ANNUAL REVIEWS

*Annual Review of Psychology*

# Catching Up on Multilevel Modeling

Lesa Hoffman[1] and Ryan W. Walters[2]

[1]Department of Psychological and Quantitative Foundations, University of Iowa, Iowa City, Iowa 52242, USA; email: Lesa-Hoffman@UIowa.edu

[2]Department of Clinical Research, Creighton University, Omaha, Nebraska 68178, USA



Annu. Rev. Psychol. 2022. 73:659–89

The *Annual Review of Psychology* is online at psych.annualreviews.org

https://doi.org/10.1146/annurev-psych-020821-103525



## Keywords

mixed-effects models, hierarchical linear models, centering, random slopes, mixed-effects location–scale models

## Abstract

This review focuses on the use of multilevel models in psychology and other social sciences. We target readers who are catching up on current best practices and sources of controversy in the specification of multilevel models. We first describe common use cases for clustered, longitudinal, and cross-classified designs, as well as their combinations. Using examples from both clustered and longitudinal designs, we then address issues of centering for observed predictor variables: its use in creating interpretable fixed and random effects of predictors, its relationship to endogeneity problems (correlations between predictors and model error terms), and its translation into multivariate multilevel models (using latent-centering within multilevel structural equation models). Finally, we describe novel extensions—mixed-effects location–scale models—designed for predicting differential amounts of variability.

## Contents

CATCHING UP ON MULTILEVEL MODELING ..... 660
PREVIOUS EPISODES: EXAMPLE USE CASES FOR MULTILEVEL MODELS ..... 661
Series Trailer ..... 661
Season 1: Clustered Samples ..... 662
Season 2: Longitudinal Samples ..... 662
Season 3: Clustered Longitudinal Samples ..... 663
Season 4: Cross-Classified Samples ..... 663
Recap: Seasons 1–4 ..... 664
A RECURRING PLOT LINE: CENTERING IN MULTILEVEL MODELS ..... 664
Season 1 Director Commentary: Centering in Clustered Samples ..... 664
Crossover Episode: Smushed Effects and Endogeneity ..... 668
Season 2 Director Commentary: Centering in Longitudinal Samples ..... 668
Plot Twist: When Centering with Observed Variables Can Fail ..... 671
A New Character: Latent-Centering within Multivariate MLMs ..... 672
Recap and Recommendations: Centering ..... 674
ANOTHER RECURRING PLOT LINE: SPECIFICATION OF RANDOM EFFECTS ..... 675
Conflict on the Set: Which Random Effects Are Needed? ..... 676
Conflict Resolution: How to Avoid Smushed Random Slopes ..... 677
Bonus Material: Effect Sizes Using Explained Variance ..... 678
Recap: Random Effects ..... 678
CURRENT EPISODES: MIXED-EFFECTS LOCATION–SCALE MODELS ..... 679
Season 5: MELSMs for Clustered Samples ..... 679
Season 6: MELSMs for Longitudinal Samples ..... 681
Recap and Spin-Offs: MELSMs ..... 684
CONCLUSION ..... 685

## CATCHING UP ON MULTILEVEL MODELING

**Cross-level interaction:** a model-estimated constant slope for the multiplicative combination of two predictor variables measured for different dimensions of sampling

Multilevel models (MLMs) are a versatile family of statistical models used to quantify and predict the distinct sources of variability that arise when sampling over multiple dimensions simultaneously, such as occasions, persons, and groups. Also known as hierarchical linear models, MLMs are so named for their expression via a layered system of equations that partition variability due to each sampling dimension into a distinct level of analysis. In MLMs one can simultaneously examine main effects and interactions of predictors measured for each sampling dimension, as well as interactions of predictors from different dimensions (the latter are known as cross-level interactions). Like traditional (single-level) regression models, MLMs also include a fixed intercept (for the expected outcome when all predictors are 0), fixed slopes of predictors (for the expected difference in the outcome per unit higher in each predictor), and a model residual (for the deviation of the actual outcome from the model-predicted outcome). But unlike in traditional regression models, the intercept and slopes of predictors in MLMs can also vary randomly across sampling units. As elaborated below, the latter additions are known as random effects (and thus the term random coefficients model is also a synonym for MLMs, as is mixed-effects model, a term used for models that include both fixed and random effects).

MLMs have become a prominent tool in psychology and many other areas (for historical overviews, see Eckardt et al. 2021, Hox & Roberts 2011), but the breadth and depth of this field can be intimidating for those looking to join in for the first time. This is similar to watching the fifth season of a television show without having watched the first four seasons, in which the characters and their plot lines can be difficult to follow without having seen every prior episode. Given that we are writing this review during a pandemic in which binge-watching television has become our new national pastime, we decided to frame our review from this perspective—what do readers need to know to start catching up on one of our favorite shows, multilevel modeling?

To help readers from different areas get up to date, we first review common use cases for MLMs. This is followed by a detailed treatment of centering (a recurring plot line throughout the series). We then provide advice about specifying random effects in MLMs (including some recent controversies) and using $R^2$ explained variance in assessing effect size. Finally, we try to equip readers for future MLM episodes by describing a relatively new variant—mixed-effects location–scale models—by which to answer research questions about differential variability. Throughout, we reference methodological research and tutorials for MLMs using both clustered and longitudinal designs; readers interested in only one of these can skip the other without losing too much of the storyline. But one review cannot cover every aspect of MLMs, and so we did not expand on issues of study planning, power analysis, or missing data. For these and other topics, we recommend two recent comprehensive handbooks (Humphrey & LeBreton 2019, Scott et al. 2013). For readers with no prior exposure to MLMs, the **Supplemental Material** provides a list of suggested introductory textbooks.

## PREVIOUS EPISODES: EXAMPLE USE CASES FOR MULTILEVEL MODELS

### Series Trailer

The notion of a simple random sample—in which all observations have unrelated model residuals—is rarely accurate in characterizing real-life research data. In educational contexts, students and teachers may be sampled from multiple schools; in health care settings, patients may be sampled from multiple hospitals. In fields such as sociology or political science, survey respondents may be sampled from multiple countries. In studying family dynamics, data are deliberately collected from multiple individuals and from multiple households. In each of these examples of clustered samples, the residuals of persons from the same cluster (school, hospital, country, or household) may be more related than those of persons from different clusters. The same is true in longitudinal samples (in which persons or groups are measured repeatedly over time), in which correlated residuals occur for occasions from the same person or group. More generally, residuals from the same sampling unit are likely to be more related than residuals from different sampling units—these patterns of residual correlation are known as dependency.

Inaccurate inferences can result when outcomes from clustered or longitudinal samples are analyzed with traditional regression models that only include a single error term (which assumes independent residuals). To solve this problem, MLMs include additional error terms—random intercepts and slopes, known more generally as random effects—that capture correlations of residuals from the same sampling unit. Random intercepts capture constant correlation through mean differences between higher-level sampling units. Random slopes capture nonconstant correlation through differences in the slopes of lower-level predictors across higher-level sampling units. (If those definitions sound like word salad right now, please stay tuned for more concrete examples.) MLMs estimate the variances of these random effects (and the covariances of random effects for the same sampling unit), along with the variance (and possibly covariance) of the lowest-level

**Fixed intercept:** a model-estimated constant that provides the expected (conditional mean) outcome when all predictor variables are 0, as also used in single-level regression models

**Fixed slope:** a model-estimated constant that provides the expected difference in an outcome variable per unit difference in a predictor variable

**Residual:** the deviation between the outcome predicted by the model's fixed and random effects and the actual outcome, with model-estimated variance across lowest-level units

**Random effect:** a latent (unobserved) variable (e.g., intercept or slope) with model-estimated variance used in linear combination with predictors to create an expected outcome for each observation

**Clustered sample:** the sampling of subjects from preexisting groups, such as children in schools, patients in hospitals, and siblings within families

**Longitudinal sample:** the repeated sampling of observations from the same unit (e.g., person or group), such as over time, over conditions, or over stimulus-specific trials

Supplemental Material >

residuals. Higher-level random effects are usually assumed to have a multivariate normal distribution, but other conditional distributions can be paired with link (transformation) functions to predict nonnormal outcomes within generalized MLMs (e.g., a logit link and Bernoulli distribution for binary responses or a log link and Poisson distribution for count responses).

**Dependency:** the presence of correlation among residuals from the same higher-level sampling unit (which can be captured by the inclusion of random effects)

**Random intercept:** the latent deviation between the fixed intercept and the intercept for a given higher-level sampling unit, with model-estimated variance across higher-level sampling units

**Random slope:** the latent deviation between a fixed slope and the slope for a given higher-level sampling unit, with model-estimated variance across higher-level sampling units

**Fixed effect:** a model-estimated constant (e.g., intercept or slope) used in linear combination with predictors to create an expected outcome for each observation

## Season 1: Clustered Samples

To describe the use of MLMs for clustered samples, consider a two-level example of level-1 persons nested in level-2 clusters. The level-1 residual variance captures within-cluster differences across persons to be explained by fixed slopes of person-level-1 predictors and their interactions. The level-2 random intercept variance captures between-cluster differences in the outcome mean across persons to be explained by fixed slopes of cluster-level-2 predictors and their interactions. Level-2 random slope variances capture between-cluster differences in the slopes of the person-level-1 predictors, which can be explained by cross-level interactions of each person-level-1 predictor with cluster-level-2 predictors. Although the sources of remaining outcome variance—level-1 residual, level-2 random intercept, and level-2 random slopes—are usually assumed constant over all observations, options for nonconstant variances are afforded by location–scale variants of MLMs (stay tuned).

More complex sampling designs can require additional levels of analysis. For example, if multiple family members are sampled in different cities, then level-1 persons are nested in level-2 families, and level-2 families are nested in level-3 cities. Level-1 (within-family) residual variance can be explained by fixed slopes of person-level-1 predictors and their interactions, level-2 (between-family, within-city) random intercept variance can be explained by fixed slopes of family-level-2 predictors and their interactions, and level-3 (between-city) random intercept variance can be explained by fixed slopes of city-level-3 predictors and their interactions. The slopes of person-level-1 predictors can vary randomly over level-2 families and/or level-3 cities, and each random slope variance can be explained by cross-level interactions of that person-level-1 predictor with family-level-2 and/or city-level-3 predictors. Likewise, the slopes of family-level-2 predictors can vary randomly over level-3 cities, and each random slope variance can be explained by cross-level interactions of that family-level-2 predictor with city-level-3 predictors.

## Season 2: Longitudinal Samples

The flexibility of MLMs for capturing patterns of dependency when sampling over multiple dimensions also extends to longitudinal (i.e., panel) designs in which level-1 occasions are modeled as nested in level-2 persons (also known as growth curve models). MLMs pool time series data across multiple persons to examine average time trends and variability in and around those trends. In this case, the level-1 residual variance captures within-person fluctuation over time to be explained by fixed slopes of time-varying predictors—time, other time-level-1 variables, and their interactions. The level-2 random intercept variance captures between-person differences in the outcome mean over time to be explained by fixed effects of time-invariant predictors—person-level-2 variables and their interactions. Level-2 random slope variances capture between-person slope differences for time or other time-level-1 predictors, each of which can be explained by cross-level interactions of that time-level-1 predictor with person-level-2 predictors. Longitudinal data over shorter intervals (e.g., daily) can also be analyzed with the same type of two-level model to examine between-person differences in slopes capturing shorter time trends (e.g., cycles by day of the week).

Longitudinal designs with more intensive measurement schedules may require additional levels of analysis. For instance, when multiple occasions per day are collected across multiple days,

level-1 occasions from the same day are nested in level-2 days, and level-2 days are nested in level-3 persons. One could examine how level-1 within-day changes (e.g., circadian rhythms) vary across days within persons (as captured by level-2 random slopes) and/or between persons on average (as captured by level-3 random slopes), as well as day-level-2 or person-level-3 moderators of those within-day changes (via corresponding cross-level interactions). Level-1 (within-day) residual variance can be explained by fixed slopes of time-level-1 predictors and their interactions, level-2 (between-day, within-person) random intercept variance can be explained by fixed slopes of day-level-2 predictors and their interactions, and level-3 (between-person) random intercept variance can be explained by fixed slopes of person-level-3 predictors and their interactions. The slopes of time-level-1 predictors can vary across level-2 days and/or level-3 persons, the slopes of day-level-2 predictors can vary across level-3 persons, and each of these random slope variances can be explained by corresponding cross-level interactions.

## Season 3: Clustered Longitudinal Samples

MLMs can also be extended for combinations of clustered and longitudinal samples. In clustered longitudinal designs of repeatedly measured persons in clusters, level-1 occasions are nested in level-2 persons, and level-2 persons are nested in level-3 clusters. But if the same cluster is measured by different persons over time, level-1 persons are instead nested in level-2 occasions, and level-2 occasions are nested in level-3 clusters. In these types of hierarchical three-level models, in addition to a level-1 residual variance and the level-2 and level-3 random intercept variances, the slopes of level-1 predictors can vary randomly over level-2 units and/or level-3 units, and the slopes of level-2 predictors can vary randomly over level-3 units.

