-level predictor. In each scenario, the fixed effect of the individual-level predictor would be considered a *systematically-varying effect* given that all individual differences in intra-individual variability would be a function of known predictors. Consistent with Leckie et al. (2014), we also found drastically increased Type I error rates for an individual-level scale-model predictor when significant scale-model random intercept variance was erroneously omitted. Further, in the absence of initially detectable individual differences in intra-individual variability, the Type I error rate for a scale-model individual-level predictor remained acceptable (~ 5%) whether the nonsignificant scale-model random intercept variance was retained or omitted, but that the predictor's power was very low (< 30%). Similarly, our results indicated that Type I error rates were approximately 5% for individual-level predictors that fully explained scale-model random intercept variance.

Advice for using the mixed-effects location-scale model in practice

Taken together, our model-building advice for testing hypotheses regarding individual differences in intra-individual variability is three-fold. First, studies should be designed to try to ensure adequate statistical power to both detect and predict those individual differences to the greatest extent possible. Although this is admittedly easier said than done, we hope that the power simulation programs we developed can help facilitate *a priori* power analysis.

Second, we recommend that analyses for the scale model for the residual variance begin by testing the significance of the scale-model random intercept variance before testing scale-model individual-level predictors. This suggestion is based on our current finding (and that of Leckie et al., 2014) of excessive Type I error rates of scale-model individual-level predictors when erroneously omitting significant scale-model random intercept variance. If scale-model random intercept variance is estimable, but not statistically significant, the researcher must also consider whether this non-statistically significant variance component is clinically or practically meaningful. Although we are unaware of benchmarks that define clinically-significant scale-model random intercept variance, the definition of non-ignorable variance undoubtedly varies by field of study (e.g., physical activity vs. fMRI). It may be useful to conduct a visual depiction of between-individual differences in intra-individual variability via a line graph of level-1 residuals (or, with identical results and perhaps more clarity, the observed outcome values) plotted against measurement occasion for a select number of individuals (e.g., see

Figure 3 in Watts, Walters, Hoffman, & Templin, 2016). If, upon review, the magnitude of individual differences is deemed clinically meaningful, a larger more powerful study will be required to allow adequate detection (and prediction) of those individual differences.

Third, if the scale-model random intercept variance is neither statistically nor clinically significant, it is still acceptable to test individual-level predictors in the scale-model for the residual variance. Alternatively, if the scale-model random intercept variance is initially significant, it should be included while testing individual-level predictors. If the scale-model random intercept variance has been fully explained, we nevertheless recommend its retention in all subsequent models at least until model convergence problems are encountered (at which point it may be removed).

Limitations and future directions

For this study, we used MixRegLS software developed by Hedeker and Nordgren (2013) to estimate all mixed-effects location-scale models. The primary limitation of MixRegLS is that it currently cannot estimate random slope variances in either the location- or scale model for the residual variance. Therefore, if random slope variances are needed, other software options will be required—Hedeker et al. (2008) have provided annotated SAS code for estimation in PROC NLMIXED that *in theory* can estimate any number of location- and/or scale-model random effects, and Bayesian approaches have been provided by Kupar et al. (2015), Rast et al. (2012), and Wang et al. (2012). Although our convergence rate across replications was high, we did experience estimation difficulties when the scale-model random intercept variance was sampled to be approximately zero—a situation that could potentially be remedied by increasing the number of quadrature points at the cost of computational efficiency (Pinheiro & Bates, 1995). In our power simulation programs, we tried our best to include code to catch estimation errors, but we urge researchers to stay diligent when examining parameter recovery and power estimates and remove erroneous replications as necessary. Finally, although a Bayesian estimator might have estimated all models for all replications, it is important to note that this estimator is not a panacea for the convergence failures encountered when estimating this model using frequentist approaches (e.g., maximum likelihood), as biased posterior parameter estimates are likely with small level-2 sample sizes and uninformative priors (Rast et al., 2012).