## Season 4: Cross-Classified Samples

For nested sampling designs that are not strictly hierarchical, MLMs with cross-classified random effects can be used when two (or more) sampling dimensions are crossed at the same level. In fact, many longitudinal designs are actually crossed designs—when observations are nested in both sampling units and occasions, units are actually crossed by occasions at level 2. But after capturing mean differences over time through the inclusion of fixed slopes for time, the random occasion dimension is no longer needed, resulting in a purely nested design of time-specific outcomes nested in units (often abbreviated as level-1 occasions in level-2 units).

As a clustered example of cross-classification, consider children sampled from multiple neighborhoods and schools, in which not all children who live in the same neighborhood attend the same school. Level-1 children are nested in crossed level-2 dimensions of neighborhoods and schools. Mean differences across each dimension can be captured by a random intercept variance (to be explained by per-neighborhood or per-school predictors). As a longitudinal example, cross-classification also occurs when cluster membership changes over time, such as when patients change therapy groups over time. After mean differences over time by fixed time slopes are accounted for, level-1 occasions are nested in crossed level-2 dimensions of patients and groups. Mean differences across each level-2 dimension can be captured by a random intercept variance (to be explained by per-patient or per-group predictors). In such cross-classified designs, the slopes of level-1 predictors can vary randomly over each level-2 dimension, and the slopes of level-2 predictors can also vary randomly over the other crossed level-2 dimensions.

MLMs with cross-classified random effects are also applicable to repeated measures experiments. If each subject responds to the same items once, level-1 trials are nested in both level-2 subjects and level-2 items, and mean differences across each dimension can be captured by a random intercept variance (to be explained by per-subject or per-item predictors). The slopes of

trial-level-1 predictors can vary over level-2 subjects and/or level-2 items, the slopes of subject-level-2 predictors can vary over level-2 items, and the slopes of item-level-2 predictors can vary over level-2 subjects. Additional random slopes are possible if subjects respond to the same items more than once (e.g., items are presented under different within-subjects conditions).

Finally, cross-classified MLMs can also be useful in analyzing multi-rater data, such as when each person provides a response (as a rater) about each other person (as a target). Level-1 responses are then nested in both level-2 raters and level-2 targets. Unlike other cross-classified designs, this scenario requires covariances for the across-dimension random effects for the same person (i.e., for a person's random intercept as a rater and as a target, given that people who provide more favorable ratings of others on average may also be rated more favorably by others).

**Constant-centering:** subtraction of a constant that relocates the meaning of a 0 value for a predictor; used to aid in the interpretation of fixed and random intercepts, as well as of fixed slopes of predictors that are also part of interactions

### Recap: Seasons 1–4

To summarize, multilevel modeling is useful for designs that include multiple dimensions of sampling simultaneously. Typical use cases for MLMs include clustered samples (of persons nested in groups), longitudinal samples (of occasions nested in persons and/or groups), and cross-classified samples (of observations nested in multiple crossed dimensions). Like single-level regression models, MLMs use fixed intercepts and predictor slopes (which are constants) to create expected outcomes. But MLMs also use random intercepts and predictor slopes (which are variables) to capture correlations among residuals from the same sampling unit. Said differently, MLMs partition the unexplained outcome variance into distinct levels of sampling, attributed either to higher-level random intercepts and slopes or to lowest-level residuals. The sources of variance for each level of sampling can then be explained by predictors at that level.

## A RECURRING PLOT LINE: CENTERING IN MULTILEVEL MODELS

We now introduce one of the most salient plot lines in the history of MLMs—the use of centering for lower-level variables. In single-level models, centering refers only to the rescaling of a predictor by subtracting a constant so that 0 becomes meaningful. Centering is used to create an interpretable intercept (the expected outcome when all predictors equal 0), as well as interpretable main effect slopes of predictors that are also included in interaction terms (in which those main effect slopes become conditional on their interacting predictors equaling 0). A common choice for a centering constant is the sample mean (known as grand-mean-centering), but any constant can be used (e.g., 12 years of education so that 0 indicates a high school degree). Consequently, this strategy can be labeled more generally as constant-centering (Hoffman 2019), or centering-at-a-constant (Rights & Sterba 2019), in which special cases are centering at the grand mean or at 0 (leaving predictors uncentered). Critically, constant-centering changes a predictor's mean, but it does not change its model predictions or variability. In MLMs, level-2 predictors can be constant-centered for the same interpretational conveniences as in single-level models, but greater consideration is needed in centering level-1 predictors (and lower-level predictors in general), as described in the next section.

### Season 1 Director Commentary: Centering in Clustered Samples

To illustrate the centering of level-1 predictors in clustered samples, let us consider a two-level example of patients (*p*) nested in clinics (*c*), in which we aim to predict a patient outcome (*anxiety*) from patient compliance with therapy activities (*comp*) in clinics that use one of two treatments (*type*; 0 = old, 1 = new) for all of their patients. Although compliance is measured per level-1

patient, a predictor for patient compliance must be included at each level, as explained below. An example MLM for this clustered design is shown in Equation 1,

$$\begin{aligned} &\text{Level 1 } (L1): anxiety_{pc} = \beta_{0c} + \beta_{1c}\left(L1comp_{pc}\right) + e_{pc} \\ &\text{Level 2 } (L2): \beta_{0c} = \gamma_{00} + \gamma_{01}\left(L2comp_c\right) + \gamma_{02}\left(type_c\right) + U_{0c}\,, \\ &\qquad\qquad \beta_{1c} = \gamma_{10} + U_{1c} \end{aligned} \qquad 1.$$

in which $anxiety_{pc}$ is the level-1 outcome for level-1 patient $p$ from level-2 clinic $c$.

In the level-1 model, the $\beta$ terms are placeholders for the two clinic outcomes: $\beta_{0c}$ holds the overall clinic intercept, and $\beta_{1c}$ holds the overall clinic slope of patient-level-1 compliance. Next, $e_{pc}$ is the level-1 residual—the deviation of the actual outcome from the model-predicted patient outcome (with constant variance $\sigma_e^2$ over patients and clinics). Each $\beta$ clinic outcome is defined by a level-2 equation using fixed effects (the $\gamma$ terms, which are constants) and random effects (the $U$ terms, which vary over clinics). The first subscript of each effect indexes the level-1 $\beta$ to which it belongs: All effects predicting $\beta_{0c}$ begin with 0, and all effects predicting $\beta_{1c}$ begin with 1. For the $\gamma$ fixed effects, the second subscript indexes its order in the level-2 equation: The fixed effects predicting $\beta_{0c}$ begin with $\gamma_{00}$, followed by $\gamma_{01}$ and $\gamma_{02}$ for its first and second predictors. For the $U$ random effects, the second subscript indicates the sampling unit it varies over (e.g., $c$ for level-2 clinics here). In predicting the level-2 $\beta_{0c}$ clinic intercept, $\gamma_{00}$ is the fixed intercept (the expected outcome when all predictors equal 0), whereas $\gamma_{01}$ and $\gamma_{02}$ are the fixed slopes of the clinic-level-2 predictors (the intercept difference per unit higher of each predictor). $U_{0c}$ is the level-2 random intercept—the deviation of the actual intercept from the predicted clinic intercept (with constant variance $\tau^2_{U_0}$ over clinics). In predicting the level-2 $\beta_{1c}$ per-clinic slope of patient-level-1 compliance, $\gamma_{10}$ is the average slope across clinics, and $U_{1c}$ is a level-2 random slope—the deviation of the actual slope from the predicted clinic slope (with constant variance $\tau^2_{U_1}$ over clinics).

**Figure 1** displays example results for the model in Equation 1, in which the prediction of patient anxiety for five example patients in each of 12 clinics is shown using separate lines (with 3 clinics in each of four combinations from $type_c$ by $L2comp_c$). The fixed intercept of $\gamma_{00} = 18$ is shown by the $y$-axis value for the center open square in the top black line (for $type_c = 0$, $L2comp_c = 0$, and for patient $comp_{pc} = 0$ on the $x$-axis).

Let us examine the fixed slopes. Because $type_c$ varies between clinics only, the meaning of $\gamma_{02}$ is straightforward—it is the between-clinic difference in mean anxiety for clinics that use the new treatment ($type_c = 1$) instead of the old treatment ($type_c = 0$). In **Figure 1**, the $type_c$ slope $\gamma_{02} = -5.0$ is shown by the vertical distance between the black lines whose symbols have the same fill (i.e., between the open triangles and open squares, or between the filled triangles and filled squares, holding $L2comp_c$ constant). In contrast, the meaning of the level-1 and level-2 slopes of patient compliance depends on how their predictor variables are created, which requires an explanation of why two versions of the predictor are needed in the first place! The rationale is as follows: Just as any person-level-1 outcome has both between- and within-cluster variability (as captured for $anxiety_{pc}$ by the level-2 random intercept variance $\tau^2_{U_0}$ and the level-1 residual variance $\sigma_e^2$, respectively), so do most person-level-1 predictors. Here, if some clinics have more compliant patients, then level-1 $comp_{pc}$ will contain between-clinic mean differences beyond just within-clinic patient differences. Each source of predictor variance (between-level-2 and within-level-1) has a distinct potential relation with the outcome, and so most level-1 predictors are really two variables, not one. Here, we represent between-clinic mean differences using the level-2 predictor: $L2comp_c = \overline{comp}_c - C_2$, in which $\overline{comp}_c$ is the level-2 mean compliance for patients in clinic $c$ and $C_2$ is a centering constant. In **Figure 1**, separate lines distinguish the six clinics with $L2comp_c = 0$ (open symbols) from the six clinics with $L2comp_c = 1$ (filled symbols).

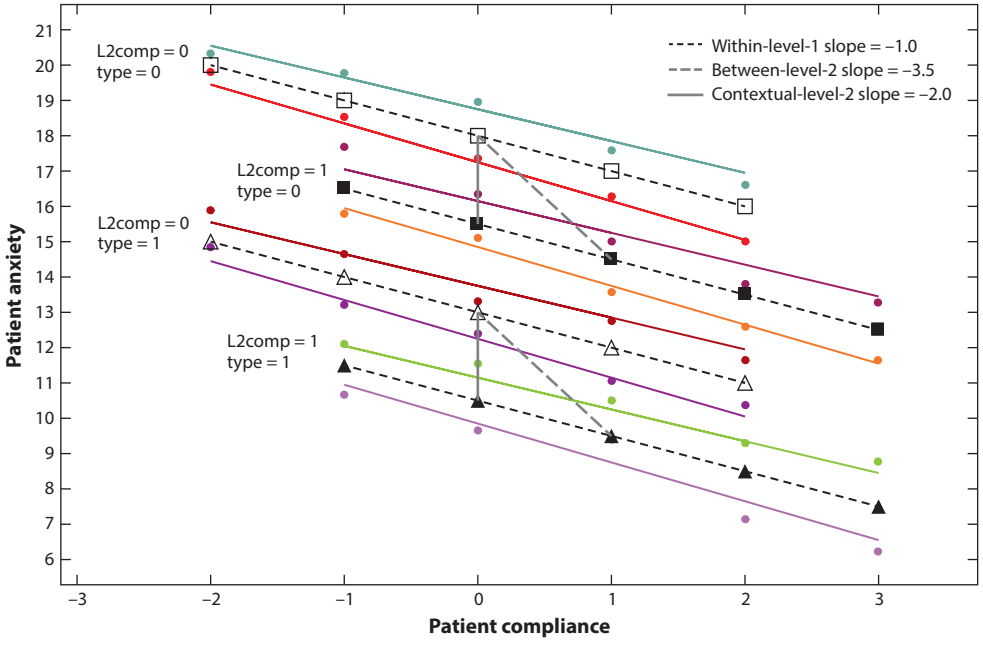


**Figure 1**

Example results for the two-level clustered model in Equation 1. The *x*-axis is original uncentered patient compliance ($comp_{pc}$). The three lines within each combination depict the $U_{0c}$ random intercept and the $U_{1c}$ random slope: Each middle dashed line (*black*) depicts values of 0, each top solid line (*various colors*) depicts positive values (resulting in relatively higher intercepts and shallower slopes), and each solid bottom line (*various colors*) depicts negative values (resulting in relatively lower intercepts and steeper slopes). The distances from the lines to the same-colored circles (the actual outcomes) show the patient-level-1 $e_{pc}$ residuals.

In deciding how to center a level-1 predictor, an alternative strategy—centering at an observed variable rather than at a constant—creates not only a meaningful 0 but also a within-level-only version of the predictor. In so-called variable-centering (Hoffman 2019), level-1 predictors are usually centered using their level-2 means (as shown below), but other options are possible, such as using the baseline variable of time-level-1 predictors in longitudinal designs (Algina & Swaminathan 2011) or centering level-1 predictors using more than one level-2 variable simultaneously (e.g., double decomposition; O'Keefe & Rodgers 2017). When used for clustered data, variable-centering using the cluster-level-2 mean has many names, such as cluster-mean-centering (e.g., Antonakis et al. 2021, Brincks et al. 2017, Loeys et al. 2018, Rights & Sterba 2019), group-mean-centering (e.g., Algina & Swaminathan 2011; Hofmann & Gavin 1998; Raudenbush & Bryk 2002, chapter 5; Snijders & Bosker 2012, chapter 5), centering-within-clusters (e.g., Enders & Tofighi 2007), and centering-within-contexts (e.g., Enders 2013, Kreft et al. 1995). Here, we can variable-center patient compliance by subtracting its clinic-level-2 mean, creating a new predictor: $L1comp_{pc} = comp_{pc} - \overline{comp}_c$.

**Variable-centering:** subtraction of a variable that not only relocates the meaning of a value of 0 for a predictor but also removes higher-level unit differences (such as mean differences)

Centering patient compliance at a level-2 variable instead of at a constant alters its variability in a helpful way. If level-2 clinics differ only in mean patient compliance, variable-centered $L1comp_{pc}$ will then contain only within-clinic level-1 variance. As such, its fixed slope $\gamma_{10}$ can only be a within-clinic level-1 effect—the difference in patient anxiety per unit higher compliance than other patients in the same clinic. In **Figure 1**, a within-level-1 slope $\gamma_{10} = -1.0$ is the slope of the black lines across the five example patients (whose $L1comp_{pc}$ values are ±2 of their

**Table 1 Predictor effect type by model specification**

| Centering strategy for level-1 predictor (constant-centered level-2 predictor) | Fixed effect type by predictors included | | |
|---|---|---|---|
| | Level-1 only | Level-2 only | Both levels |
| **Variable-centered level-1** | | | |
| Level-1 predictor: $L1x_{wb} = x_{wb} - \bar{x}_b$ | Within | (= 0) | Within |
| Level-2 predictor: $L2x_b = \bar{x}_b - C_2$ | (= 0) | Between | Between |
| **Constant-centered level-1** | | | |
| Level-1 predictor: $L1x_{wb} = x_{wb} - C_1$ | Smushed | (= 0) | Within |
| Level-2 predictor: $L2x_{wb} = \bar{x}_b - C_2$ | (= Within) | Between | Contextual |

Abbreviations: $w$, within; $b$, between; $C_1$, level-1 centering constant; $C_2$, level-2 centering constant.
Parentheses indicate assumptions about the fixed slopes of omitted predictors.



clinic-level-2 mean, creating the original patient $comp_{pc}$ values on the $x$-axis). Further, given that $L2comp_c$ ($= \overline{comp}_c - C_2$) has only between-clinic level-2 variance, its fixed slope $\gamma_{01}$ can only be a between-clinic level-2 effect—the difference in clinic-mean anxiety per unit higher average patient compliance than other clinics. In **Figure 1**, a between-level-2 slope $\gamma_{01} = -3.5$ is shown by the slope of the dashed gray lines connecting the clinics with $L2comp_c = 0$ and 1 (i.e., a one-unit difference in $L2comp_c$) with the same symbol (holding $type_c$ constant) when patient-level-1 compliance is at the clinic mean ($L1comp_{pc} = 0$).

Although some have cautioned against variable-centering lower-level predictors (e.g., Hox 2010, chapter 4; Snijders & Bosker 2012, chapter 5), we actively encourage it because it simplifies interpretation. Variable-centering leads to an orthogonal between–within partitioning of a level-1 predictor's variance that directly parallels what happens to a level-1 outcome: The level-2 mean predictor approximates its level-2 random intercept (which contains between-level-2 variance only), and the variable-centered level-1 predictor approximates its level-1 residual (which contains within-level-1 variance only). This partitioning greatly simplifies the meaning of the fixed slopes of the level-1 and level-2 predictors into one possible interpretation for each, as shown in **Table 1**. It also simplifies the interpretation of the random slope for the level-1 predictor (stay tuned).

What if we had chosen to constant-center patient-level-1 compliance instead, as $L1comp_{pc} = comp_{pc} - C_1$ (in which $C_1$ is the constant)? As shown in **Table 1**, the meaning of its fixed slope will depend on whether a fixed slope for $L2comp_c$ is still included. On the surface, a fixed slope for $L2comp_c$ may then seem unnecessary given that clinic-level-2 mean variability is still included in this version of patient-level-1 compliance. But as described in many sources (e.g., Raudenbush & Bryk 2002, chapter 5), without a fixed slope for its corresponding level-2 mean predictor, the fixed slope of a constant-centered level-1 predictor is an uninterpretable blend of its within-level-1 and between-level-2 fixed slopes. This useless blended effect has many names, including total effect (e.g., Burstein 1980; Raudenbush & Bryk 2002, chapter 5; Snijders & Bosker 2012, chapter 3), conflated effect (e.g., Preacher et al. 2010), composite effect (e.g., Wang & Maxwell 2015), convergence effect (e.g., Hoffman 2012, Sliwinski et al. 2010), and smushed effect (e.g., Hoffman 2015, chapter 8); we adopt the informal (yet descriptive) term smushed effect here as well.

**Smushed effect:** the unintended blending of a predictor variable's slopes across multiple levels of sampling by which distinct predictor slopes are constrained equal

Adding a fixed slope for the level-2 mean of the constant-centered level-1 predictor prevents its smushed effect (with respect to its fixed slope; stay tuned for what happens to its random slope), but the fixed slope for the level-2 mean predictor is then not a between-level-2 effect as before. Instead, it is a contextual-level-2 effect—it is how the between-level-2 fixed slope differs from the within-level-1 fixed slope (as given by $\gamma_{01} - \gamma_{10}$ in the variable-centered level-1 model using $L1comp_{pc}$

instead). Also known as an emergent-level-2 effect, a contextual-level-2 effect is the incremental between-level-2 effect after controlling for the level-1 predictor. Here, a contextual-level-2 fixed slope is the difference in clinic-mean anxiety per unit higher average patient compliance than other clinics after controlling for patient compliance. In **Figure 1**, a contextual-level-2 slope of $\gamma_{01} - \gamma_{10} = -3.5 + 1.0 = -2.0$ is shown by the vertical distance of the solid gray lines connecting the clinics with $L2comp_c = 0$ and 1 (a one-unit difference in $L2comp_c$) with the same symbol (holding $type_c$ constant) when the original patient $comp_{pc} = 0$ on the $x$-axis. More generally, a contextual-level-2 effect indicates to what extent cluster membership still matters after controlling for each person's predictor, a quantity often of interest in clustered designs (Blaklock 1984, Burstein 1980, Hofmann & Gavin 1998). Excellent examples of real-world contextual-level-2 effects are given in Antonakis et al. (2021) and Bliese et al. (2018).

**Endogeneity:** the presence of an unmodeled relation between a model predictor and a random effect or residual term (which can be caused by smushed lower-level fixed slopes)

## Crossover Episode: Smushed Effects and Endogeneity

The need to disaggregate fixed slopes across levels of analysis parallels debates of fixed versus random effects (e.g., in sociology, econometrics) for preventing endogeneity bias, which can occur if a level-1 predictor is correlated with a level-2 random intercept. One solution is to remove all level-2 mean differences using a fixed effects model—by including level-2 cluster membership as a categorical predictor (i.e., by including $C - 1$ dummy codes for the $C$ clusters). Although useful for making inferences about each specific level-2 unit (see Cushing et al. 2014), an unfortunate limitation is that other cluster-level-2 predictors then cannot be examined. Recent work has shown how endogeneity tests indicating the superiority of a fixed effects model over a random effects model actually indicate that fixed slopes of constant-centered level-1 predictors have been smushed (Bell et al. 2019, Hamaker & Muthén 2020, McNeish & Kelly 2019, Townsend et al. 2013). The term correlated random effects model describes the addition of a contextual-level-2 fixed slope to remove the correlation of a constant-centered level-1 predictor with a level-2 random intercept, thus preventing predictor endogeneity bias (Antonakis et al. 2021).

## Season 2 Director Commentary: Centering in Longitudinal Samples

Spoiler alert: In MLMs for longitudinal samples, the same centering process described above for clustered samples may not be sufficient to create purely within-level-1 predictors. Let us begin a new example—a two-level longitudinal model of level-1 occasions ($t$) nested in level-2 patients ($p$)—in which we predict a time-varying outcome (*anxiety*) from time-varying patient compliance with therapy activities (*comp*) and time-in-treatment (*time*) in which patients (from the same clinic) were given one of two treatments (*type*; 0 = old, 1 = new). A predictor for time-level-1 compliance is needed at both levels to properly specify its effects. As in all MLMs, this model requires a long data structure, with one row per level-1 unit (here, occasion) and one column per variable. An example MLM for this longitudinal design is shown in Equation 2,

$$\begin{aligned}
\text{Level 1 } (L1): anxiety_{tp} &= \beta_{0p} + \beta_{1p}\left(L1comp_{tp}\right) + \beta_{2p}\left(time_{tp}\right) + e_{tp} \\
\text{Level 2 } (L2): \beta_{0p} &= \gamma_{00} + \gamma_{01}\left(L2comp_p\right) + \gamma_{02}\left(type_p\right) \\
&\quad + \gamma_{03}\left(L2comp_p\right)\left(type_p\right) + U_{0p} \\
\beta_{1p} &= \gamma_{10} + \gamma_{12}\left(type_p\right) + U_{1p} \\
\beta_{2p} &= \gamma_{20} + \gamma_{22}\left(type_p\right) + U_{2p}
\end{aligned} \quad , \qquad 2.$$

in which $anxiety_{tp}$ is now the level-1 outcome for level-1 occasion $t$ from level-2 patient $p$.

In the level-1 model, $\beta_{0p}$ holds the overall patient intercept, $\beta_{1p}$ and $\beta_{2p}$ hold the overall patient slopes of time-level-1 compliance and time-in-treatment, and $e_{tp}$ is the level-1 residual (the deviation of the actual outcome from the model-predicted per-occasion outcome, with constant variance $\sigma_e^2$ over occasions and persons). In predicting the level-2 $\beta_{0p}$ intercept, $\gamma_{00}$ is the

fixed intercept; $\gamma_{01}$, $\gamma_{02}$, and $\gamma_{03}$ are the fixed slopes of the patient-level-2 predictors; and $U_{0p}$ is the level-2 random intercept (the deviation of the actual intercept from the predicted patient intercept, with constant variance $\tau^2_{U_0}$ over patients). In predicting the level-2 $\beta_{1p}$ slope of time-level-1 compliance, $\gamma_{10}$ is the slope for patients given the old treatment, $\gamma_{12}$ is the slope difference for patients given the new treatment, and $U_{1p}$ is a level-2 random slope (the deviation of the actual slope from the predicted patient slope, with constant variance $\tau^2_{U_1}$ over patients). Likewise, in predicting the level-2 $\beta_{2p}$ slope of time-in-treatment, $\gamma_{20}$ is the slope for patients given the old treatment, $\gamma_{22}$ is the slope difference for patients given the new treatment, and $U_{2p}$ is a level-2 random time slope (the deviation of the actual slope from the predicted patient slope, with constant variance $\tau^2_{U_2}$ over patients).

**Figure 2** displays example results for the model in Equation 2, in which the prediction of time-varying anxiety for five example occasions in each of 12 patients is shown using separate lines (with 3 patients in each of four combinations from $type_p$ by $L2comp_p$). **Figure 2*a*** depicts the effects of time-level-1 compliance ($L1comp_{tp}$) at $time_{tp} = 4$; **Figure 2*b*** depicts change over time-in-treatment ($time_{tp}$) holding $L1comp_{tp} = 0$. In **Figure 2*b***, the fixed intercept of $\gamma_{00} = 26$ is shown by the *y*-axis value for the first open square in the top black line (for $type_c = 0$ and $L2comp_c = 0$). Also in **Figure 2*b***, the $type_p$ slope of $\gamma_{02} = -0.5$ is shown by the vertical distance from the first open square to the first open triangle (at $time_{tp} = 0$ and conditional on $L2comp_p = 0$).