The present work focused specifically on a two-level sampling design in which the repeated measurements of a level-1 conditionally normally distributed continuous outcome was nested within level-2 individuals; we also focused only on designs in which the outcome fluctuated within individuals (i.e., no systematic change or other time-related dependence was expected across the repeated occasions) and no additional random slopes were included. Thus, it is important to examine the generalizability of our findings for other types of outcomes when collected under alternative sampling designs and under more salient influences of time. Further, all simulation studies constrained the correlation between the location- and scale-model random intercepts to zero to ensure a one degree of freedom likelihood ratio test; however, we acknowledge that this correlation will likely be non-zero in practice. Although Walters (2015) found a correlation of .50 to have no association with the power to detect or predict individual differences in intra-individual variability in a Bayesian framework using deviance information criterion, it is unknown whether our simulation study results (conducted in a frequentist framework) would have been different had this correlation been estimated.

In addition, our simulation did not address the ubiquitous problem of missing data, which has consistently been shown to deteriorate statistical power and to have disastrous effects on the accuracy and recovery of fixed and random effects (Enders, 2010). Further, because we did not consider occasion-level predictors in either the location model or scale model for the residual variance, we constrained residual variance estimates to be constant within an individual. An open question remains to what extent the detection of fixed and random effects in the scale model for the residual variance is impacted by occasion-level predictors in the location model. Considering the results of our simulation studies, and given that the purpose of location-model occasion-level predictors is to explain residual variance, we expect that their inclusion could make individual differences in intra-individual variability more difficult to then detect and predict.

In conclusion, we hope to have provided empirical scientists with some practical tools and advice with which to form and test hypotheses related to individual differences in intra-individual variability. We have admittedly explored only a small segment of the methodological landscape for the mixed-effects location-scale model, but we hope our work helps inspire others to become more confident explorers of this useful analytic approach.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported

any financial or other conflicts of interest in relation to the work described.

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

Funding: This work was not supported.

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

Acknowledgments: The authors would like to thank Associate Editor Dr. Keith Widaman, two anonymous reviewers, as well as Dr. Walt Stroup and Dr. Cal Garbin from the University of Nebraska-Lincoln for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

References

Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Journal of the Royal Statistical Society*, *36*(3), 332–339. doi:10.2307/2347792

Cleveland, W., Denby, L., & Liu, C. (2000). Random location and scale effects: Model building methods for a general class of models. *Computing Science and Statistics*, *32*, 3–10.

Cleveland, W. S., Devlin, S. J., & Grosse, E. (1988). Regression by local fitting: Methods, properties, and computational algorithms. *Journal of Econometrics*, *37*(1), 87–114. doi:10.1016/0304-4076(88)90077-2

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measures data*. Boca Raton, FL: CRC Press.

Dixon, R. A., Lentz, T. L., Garrett, D. D., MacDonald, S. W. S., Strauss, E., & Hultsch, D. F. (2007). Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology*, *21*, 381–399. doi:10.1037/0894-4105.21.3.381

Dunton, G. F., Huh, J., Leventhal, A. M., Riggs, N., Hedeker, D., Spruijt-Metz, D., & Pentz, M. A. (2013). Momentary assessment of affect, physical feeling states, and physical activity in children. *Health Psychology*, *33*(3), 255–263. doi:10.1037/a0032640

Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.

Foulley, J. L., & Quaas, R. L. (1995). Heterogeneous variances in Gaussian linear mixed models. *Genetics Selection Evolution*, *27*, 211–228. doi:10.1186/1297-9686-27-3-211

Gallagher, P., Nilsson, J., Finkelmeyer, A., Goshawk, M., Macritchie, K. A., Lloyd, A. J., … Watson, S. (2015). Neurocognitive intra-individual variability in mood disorders: Effects on attentional response time distributions. *Psychological Medicine*, *45*(14), 2985–2997. doi:10.1017/S0033291715000926

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, *44*(3), 461–465. doi:10.2307/1913974

Hedeker, D., & Mermelstein, R. J. (2007). Mixed-effects regression models with heterogeneous variance: Analyzing Ecological Momentary Assessment (EMA) data. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies*. (pp. 183–206). New York: Psychology Press.