Following the clustered example from Season 1, the time-level-1 compliance predictor can be constant-centered or variable-centered. When using the person-level-2 mean in longitudinal studies, the latter is called person-mean-centering (e.g., Algina & Swaminathan 2011; Curran & Bauer 2011; Hoffman 2015, chapter 8; Wang & Maxwell 2015) or subject-mean-centering (e.g., Loeys et al. 2018). Here, we center time-varying compliance using its patient-level-2 mean, creating time-level-1 compliance as $L1comp_{tp} = comp_{tp} - \overline{comp}_p$. If level-2 patients differ randomly only in their mean compliance over time (and not in their change in compliance over time; stay tuned), the new variable-centered $L1comp_{tp}$ will contain only within-patient level-1 variance. If so, the $\gamma_{10}$ fixed slope of $L1comp_{tp}$ must be a within-patient level-1 effect: the difference in that occasion's anxiety per unit greater compliance than usual (for $type_p = 0$). In **Figure 2*a***, a within-level-1 slope $\gamma_{10} = -1.5$ for $type_p = 0$ is shown by the slope of the black lines with squares across five values of time-varying compliance (with $L1comp_{tp}$ values $\pm 2$ of the patient-level-2 mean, creating the original time-varying $comp_{tp}$ values on the *x*-axis).

In turn, the $\gamma_{01}$ fixed slope of $L2comp_p$ ($= \overline{comp}_p - C_2$) must be a between-patient level-2 effect: the difference in patient anxiety per unit greater compliance on average than other patients (in $type_p = 0$). In **Figure 2*a***, a between-level-2 slope $\gamma_{01} = -3.5$ for $type_p = 0$ is shown by the slope of the dashed gray lines with squares connecting the patients with $L2comp_p = 0$ and 1 (a one-unit difference in $L2comp_p$) when time-level-1 compliance is at the patient mean ($L1comp_{tp} = 0$). But if time-level-1 compliance were constant-centered ($L1comp_{tp} = comp_{tp} - C_1$), then the fixed slope of $L2comp_p$ would instead carry a contextual-level-2 effect: the incremental between-patient level-2 compliance effect after controlling for time-varying compliance ($comp_{tp}$). Also in **Figure 2*a***, a contextual-level-2 slope for $type_p = 0$ of $\gamma_{01} - \gamma_{10} = -3.5 + 1.5 = -2.0$ is shown by the vertical distance of the solid gray lines with squares connecting the patients with $L2comp_p = 0$ and 1 (a one-unit difference in $L2comp_p$) when the original time-varying $comp_{tp} = 0$ on the *x*-axis. But given that occasions are not distinct entities like persons are, contextual-level-2 effects are often of less interest in longitudinal designs than they might be in clustered designs (see Hoffman & Stawski 2009).

The models in Equations 1 and 2 differ in two notable ways. First, the longitudinal model in Equation 2 includes level-1 $time_{tp}$ without a predictor for its level-2 mean. This practice will yield pure level-1 effects only for predictors that do not contain any level-2 variance, such as

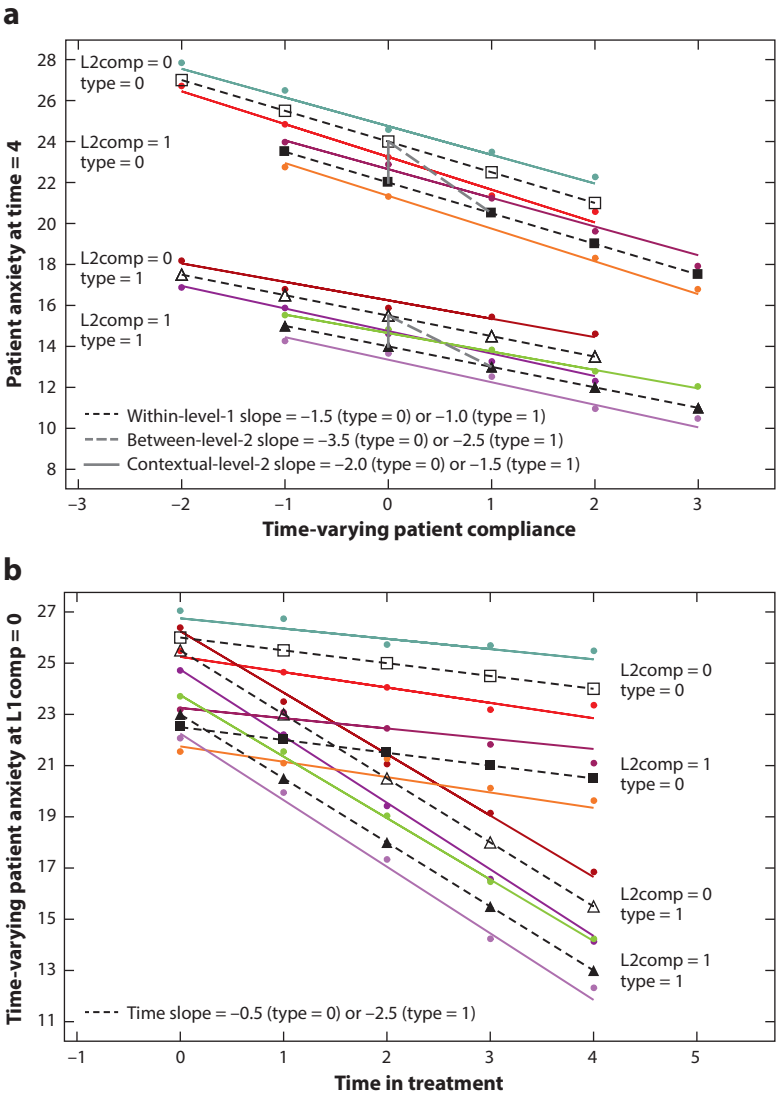


**Figure 2**

Example results for the two-level longitudinal model in Equation 2 by (*a*) time-varying patient compliance ($comp_{tp}$) and (*b*) time-in-treatment ($time_{tp}$). The three lines within each combination depict the $U_{0p}$ random intercept and the $U_{1p}$ (panel *a*) or $U_{2p}$ (panel *b*) random slope: Each middle dashed line (*black*) depicts values of 0, each top solid line (*various colors*) depicts positive values (resulting in relatively higher intercepts and shallower slopes), and each bottom solid line (*various colors*) depicts negative values (resulting in relatively lower intercepts and steeper slopes). The distances from the lines to their same-colored circles (the actual outcomes) show the time-level-1 $e_{tp}$ residuals.

the time predictor in balanced designs (in which everyone is measured at the same occasions). Otherwise, if persons differ at baseline in their place along the metric of time (e.g., in age or grade, as found in accelerated longitudinal designs; Estrada et al. 2020), then the fixed slope for the level-1 time predictor has the same potential for a smushed effect as any other constant-centered level-1 predictor (Hoffman 2012, 2015, chapter 10; Sliwinski et al. 2010).

Second, the longitudinal model in Equation 2 adds three interactions with treatment type. Two are cross-level interactions whose fixed slopes are straightforward: They describe how patients given treatment $type_p = 1$ differ from those given treatment $type_p = 0$ in their within-patient fixed slopes of time-level-1 compliance ($\gamma_{12}$) and time-in-treatment ($\gamma_{22}$). In **Figure 2*a***, the $type_p$ by $L1comp_{tp}$ interaction of $\gamma_{12} = 0.5$ is shown by how much shallower the $L1comp_{tp}$ slopes are for $type_p = 1$ (triangles) than for $type_p = 0$ (squares)—there is a weaker effect of greater compliance than usual on that occasion's anxiety for patients given the new treatment (whose $L1comp_{tp}$ slope is then $\gamma_{10} + \gamma_{12} = -1.5 + 0.5 = -1.0$). In **Figure 2*b***, the $type_p$ by $time_{tp}$ interaction of $\gamma_{22} = -2.0$ is shown by how much steeper the $time_{tp}$ slopes are for $type_p = 1$ (triangles) than for $type_p = 0$ (squares)—there is a stronger effect of time-in-treatment on that occasion's anxiety for patients given the new treatment (whose $time_{tp}$ slope is then $\gamma_{20} + \gamma_{22} = -0.5 - 2.0 = -2.5$).

In contrast, the meaning of the level-2 interaction of treatment type with patient-level-2 compliance ($\gamma_{03}$) depends on how time-level-1 compliance is centered. Because it has been variable-centered (at the patient-level-2 mean), here $\gamma_{03}$ describes treatment type differences in the between-level-2 fixed compliance slope. In **Figure 2*a***, the $type_p$ by $L2comp_p$ interaction of $\gamma_{03} = 1.0$ is shown by how much shallower the slope of the dashed gray line is for $type_p = 1$ (triangles) than for $type_p = 0$ (squares)—there is a weaker effect of higher average compliance than others on mean anxiety for patients given the new treatment (whose between-level-2 compliance slope is then $\gamma_{01} + \gamma_{03} = -3.5 + 1.0 = -2.5$). But if time-level-1 compliance were constant-centered instead, then $\gamma_{03}$ would describe treatment type differences in the contextual-level-2 fixed compliance slope. Also in **Figure 2*a***, this implied $type_p$ by contextual-level-2 slope interaction (given by the variable-centered level-1 model as $\gamma_{03} - \gamma_{12} = 1.0 - 0.5 = 0.5$) is shown by the smaller vertical distance of the solid gray line connecting the patients given $type_p = 1$ (triangles) compared with the solid gray line connecting patients given $type_p = 0$ (squares). For $type_p = 1$, the implied contextual-level-2 compliance slope would then be given as $\gamma_{01} - \gamma_{10} + \gamma_{03} - \gamma_{12} = -3.5 + 1.5 + 1.0 - 0.5 = -1.5$. Said differently, if level-1 compliance were constant-centered and the $\gamma_{03}$ level-2 interaction were omitted, the $\gamma_{12}$ cross-level interaction would be smushed—it would force equal moderation of the between-patient and within-patient fixed compliance slopes by treatment type (for elaboration, see Hoffman 2015, chapter 8, 2019). Analogous problems of smushed cross-level interactions have been described elsewhere for clustered MLMs (Enders 2013, Hofmann & Gavin 1998, Preacher et al. 2016). Likewise, interactions among level-1 predictors must also include interactions among their level-2 mean counterparts to prevent smushed level-1 interactions (Hoffman 2015, chapter 8; Loeys et al. 2018).

## Plot Twist: When Centering with Observed Variables Can Fail

As hinted at above, variable-centering using a level-2 mean may not remove all between-level-2 variance from a level-1 predictor, especially in longitudinal designs. Just as time-level-1 outcomes may contain between-person level-2 differences in both the intercept and the slopes for change over time, so may time-level-1 predictors. For instance, if persons change differently in compliance during treatment (some grow more compliant, whereas others grow less compliant), then the time-level-1 compliance predictor will contain two sources of level-2 variance. While person-mean-centering removes between-person mean (intercept) variance, the between-person time slope variance remaining in the time-level-1 predictor will be confounded with its within-person variance (see Curran & Bauer 2011). This confounding can cause two problems.

First, as noted by Hedeker & Gibbons (2006, chapter 4), artifactual interactions of time with time-level-1 predictors may result from their unequal variance across occasions (which can be caused by between-person time slope differences in the time-level-1 predictor). Second, to the

extent that the time-level-1 predictor's random time slope relates to that of the outcome, that between-person level-2 time-to-time slope relationship will be conflated with the within-person level-1 effect of the time-level-1 predictor, resulting in a time-smushed effect (Hoffman 2015, chapter 9; 2021). The same problem can also occur in shorter-term longitudinal studies, in which relations of between-person differences in cyclic patterns (e.g., day-of-the-week trends) can confound within-level-1 effects (Lui & West 2015), and potentially in clustered designs (although this is less likely in the absence of a third variable as salient as time). The solution to this problem is a multivariate extension to MLMs (with many different names), as described in the next section.

**Latent-centering:** model-based partitioning of a lower-level predictor into latent variables (e.g., a random intercept and a residual), the same as that which occurs for lower-level outcome variables in multilevel models

## A New Character: Latent-Centering within Multivariate MLMs

Given that variable-centering using an observed level-2 mean may fail to fully partition the between-level-2 and within-level-1 sources of a level-1 predictor's variance, why not just treat the level-1 predictor the same as the level-1 outcome? We can do so using a multivariate MLM in three steps: (*a*) replace the level-1 predictor's observed level-2 mean with its latent analog, a model-estimated level-2 random intercept; (*b*) replace the variable-centered observed level-1 predictor with its latent analog, a model-estimated level-1 residual; and (*c*) add any level-2 random slopes needed to capture other sources of between-level-2 variance (e.g., in change over time). This results in separate but simultaneous models predicting each level-1 variable—the level-1 predictor is now another level-1 outcome. This multivariate MLM strategy for level-1 predictors is also known as latent-centering (e.g., Asparouhov & Muthén 2019), in which the level-2 random intercept that replaced the observed level-2 mean of the level-1 predictor is called a latent mean (e.g., Preacher et al. 2016) or a latent covariate (e.g., Lüdtke et al. 2008).

The correspondence between univariate and multivariate treatments of level-1 predictors in MLMs is shown in **Figure 3**, which borrows graphical conventions used in structural equation models (SEMs). **Figure 3*a*** (univariate) depicts the use of observed predictors (created by variable-centering level-1 compliance with the patient-level-2 mean), whereas **Figure 3*b*** (multivariate) shows their replacement by latent predictors (and adds a third subscript to distinguish the fixed intercept, random intercept, and residual for each outcome).

How the relations across multivariate outcomes can be modeled differs by software. If multivariate MLMs are estimated using software for univariate MLMs (SAS, SPSS, or Stata mixed; HLM; lme in R; see Hoffman 2015, chapter 9), across-variable relations (among the level-2 random effects or among the level-1 residuals) can only be phrased as covariances. For directed effects instead (e.g., compliance predicting anxiety, as in **Figure 3**), multivariate MLMs require software for single-level SEMs or multilevel SEMs (M-SEMs), such as *Mplus*, Stata gsem, and the R packages lavaan, OpenMx, and xxM. Many sources have described how fixed and random effects in MLM can be represented equivalently by latent factor means and variances in SEMs and M-SEMs (Bauer 2003, Berry & Willoughby 2017, Curran et al. 2012, Hamaker et al. 2015, Hoffman 2019, McNeish & Matta 2018, Mehta & Neale 2005). One distinction is that SEMs use a wide data structure, with one row per level-2 sampling unit and one column per level-1 unit per variable. In contrast, M-SEMs use the same long data structure as MLMs, with one row per level-1 sampling unit and one column per variable. Accordingly, M-SEMs can be more convenient than SEMs for unbalanced designs (of different numbers of level-1 units per level-2 unit), more than two levels of analysis, cross-classified samples, and when including random slopes.

Even for predictors in which an observed level-2 mean would be sufficient to capture all between-level-2 variability (i.e., using a univariate MLM), there are several situations in which a multivariate MLM estimated as a SEM or M-SEM (via latent-centering) may be preferable. First, by treating predictors as outcomes (bringing them into the model likelihood), cases with

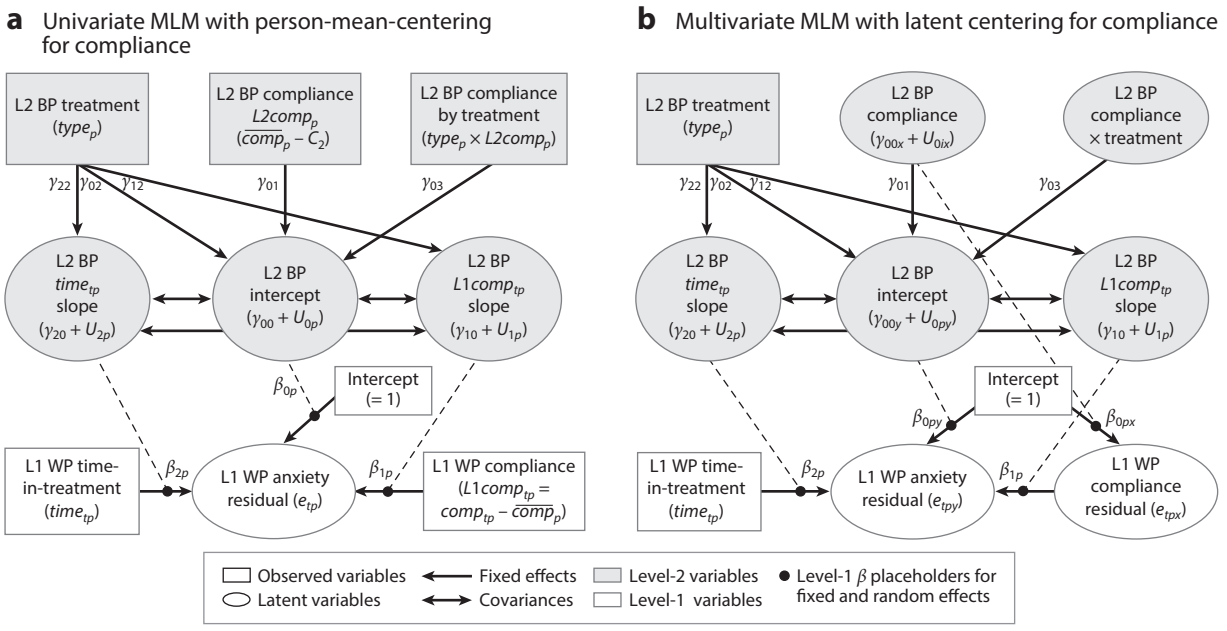


**Figure 3**

(*a*) Univariate and (*b*) multivariate MLMs for level-1 predictors based on Equation 2. Abbreviations: BP, between-person; MLM, multilevel model; L, level; WP, within-person.

missing predictors can then be included under the same assumptions as for missing outcomes (see Grund et al. 2019). Second, using an observed level-2 mean assumes it is a perfectly reliable version of a variable's level-2 random intercept, when in reality the level-2 mean's reliability is a function of the size of the level-2 unit and the variable's intraclass correlation (ICC) (i.e., the proportion of its total variance due to level-2 mean differences). Consequently, using an observed level-2 mean can result in too-small level-2 fixed effects relative to those provided by a latent level-2 mean, more so with fewer level-1 units and smaller predictor ICCs (Lüdtke et al. 2008). In the context of multilevel mediation, the same problem can result in too-small level-2 indirect fixed effects when using observed rather than latent level-2 means (Preacher et al. 2011). Third, M-SEMs can be expanded into so-called dynamic SEMs to include autoregressive and cross-lag effects given unbalanced occasions (see McNeish & Hamaker 2020). Last, SEMs can be expanded to include item-level measurement models for latent constructs, and M-SEMs can do so at multiple levels (e.g., Vandenberg & Richardson 2019), although in practice this extension may be limited to larger level-2 samples (Lüdtke et al. 2011, Meuleman & Billiet 2009).

Yet SEM and M-SEM approaches are not a panacea. Using a latent level-2 mean can provide level-2 fixed effects that are more inconsistent than those provided by an observed level-2 mean, more so with smaller level-2 samples and smaller predictor ICCs (Lüdtke et al. 2008), as well as lower power for level-2 indirect fixed effects (McNeish 2017a, Zigler & Ye 2019). Further, SEM and M-SEM software do not offer residual maximum-likelihood estimation, which provides more accurate random effects variances in smaller level-2 samples (McNeish 2017b). Bayesian estimation has been proposed as a solution for SEMs and M-SEMs with small samples, but different choices for noninformative prior distributions can have a large unintended impact on the results (Smid et al. 2020, Zitzmann et al. 2020).

Finally, there is potential for much greater confusion regarding which type of level-2 fixed effect—between or contextual—is provided by SEMs and M-SEMs, as detailed in Hoffman (2019). In SEMs that include within-level-1 fixed effects, directed paths among the random effect latent variables will be contextual-level-2 fixed effects by default. Between-level-2 fixed effects can be created instead by using level-1 structured residuals (i.e., by moving the level-1 residual variances and their paths to new latent variables; Curran et al. 2014), but this strategy may not be possible for level-1 effects with random slopes. In M-SEMs, which level-2 effect is provided differs by software, estimator, and whether the level-1 effect also has a random slope! Thus, the onus is on the analyst to ensure they know exactly what the software is doing.

## Recap and Recommendations: Centering

This section's examples showed how to create distinct within-level-1 and between-level-2 fixed effects of observed level-1 predictors—this occurs in two-level MLMs whenever a fixed slope for the level-2 mean of a level-1 predictor is included, but in different ways on the basis of how the level-1 predictor is centered. Direct within-level-1 and between-level-2 fixed slopes result in variable-centered level-1 models from using level-specific uncorrelated predictors; the contextual-level-2 effect for their difference can be found as *between – within*. But in constant-centered level-1 models, distinct effects per level are a by-product of controlling for covariance among predictors: For a constant-centered level-1 predictor, its unique effect (after controlling for the level-2 mean predictor) is a purely within-level-1 effect; for the level-2 mean predictor, its unique effect (after controlling for the level-1 predictor) is a contextual-level-2 effect. The between-level-2 effect can be found as *within + contextual*. Similar equivalences can be found across centering strategies for MLMs with interactions so long as the same interactions specified with the level-1 predictor are also specified with the level-2 mean predictor. Fortunately, routines that provide estimates and standard errors of such linear combinations of fixed effects in MLMs, SEMs, or M-SEMs are widely available (e.g., using ESTIMATE in SAS MIXED, TEST in SPSS MIXED, lincom in Stata mixed, general linear hypothesis in HLM, glht or contest1D in R, and NEW in M*plus*). Multivariate MLMs using latent-centering (in software for SEMs or M-SEMs) may offer some advantages over univariate MLMs using observed predictors, the particulars of which remain an active area of research.

In general, smushed fixed slopes can be prevented by including predictors at each level of the model at which they have variability. In three-level nested models, level-1 predictors may need fixed slopes for their level-2 and level-3 means, and level-2 predictors may need fixed slopes for their level-3 means (see Brincks et al. 2017; Hoffman 2015, chapter 11). In cross-classified models, level-1 predictors may need fixed slopes for the level-2 mean of each crossed dimension. Avoiding smushed effects is also important in testing mediation, which can logically only occur for variables at the same level. For instance, consider a 2-1-1 design (Preacher et al. 2010) with a level-2 predictor, a level-1 mediator, and a level-1 outcome. Only between-level mediation is then logical, the indirect effect for which would start at the level-2 predictor, go to the level-2 random intercept for the mediator, and then to the level-2 random intercept for the outcome. In a 1-1-1 design, separate indirect effects would be logical at both the between and the within levels. All multilevel analyses should begin by carefully examining the sources of variability for each variable of interest—regardless of whether it is conceptualized as a predictor or an outcome—to inform at which levels of the model each predictor has the potential to provide a distinct effect.

There are (at least) two reasons why the same predictor can have different effects across levels. The first reason is conceptual—different effects result from different theoretical causes of predictor variability at each level. For instance, the more transient reasons why a patient may differ in

compliance from occasion to occasion are likely not the same reasons why patients differ in compliance on average from each other (e.g., due to stable traits, beliefs, and lifestyle differences). The second reason predictor effects differ across levels is simply because of scaling—fixed slopes are unstandardized estimates scaled in units of the outcome per units of the predictor. To the extent that the standard deviations of either the predictor or the outcome differ across levels, this discrepancy will result in different fixed slope coefficients for their relations at each level (Hoffman 2015, chapter 8).

The insistence on avoiding smushed effects endorsed here continues to gain support (e.g., Antonakis et al. 2021, Bliese et al. 2018, Curran et al. 2012, Hoffman 2019, Preacher et al. 2016, Wang & Maxwell 2015). It does not matter whether a lower-level predictor is of theoretical interest, is merely a control (Rights et al. 2020), or is quantitative or categorical (Yaremych et al. 2021). What matters is whether the lower-level predictor contains systematic higher-level variability. If it does, and only if a lowest-level fixed slope is included, that smushed slope (which forces between-level-2 and within-level-1 fixed slopes to be equal) will likely result in a misspecified level-2 model. Because smushed effects are usually heavily weighted toward the within-level-1 fixed slope (Raudenbush & Bryk 2002, chapter 5), the level-1 model will be less affected. But the model will have the wrong between-level-2 fixed slope, which can bias the random effects variances and thus may result in incorrect inferences for their corresponding fixed effects. Consumers of research using MLMs should be wary whenever results using constant-centered lower-level predictors (that still contain higher-level variability) are reported without separate fixed slopes for their higher-level means, as this assumes no higher-level contextual effects are missing!

What if the between-level-2 and within-level-1 fixed slopes do not differ significantly—then is a single smushed level-1 slope okay to retain? Perhaps so in that particular model, but their tenuous equivalence may be broken upon adding other predictors or interactions, which can create unique effects or conditional main effects that are no longer equivalent across levels. In models with constant-centered lower-level predictors, retaining nonsignificant contextual higher-level fixed slopes can help maintain clarity when interpreting the lower-level fixed slopes. In models with variable-centered lower-level predictors, we suggest between-higher-level fixed slopes be retained when possible as well, although they are not needed to prevent smushed lower-level fixed slopes. Because the higher-level mean variables carry the differential reference point for what is usual with respect to the lower-level predictor (more than a person's own average in longitudinal designs, more than the cluster average in clustered designs), it can be awkward to ignore the effect of variability in this usual reference point in describing the model results.