Hedeker, D., Mermelstein, R. J., Berbaum, M. L., & Campbell, R. T. (2009). Modeling mood variation associated with smoking: An application of heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction*, *104*(2), 297–307. doi:10.1111/j.1360-0443.2008.02435.x

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*(2), 627–634. doi:10.1111/j.1541-0420.2007.00924.x

Hedeker, D., & Nordgren, R. (2013). MIXREGLS: A program for mixed-effects location scale analysis. *Journal of Statistical Software*, *52*(12), 1–38. doi:10.18637/jss.v052.i12

Hedeker, D., Mermelstein, R. J., & 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

Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, *42*(4), 609–629. doi:10.1080/00273170701710072

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York: Routledge.

Hultsch, D. F., MacDonald, S. W., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*, *14*(4), 588–598. doi: 10.1037/0894-4105.14.4.588

Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. (2008). Intraindividual variability, cognition, and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 491–556). New York: Psychology Press.

Kupar, K., Li, X., Blood, E. A., & Hedeker, D. (2015). Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: An application to ecological momentary assessment data. *Statistics in Medicine*, *34*(4), 630–651. doi:10.1002/sim.6345

Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics*, *39*(5), 307–332. doi:10.3102/1076998614546494

Lee, Y., & Noh, M. (2012). Modelling random effect variance with double hierarchical generalized linear models. *Statistical Modelling*, *12*(6), 487–502. doi:10.1177/1471082X12460132

Littell, R. C., Pedergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, *19*(13), 1793–1819. doi:10.1002/1097-0258(20000715)19:13<1793::AID-SIM482>3.0.CO;2-Q

MacDonald, S. W., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychology and Aging*, *18*(3), 510–523. doi:10.1037/0882-7974.18.3.510

Ong, J. C., Hedeker, D., Wyatt, J. K., & Manber, R. (2016). Examining the variability of sleep patterns during treatment for chronic insomnia: Application of a location-scale mixed model. *Journal of Clinical Sleep Medicine*, *12*(6), 797–804. doi:10.5664/jcsm.5872

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*(1), 12–35. doi:10.2307/1390625

Rast, P., Hofer, S. M., & Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research*, *47*(2), 177–200. doi:10.1080/00273171.2012.658328

Rast, P., & Zimprich, D. (2011). Modeling within-person variance in reaction time data of older adults. *The Journal of Gerontopsychology and Geriatric Psychiatry*, *24*(4), 169–176. doi:10.1024/1662-9647/a000045

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.). Thousand Oaks, CA: Sage.

Schneider, S., Junghaenel, D. U., Keefe, F. J., Schwartz, J. E., Stone, A. A., & Broderick, J. E. (2012). Individual differences in the day-to-day variability of pain, fatigue, and well-being in patients with rheumatic disease: Associations of psychological variables. *Pain*, *153*(4), 813–822. doi:10.1016/j.pain/2012.01.001

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods Research*, *22*(3), 342–363. doi:10.1177/0049124194022003004

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*(4), 1171–1177. doi:10.2307/2533455

Walters, R. W. (2015). *Mixed-effects location-scale models for conditionally normally distributed repeated-measures data* (Doctoral dissertation). Retrieved from http://digitalcommons.unl.edu/psychdiss/75

Wang, L., Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods*, *17*(4), 567–581. doi:10.1037/a0029317

Watts, A., Walters, R. W., Hoffman, L., & Templin, J. (2016). Intra-individual variability of physical activity in older adults with and without mild Alzheimer's disease. *PLoS ONE*, *11*(4), e0153898. doi:10.1371/journal.pone.0153898