## ANOTHER RECURRING PLOT LINE: SPECIFICATION OF RANDOM EFFECTS

In the previous sections we focused on fixed slopes and centering strategies by which to ensure their correct interpretations. But the accuracy of the standard errors (SEs) of these fixed slopes (and thus their *p*-values) depends on having the right model for the variance—the set of random effects and residuals needed to accurately capture the unexplained outcome variance and covariance across observations (but without adding unnecessary parameters). But given that in real data one can never know what the right model is, there are a variety of opinions on how to avoid the wrong model. We now turn to recent developments and recommendations on this topic, including how to know when a random intercept or random slope might be needed, how to make sure a random slope is interpretable as intended, and how to provide effect sizes for fixed and random slopes through measures of explained variance (indices of $R^2$ values for MLMs).

## Conflict on the Set: Which Random Effects Are Needed?

Determining the optimal set of random effects for a given design is not always easy, but we offer some guiding principles, first for random intercepts (for mean differences between units) and then for random slopes (for differences between units in the effects of predictors).

A random intercept (that creates a new level) is needed for each sampling dimension in which mean differences remain after accounting for the fixed slopes of predictors. Consider a clustered example of persons nested in one of two treatments (*type*; old = 0, new = 1). A random intercept is not needed here because after including a fixed slope for *type*, the residuals of persons from the same treatment should be independent. But what if there were multiple clusters within each treatment type? After controlling for *type*, mean differences may still remain among clusters of the same *type*, in which case the residuals of persons from the same cluster will still be related. Adding a random intercept to capture cluster mean differences creates a two-level model of level-1 persons in level-2 clusters (with *type* as a level-2 predictor, as in Equation 1).

The same logic applies to cross-classified samples, such as when countries are measured over multiple years but different respondents are used in each year (Schmidt-Catran & Fairbrother 2016). Level-1 persons are nested in level-2 country-years (years from the same country), and country-years are nested within crossed level-3 dimensions of years and countries. Thus, mean differences across years and countries require a level-3 random intercept; mean differences across country-years require a level-2 random intercept. Alternatively, fixed slopes for *year* could replace the level-3 year random intercept, creating a three-level nested model. Likewise, fixed slopes for *country* could replace the level-3 country random intercept, creating a two-level nested model (of level-1 persons in level-2 country-years). Otherwise, if a sampling dimension's mean differences are unaccounted for (a random intercept is forgotten), the SEs of the fixed slopes will be too large or too small, depending on the level(s) at which a predictor contains variability (Berkhof & Kampen 2004, Luo & Kwok 2009, Moerbeek 2004, Van Landeghem et al. 2005). A random intercept for mean differences by combinations of crossed sampling dimensions (e.g., country-years) is needed whenever there is more than one observation per combination (Shi et al. 2010). Custom random intercepts are also needed to address cross-classification created by changing clusters or multiple cluster memberships (Meyers & Beretvas 2006, Ye & Daniel 2017).

Analogously, a random slope is needed for each sampling dimension in which variability in a predictor's slope remains after accounting for the fixed slopes of cross-level interactions. For instance, consider the longitudinal model in Equation 2. After a fixed slope for level-1 $time_{tp}$ accounts for occasion mean differences, level-1 occasions are nested in level-2 patients, which at a minimum should include a level-2 random intercept to capture patient mean differences ($U_{0p}$). But if patients differ in change over time, then a level-2 random time slope across patients ($U_{2p}$) is also needed to ensure correct SEs for fixed slopes of level-1 $time_{tp}$ ($\gamma_{20}$) and its cross-level interaction ($\gamma_{22}$) with patient-level-2 $type_p$. Likewise, a level-2 random slope ($U_{1p}$) is needed for accurate SEs of the fixed slopes of within-level-1 compliance ($\gamma_{10}$) and its cross-level interaction with $type_p$ ($\gamma_{12}$). Otherwise, ignoring sizeable remaining random slope differences will underestimate the SEs for that predictor's level-1 fixed slope and its cross-level interactions (Algina & Swaminathan 2011, Barr et al. 2013, Bell et al. 2019, Heisig & Schaeffer 2019, LaHuis et al. 2020, Ye & Daniel 2017). Parallel concerns for random slope variability in repeated measures experiments are described by Judd et al. (2017) and Brauer & Curtin (2018).

In practice, the need for random effects to capture unexplained variability can be assessed using likelihood ratio tests (LRTs) (the difference in the −2 log-likelihood of nested models; Ke & Wang 2015, Snijders & Bosker 2012). But what if the test indicates that a random slope variance is not needed? This can happen when it is tested initially (i.e., no random slope variance is detectable to

begin with), or it can result from adding cross-level interactions (i.e., no detectable unexplained slope variance remains). In such cases, can one proceed safely without the random slope?

Lower-level effects that vary deterministically via cross-level interactions but not randomly otherwise are called systematically or nonrandomly varying effects (Raudenbush & Bryk 2002, chapter 9). Several authors have suggested these effects are permissible (Aguinis et al. 2013; LaHuis & Ferguson 2009; Snijders & Bosker 2012, chapter 6), given that fixed slopes of cross-level interactions can have greater power than the random slope variances they should have explained. To prevent Type I errors for their level-1 fixed slopes and cross-level interactions, other authors have suggested all random slopes be included by default (Barr et al. 2013, Heisig & Schaeffer 2019). But this so-called maximal model strategy may result in more nonconvergence or lower power than when using LRTs or related fit indices to select which random slopes are needed (Matuschek et al. 2017, Park et al. 2020). As a compromise, using effect size (slope reliability) rather than LRTs as a basis for retaining random slopes in small level-2 samples may prevent inflated Type I error rates for their level-1 fixed slopes and cross-level interactions (L. Hoffman & J. Templin, unpublished manuscript).



## Conflict Resolution: How to Avoid Smushed Random Slopes

As promised, we now turn to the interpretation of random slopes of level-1 predictors. Variable-centering avoids smushed level-1 and level-2 fixed slopes and simplifies interpretation, and it also prevents smushed random slopes, a topic we have ignored so far. As shown in many sources (e.g., Kreft et al. 1995; Preacher et al. 2016; Raudenbush & Bryk 2002, chapter 5; Snijders & Bosker 2012, chapter 5), adding fixed slopes for the level-2 means of level-1 predictors creates equivalent models across level-1 centering strategies for the fixed effects but not for the random effects. For instance, in Equation 2, the random slope term for $L1comp_{tp}$ when variable-centered is $U_{1p}(comp_{tp} - \overline{comp}_p)$, but when constant-centered it is $U_{1p}(comp_{tp} - C_1)$. Because level-2 mean predictor variance is still included in the latter random slope term, the estimated variances and covariances for a model's random effects can differ across level-1 centering strategies.

Raudenbush & Bryk (2002, chapter 5) indicated that discrepancies in the random intercept variance across models could result from different interpretations of what 0 means for the level-1 predictor with the random slope: at the centering variable = 0 (e.g., at the patient-level-2 mean in $L1comp_{tp} = comp_{tp} - \overline{comp}_p$) or at the centering constant = 0 (e.g., at $C_1$ in $L1comp_{tp} = comp_{tp} - C_1$). If the centering constant is not observed for some level-2 units, this could lead to greater shrinkage to the mean for their random intercepts, which could homogenize their random slopes. In support of this conjecture, in their example Raudenbush & Bryk (2002, p. 144) found a smaller random slope variance in their constant-centered level-1 models than in their variable-centered level-1 models (a pattern also reported anecdotally by Hoffman 2012). Raudenbush & Bryk recommended variable-centering all lower-level predictors to be given random slopes, especially those with large mean variation.

An alternative explanation for the discrepancy in random slope variances across level-1 centering strategies was provided by Rights and Sterba (J.D. Rights & S.K. Sterba, unpublished manuscript). They showed that the discrepancy can result from smushed random slopes in the constant-centered level-1 model, a problem parallel to but separate from that of smushed fixed slopes described above. More specifically, a random slope predicts outcome variance with a quadratic form of heterogeneity along its predictor (Hedeker & Gibbons 2006, chapter 4). For example, in the longitudinal model in Equation 2, the total variance (holding $time_{tp} = 0$) at a given $L1comp$ is predicted to be $Var|L1comp = \sigma_e^2 + \tau_{U_0}^2 + 2\tau_{U_{01}}(L1comp) + \tau_{U_1}^2(L1comp)^2$. If a level-2 random slope variance is estimated for variable-centered $L1comp_{tp}$, this quadratic variance heterogeneity applies only to within-level-1 compliance, and the level-2 random slope variance $\tau_{U_1}^2$

contains only between-level-2 differences in the slope of within-level-1 compliance as intended. But when estimating a level-2 random slope variance for constant-centered $L1comp_{tp}$ instead, this forces the same variance heterogeneity for within-level-1 compliance as for between-level-2 mean compliance because both variables are still included in constant-centered $L1comp_{tp}$. The level-2 random slope variance $\tau^2_{U_1}$ is then smushed—its estimate conflates between-level-2 differences in the slope of within-level-1 compliance with quadratic variance heterogeneity for between-level-2 mean compliance. Thus, $\tau^2_{U_1}$ will be inaccurate whenever these sources of variance differ in reality.

Rights and Sterba showed how to resolve the discrepancy between models by adding a second random slope for the level-2 mean predictor, e.g., $U_{3p}(L2comp_p)$ (J.D. Rights & S.K. Sterba, unpublished manuscript). When added to the constant-centered level-1 model, it serves as a contextual-level-2 random slope that allows differential quadratic variance heterogeneity for the level-2 mean predictor. When added to the variable-centered level-1 model, it introduces between-level quadratic variance heterogeneity for the level-2 mean predictor. But given that level-2 terms related to variance heterogeneity can be difficult to estimate, the easiest way to prevent smushed random slopes is to estimate them using a variable-centered lower-level predictor. We note that a constant-centered lower-level predictor could still be used to estimate the fixed slope (and fixed cross-level interactions) for a desired interpretation (i.e., to create contextual-level-2 fixed slopes as direct parameters or when using lower-level predictors that already have a natural value of 0, such as binary variables).

## Bonus Material: Effect Sizes Using Explained Variance

A proper specification of random intercepts and unconfounded random slopes is important not only for fixed slope inference but also for using explained variance ($R^2$) indices of effect size. Multiple sources of variance are quantified and explained simultaneously in MLMs, leading to multiple $R^2$ values. Early attempts to create an overall $R^2$ that aggregated the random intercept and residual variances (Snijders & Bosker 2012, chapter 7) were not easily extendable to models with random slopes. An alternative total-$R^2$ (Hoffman 2015, chapter 7; Singer & Willett 2003, chapter 4) can be created by the square of the correlation between the actual outcome and the model-predicted outcome (using only the fixed effects). But total-$R^2$ does not inform how well the model fixed effects have explained each source of variability. For that purpose, one can also compute separate pseudo-$R^2$ indices for the proportion of explained variance for each source—random intercepts, random slopes, and residuals—in a target model relative to a null model without the fixed slopes of interest (Hoffman 2015, chapter 7; Raudenbush & Bryk 2002, chapter 4; Singer & Willett 2003, chapter 4). Unfortunately, pseudo-$R^2$ values can become negative, such as for predictors with small effects, which can result in confusion and less precision in practice.

An alternative strategy was introduced by Rights & Sterba (2019, 2020, 2021), who provided formulas (and R packages) for computing $R^2$ values that use model-implied total variance rather than null model variance. This strategy eliminates negative $R^2$ values while still following the same logic for which source of variance should be reduced by fixed slopes at each level as given in the MLM use cases above. Their $R^2$ indices include total and level-specific marginal versions (for fixed slopes only) and conditional versions (that also include random slopes, which seem less useful given that random slopes are actually unexplained differences across units in the slopes of predictors). However, they do not provide an $R^2$ for the random slope variance explained by cross-level interactions, a quantity that is frequently of interest.

## Recap: Random Effects

Our goal in the previous section was to describe how careful attention to the presence and contents of MLM variance components (higher-level random intercept and slopes and lowest-level

residuals) can help capture all sampling-related dependency, thereby protecting our inferences about the significance and effect size of our model predictors. Given that we are largely caught up to date on the MLM show, it is now time to introduce a more current storyline.

## CURRENT EPISODES: MIXED-EFFECTS LOCATION–SCALE MODELS

In each of our examples so far, we have implicitly assumed constant (also known as homogeneous) variances—that the same estimated variance components (e.g., residual variance, random effect variances) apply equally to all observations. Like any model assumption, this one is testable, and such testing also introduces the potential for a whole new type of research question. We now consider a new extension of MLMs for predicting differences in the extent of variability.

For instance, why might clusters differ in the similarity of their members? Why do persons differ in the inconsistency of their behavior over time? Such questions about the presence of (and reasons for) heterogeneity of variance across sampling units can be answered using mixed-effects location–scale models (MELSMs), whose purpose is to extend MLMs to concurrently quantify and predict differences in the amounts of outcome variability. Although MELSMs have existed for decades (Cleveland et al. 2000, Foulley & Quaas 1995, Lee & Nelder 2006), recent efforts have made their estimation and practical use more feasible (Hedeker & Nordgren 2013; Hedeker et al. 2008, 2009; Leckie et al. 2014; Lester et al. 2021; Rast et al. 2012; Walters et al. 2018). Consequently, we see MELSMs as an important new storyline in the future of MLMs.

A MELSM contains fixed and random effects in its location model, as well as in multiple possible scale models. The location model is the traditional MLM as described previously in Seasons 1 and 2 (Equations 1 and 2, respectively), and it quantifies and explains distinct sources of outcome variability across sampling units. In contrast, a scale model quantifies and explains differences between sampling units in the amounts of outcome variability. MLMs with only location models assume that the same variance components apply to all sampling units (an assumption known as homogeneity of variance or homoscedasticity). MELSMs relax these assumptions through scale models that create heterogeneity of each variance component as a function of fixed and random effects of predictor variables (which may or may not also be included in the location model). In explaining amounts of variance, scale models can include predictors only at the same level or higher. Further, because variances have a lower bound at 0, scale models ensure that all predicted variances remain positive by using nonnegative distributions, such as the lognormal or gamma distributions, or nonlinear transformations of the predicted variances using link functions, such as the log link (Culpepper 2010, Hamaker & Klugkist 2011, Hedeker et al. 2008, Lee & Noh 2012, Rast et al. 2012, Wang et al. 2012). MELSMs are applicable to both clustered and longitudinal designs, as illustrated in the next two sections.

**Mixed-effects location–scale model (MELSM):** a statistical model that extends the multilevel model by relaxing the assumption of homogeneity of variance and allowing fixed and random effects to predict both the amount and the variability of an outcome

**Location model:** the part of the mixed-effects location–scale model that allows quantification and prediction of between-unit differences in the mean outcome and in the effects of predictors (like the traditional multilevel model)

**Scale model:** the part of the mixed-effects location–scale model that allows quantification and prediction of between-unit differences in the variability (dispersion) of an outcome or in the effects of predictors (unlike the traditional multilevel model)

### Season 5: MELSMs for Clustered Samples

We can build a clustered MELSM by returning to Equation 1 (from Season 1), in which $anxiety_{pc}$ is the outcome for level-1 patient $p$ from level-2 clinic $c$, $comp_{pc}$ is patient-level-1 compliance that has been variable-centered using the clinic mean into $L1comp_{pc}$ and $L2comp_c$, and $type_c$ is a binary clinic treatment indicator. In Season 1, we had assumed that the level-1 residual $e_{pc}$ had constant variance $\sigma_e^2$ over patients and clinics (shown in **Figure 1** by the equal dispersion of circles around the same-colored line for each clinic) and that the level-2 random effects $U_{0c}$ and $U_{1c}$ (for between-clinic variability in the intercept and within-clinic slope of patient compliance, respectively) had constant variances $\tau_{U_0}^2$ and $\tau_{U_1}^2$ over clinics. This is no longer a requirement.

Using Equation 1 as our MELSM location model, we can add scale models to predict different amounts of each variance. For instance, what if there are fewer differences between patients from

the same clinic (as captured by the level-1 residual variance in the location model) in patients with better compliance or who attend clinics using the old treatment? To test these hypotheses, a scale model for the level-1 residual variance is shown in Equation 3,

$$\begin{aligned} &\text{Level 1 } (L1): \log_e\left(\sigma^2_{e_{pc}}\right) = T_{0c} + T_{1c}\left(L1comp_{pc}\right) \\ &\text{Level 2 } (L2): T_{0c} = \tau_{00} + \tau_{01}\left(L2comp_c\right) + \tau_{02}\left(type_c\right) + \omega_{0c}, \\ &\qquad\qquad T_{1c} = \tau_{10} \end{aligned} \qquad 3.$$

in which $\log_e(\sigma^2_{e_{pc}})$ is the natural log of the level-1 residual variance for level-1 patient $p$ in level-2 clinic $c$. Rather than predicting the level-1 anxiety outcome, Equation 3 predicts the amount of level-1 residual variance for each patient and clinic, and it does so on the natural log scale to ensure the predicted variances remain positive, as they should be.

The scale-model fixed and random effects have interpretations similar to those in the location model in Equation 1 (in Season 1). In this level-1 model, $T_{0c}$ holds the overall clinic intercept and $T_{1c}$ holds the overall clinic slope of patient-level-1 compliance (analogous to location-model $\beta$ terms). Each $T$ clinic outcome is defined in a level-2 model using $\tau$ fixed effects (analogous to location-model $\gamma$ terms) and possibly $\omega$ random effects that vary over clinics (analogous to location-model $U$ terms). In predicting the level-2 $T_{0c}$ clinic intercept, $\tau_{00}$ is the fixed intercept for the expected log level-1 residual variance when all predictors equal 0, $\tau_{01}$ and $\tau_{02}$ are the between-clinic fixed slopes of the clinic-level-2 predictors, and $\omega_{0c}$ is the scale-model level-2 random intercept (the deviation of the actual intercept from the predicted clinic intercept with variance $\tau^2_{\omega_0}$). Analogous to the location-model level-2 random intercept variance $\tau^2_{U_0}$, the scale-model level-2 random intercept variance $\tau^2_{\omega_0}$ captures between-clinic differences in log level-1 residual variance that remain after controlling for the clinic-level-2 predictors. In predicting the level-2 $T_{1c}$ clinic slope of patient-level-1 compliance, $\tau_{10}$ is the scale-model fixed slope for within-level-1 compliance. We also could have included a scale-model level-2 random slope for within-level-1 compliance, which would have allowed unexplained differences across clinics in the effect of greater compliance on log level-1 residual variance than other patients in the same clinic (Leckie et al. 2014, Lester et al. 2021).

Example results for the model in Equation 3 are displayed in **Figure 4**, in which the prediction of patient anxiety is shown for five example patients in each of four clinics. Instead of focusing on differences in the outcome per unit change in compliance or clinic treatment type (as were defined by the location model in Equation 1 and illustrated in **Figure 1**), the scale model for the log level-1 residual variance focuses on the dispersion of patient anxiety outcomes around their clinic-predicted anxiety (shown in **Figure 4** as the spread of the circles around their same-colored lines). The scale-model fixed intercept $\tau_{00} = -1$ is the expected log level-1 residual variance for a patient with compliance at their clinic's mean ($L1comp_{pc} = 0$) in a clinic with reference mean compliance ($L2comp_c = 0$) that uses the old treatment ($type_c = 0$), represented in **Figure 4** by the amount of dispersion of the third dark green circle from the dark green line.

Focusing on the fixed slopes, the $type_c$ slope $\tau_{02} = 1$ is the between-clinic difference in log level-1 residual variance for clinics that use the new treatment ($type_c = 1$) instead of the old treatment ($type_c = 0$). This effect is shown in **Figure 4** by the greater dispersion of circles around their same-colored lines for clinics with $type_c = 1$ than for clinics with $type_c = 0$ (holding $L2comp_c$ constant). In addition, because patient compliance was clinic-mean-centered for each patient, the fixed effects for $L1comp_{pc}$ and $L2comp_c$ represent level-specific effects on log level-1 residual variance. The $L1comp_{pc}$ slope $\tau_{10} = -0.25$ indicates that log level-1 residual variance is lower by 0.25 per unit higher patient compliance beyond the clinic's mean: Anxiety in more compliant patients is more consistent (less variable). This effect is shown in **Figure 4** by circles closer to their same-colored line for patients with higher compliance (higher $x$-axis values). The $L2comp_c$ slope

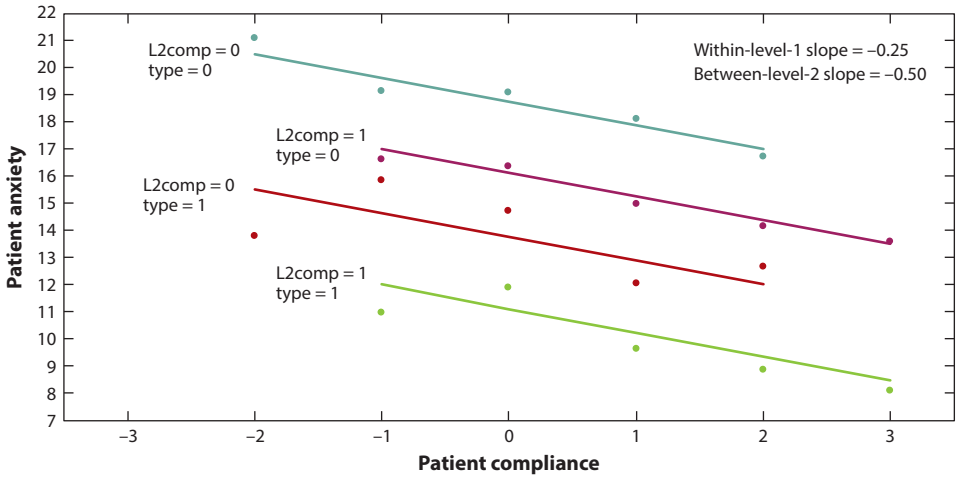


**Figure 4**

Example results for the mixed-effects location–scale model for clustered data in Equation 3 for four clinics each with five patients. Circles represent an individual patient outcome and lines represent the clinic-specific slopes across patient outcomes. The scale model for the level-1 residual variance is shown, with residual variance represented by the dispersion of the circles around the lines.

$\tau_{01} = -0.5$ indicates that log level-1 residual variance is lower by 0.5 per unit higher clinic mean compliance: Patient anxiety is less variable in clinics with more compliant patients overall. This effect is shown in **Figure 4** by circles that are closer to their same-colored lines for clinics with $L2comp_c = 1$ than for clinics with $L2comp_c = 0$ (holding $type_c$ constant).

In addition to (or separate from) the scale model for the level-1 residual variance, what if there are fewer differences in average anxiety between clinics (as captured by the location-model level-2 random intercept) that use the old treatment? To test this hypothesis, a second scale model—this time for the location-model level-2 random intercept variance—is shown in Equation 4,

$$\text{Level 2 } (L2) : \log_e\left(\tau^2_{U_{0c}}\right) = \alpha_{00} + \alpha_{01}\,(type_c), \qquad 4.$$

in which $\log_e(\tau^2_{U_{0c}})$ is the natural log location-model level-2 random intercept variance for clinic $c$. This scale model focuses on the dispersion of each clinic's mean anxiety around its treatment type mean (specifically when patient compliance is at the clinic's mean, given that the location-model random intercept variance is conditional on the location-model random slope for clinic-mean-centered patient compliance equaling 0; see Equation 1). As an example, a scale-model fixed intercept $\alpha_{00} = -0.5$ would be the expected log location-model level-2 random intercept variance for clinics using the old treatment, and the $type_c$ slope $\alpha_{01} = 0.75$ would indicate that log location-model level-2 random intercept variance is higher by 0.75 in clinics using the new treatment. As such, clinics using the new treatment would differ more from each other in average anxiety than from clinics using the old treatment. Further, although it would substantially increase model complexity, it would also be possible for each location-model random effect variance and covariance to have their own scale-model fixed effects as well (Leckie et al. 2014).

## Season 6: MELSMs for Longitudinal Samples

We can build a longitudinal MELSM by returning to Equation 2 (from Season 2), in which $anxiety_{tp}$ is the outcome for level-1 occasion $t$ from level-2 patient $p$, $time_{tp}$ is level-1

time-in-treatment, $comp_{tp}$ is level-1 time-varying patient compliance that has been variable-centered using the patient mean into $L1comp_{tp}$ and $L2comp_p$, and $type_p$ is the binary treatment indicator for patient $p$. In Season 2 we assumed that the level-1 residual $e_{tp}$ had constant variance $\sigma_e^2$ over occasions and patients (shown in **Figure 2*a*,*b*** as equal dispersion of circles around their same-colored lines for each patient) and that the level-2 random effects $U_{0p}$, $U_{1p}$, and $U_{2p}$ (for between-patient variability in the intercept and for between-patient slope variability in the within-patient effects of compliance and time-in-treatment, respectively) had constant variances $\tau_{U_0}^2$, $\tau_{U_1}^2$, and $\tau_{U_2}^2$ over patients. Below, we remove some of these restrictions to the model.

Using Equation 2 as our MELSM location model, we can add scale models to predict differences in the amounts of each variance. For instance, what if there is less fluctuation over time in anxiety (as captured by the level-1 residual variance in the location model) given longer time-in-treatment or better compliance, or in patients who were given the old treatment? To test these hypotheses, a scale model for the level-1 residual variance is shown in Equation 5,

$$\begin{aligned}
&\text{Level 1 } (L1): \log_e\left(\sigma_{e_{tp}}^2\right) = T_{0p} + T_{1p}\left(L1comp_{tp}\right) + T_{2p}\left(time_{tp}\right)\\
&\text{Level 2 } (L2): T_{0p} = \tau_{00} + \tau_{01}\left(L2comp_p\right) + \tau_{02}\left(type_p\right)\\
&\qquad\qquad + \tau_{03}\left(L2comp_p\right)\left(type_p\right) + \omega_{0p}\\
&\qquad\quad T_{1p} = \tau_{10} + \tau_{12}\left(type_p\right)\\
&\qquad\quad T_{2p} = \tau_{20} + \tau_{22}\left(type_p\right)
\end{aligned} \quad , \qquad 5.$$

in which $\log_e(\sigma_{e_{tp}}^2)$ is the natural log of the level-1 residual variance for level-1 occasion $t$ in level-2 patient $p$. Rather than predicting the level-1 anxiety outcome, Equation 5 predicts the amount of level-1 residual variance for each occasion and patient on the natural log scale to ensure the predicted variances remain positive, as they should be.

The scale-model fixed and random effects have interpretations similar to those in the location model in Equation 2 from Season 2. In this level-1 model, $T_{0p}$ holds the overall intercept, whereas $T_{1p}$ and $T_{2p}$ hold the overall slope of time-level-1 compliance and time-in-treatment (analogous to location-model $\beta$ terms). Each $T$ patient outcome is defined in a level-2 model using $\tau$ fixed effects (analogous to location-model $\gamma$ terms) and possibly $\omega$ random effects that vary over patients (analogous to location-model $U$ terms). In predicting the level-2 $T_{0p}$ patient intercept, $\tau_{00}$ is the fixed intercept; $\tau_{01}$, $\tau_{02}$, and $\tau_{03}$ are the scale-model fixed slopes of the patient-level-2 predictors; and $\omega_{0p}$ is the scale-model level-2 random intercept (the deviation of the actual intercept from the predicted patient intercept with variance $\tau_{\omega_0}^2$). Analogous to the location-model level-2 random intercept variance $\tau_{U_0}^2$, the scale-model level-2 random intercept variance $\tau_{\omega_0}^2$ captures between-patient differences in log level-1 residual variance that remain after controlling for the patient-level-2 predictors. In predicting the level-2 $T_{1p}$ slope of time-level-1 compliance, $\tau_{10}$ is the within-patient fixed slope for patients given the old treatment and $\tau_{12}$ is the difference in the within-patient compliance slope for patients given the new treatment. In predicting the $T_{2p}$ slope of time-in-treatment, $\tau_{20}$ is the change in log level-1 residual variance at each subsequent occasion for patients given the old treatment and $\tau_{22}$ is the time slope difference for patients given the new treatment. And, as if this model were not already complex enough, we could have added scale-model level-2 random slopes for the within-level-1 effect of compliance and/or time-in-treatment (which would have allowed unexplained differences across level-2 patients in the effect on log level-1 residual variance of greater compliance than usual and/or time-in-treatment; see Rast et al. 2012).

Example results for the model in Equation 5 are displayed in **Figure 5**, in which the prediction of patient anxiety is shown at five occasions for one patient in each of four clinics. **Figure 5*a*** depicts the effect of time-level-1 compliance specifically at $time_{tp} = 4$, whereas **Figure 5*b*** depicts change over time-in-treatment holding $L1comp_{tp} = 0$. Instead of focusing on differences in anxiety per

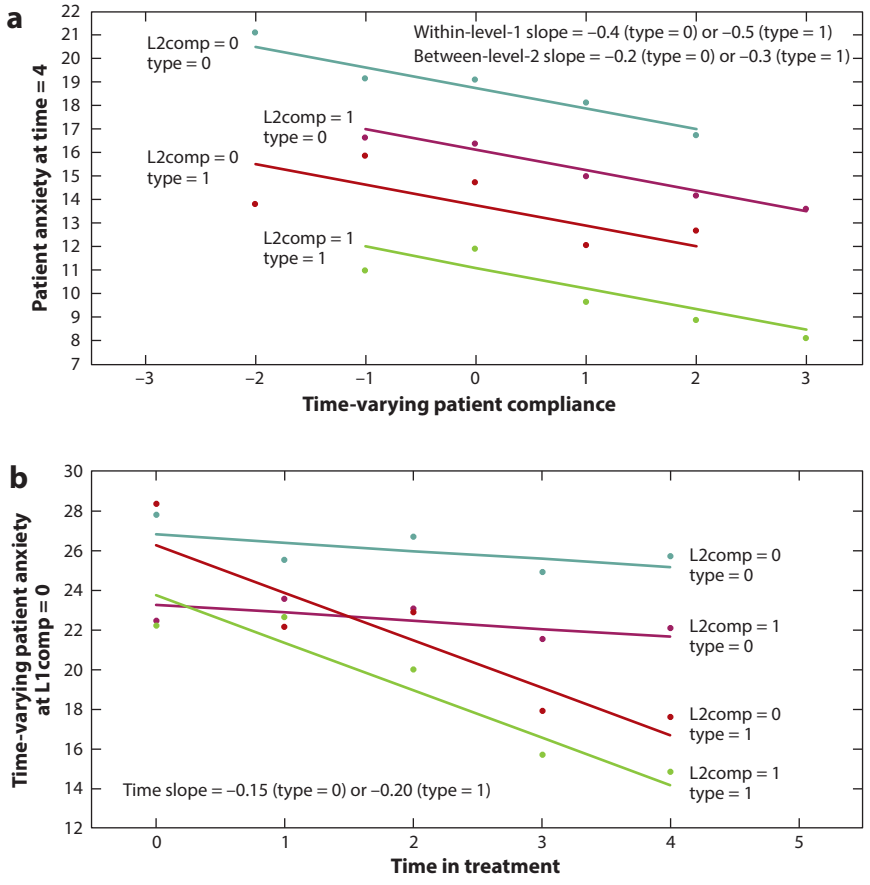


**Figure 5**

Example results for the mixed-effects location–scale model for longitudinal data in Equation 5 for four individual patients each with five occasions. Circles represent within-patient outcomes and lines represent the patient-specific slopes across their outcomes. The scale model for the level-1 residual variance is shown for (*a*) time-varying patient compliance and (*b*) time-in-treatment, with residual variance represented by the dispersion of the circles around the lines.



unit change in time-in-treatment, compliance, or treatment type (as were defined by the location model in Equation 2 and shown in **Figure 2**), the scale model for the level-1 residual variance focuses on the dispersion of level-1 residuals around predicted anxiety, as shown in **Figure 5*a*,*b*** by the dispersion of the circles around their same-colored lines.

The scale-model fixed intercept $\tau_{00} = -0.2$ is the expected log level-1 residual variance specifically at the first occasion ($time_{tp} = 0$) for a patient with compliance at their own mean ($L1comp_{tp} = 0$) who is at the reference patient mean compliance ($L2comp_p = 0$) and was given the old treatment ($type_p = 0$), represented in **Figure 5*b*** by the amount of dispersion of the first dark green circle from the dark green line. The $type_p$ slope $\tau_{02} = 0.75$ is the between-patient difference in log level-1 residual variance for patients given the new treatment instead of the old treatment (for $L2comp_p = 0$, $L1comp_{tp} = 0$, and $time_{tp} = 0$, given their interactions with $type_p$, as described below). This $type_p$ effect is shown in **Figure 5*b*** by the greater dispersion of the first brown circle compared with that of the first dark green circle.

Because patient compliance was patient-mean-centered at each occasion, the scale-model fixed effects for $L1comp_{tp}$ and $L2comp_p$ represent level-specific effects on log level-1 residual variance, each of which is moderated by $type_p$. The $L1comp_{tp}$ slope $\tau_{10} = -0.4$ indicates that, for patients given the old treatment, log level-1 residual variance is lower by 0.4 per unit higher patient compliance than the patient's mean; the $L1comp_{tp}$ by $type_p$ cross-level interaction indicates that this effect is more negative (stronger) by $\tau_{12} = -0.1$ in patients given the new treatment (whose $L1comp_{tp}$ slope is then $\tau_{10} + \tau_{12} = -0.4 - 0.1 = -0.5$). As such, anxiety is more consistent (less variable) when patients are more compliant than usual, more so in patients given the new treatment, as shown in **Figure 5*a*** by the greater reduction in dispersion as time-varying compliance increases from –1 to 2 for $type_p = 1$ compared with that for $type_p = 0$. Similarly, the $L2comp_p$ slope $\tau_{01} = -0.2$ indicates that, for patients given the old treatment, the log level-1 residual variance is lower by 0.2 per unit higher patient mean compliance; the $L2comp_p$ by $type_p$ level-2 interaction indicates that this effect is more negative (stronger) by $\tau_{03} = -0.1$ in patients given the new treatment (whose $L2comp_p$ slope is then $\tau_{01} + \tau_{03} = -0.2 - 0.1 = -0.3$). Thus, anxiety is more similar for patients with greater average compliance across occasions, more so for patients given the new treatment (as shown in **Figure 5*a*** by the smaller dispersions of circles for $L2comp_p = 1$ than for $L2comp_p = 0$) and slightly more so for patients with $type_p = 1$ instead of $type_p = 0$.

The level-1 time slope $\tau_{20} = -0.15$ (also moderated by treatment type) indicates that log level-1 residual variance decreases by an average of 0.15 per occasion in patients given the old treatment; the $time_{tp}$ by $type_p$ cross-level interaction indicates that this effect is more negative (stronger) by $\tau_{21} = -0.05$ in patients given the new treatment (whose $time_{tp}$ slope is then $\tau_{20} + \tau_{22} = -0.15 - 0.05 = -0.20$). Thus, anxiety becomes more consistent with each subsequent occasion, as shown in **Figure 5*b*** by circles closer to their same-colored lines at later occasions, more so in patients given the new treatment instead of the old treatment.

In addition to (or separate from) the scale model for the level-1 residual variance, what if there are fewer differences in average anxiety between patients (as captured by the location-model level-2 random intercept) who were given the new treatment? To test this hypothesis, a second scale model—this time for the location-model level-2 random intercept variance—is shown in Equation 6,

$$\text{Level 2 } (L2): \ \log_e\left(\tau^2_{U_{0p}}\right) = \alpha_{00} + \alpha_{01}\left(type_p\right), \qquad 6.$$

in which $\log_e(\tau^2_{U_{0p}})$ is the natural log of the location-model level-2 random intercept variance for patient $p$. This scale model focuses on the dispersion of each patient's mean anxiety around their treatment type mean (specifically at time 0 and when time-varying compliance is at the patient's mean, given that the location-model random intercept variance is conditional on the location-model random slopes for both time and patient-mean-centered compliance; see Equation 2). As an example, the scale-model fixed intercept $\alpha_{00} = -0.10$ would be the expected log location-model level-2 random intercept variance for patients given the old treatment, and the $type_p$ slope $\alpha_{01} = -0.15$ would indicate that log location-model level-2 random intercept variance is lower by 0.15 units in patients given the new treatment. As such, patients given the new treatment would have less variable levels of anxiety (at time 0 and when patient compliance is at the patient's mean) than patients given the old treatment. Finally, we note that each location-model random slope variance and covariance could also have their own scale model (Rast et al. 2012, Walters 2015).

## Recap and Spin-Offs: MELSMs

In this section, we described MELSMs by which to formally test variability-related hypotheses. MELSMs extend traditional MLMs (as location models) to concurrently estimate separate scale

models that allow the level-1 residual variance and/or level-2 random effect variance(s) to be heterogeneous across sampling units as a function of predictors. Scale-model random effects can quantify between-cluster differences in within-cluster variance in clustered designs; they can also quantify between-person differences in within-person fluctuation over time in longitudinal designs. Scale-model random effects can have covariances with location-model random effects (or directed relationships can be specified between them, depending on software).

In our examples, both the clustered models and the longitudinal models were specific to continuous, conditionally normally distributed outcomes with fixed effects that had a linear functional form. However, MELSMs are more broadly applicable, such as for ordinal outcomes (Hedeker et al. 2016), time-to-event and censored outcomes (Courvoisier et al. 2019, Lu 2018), semicontinuous outcomes (i.e., with excess zeros; Blozis et al. 2020), or random effects (Ma et al. 2021). MELSMs have also been extended to dyadic data (Rast & Ferrer 2018), cross-classified designs (Brunton-Smith et al. 2017), three-level designs (Li & Hedeker 2012, Lin et al. 2018), and to include nonlinear fixed effects (Bürkner 2018, Williams et al. 2019). Finally, MELSMs can also be estimated using latent-centering in M-SEMs (McNeish 2021), which enables scale-model random effects to become predictors or outcomes as needed.

## CONCLUSION

Catching up on a series as long running as multilevel modeling is not an easy task. But we hope our readers have gained the background and encouragement necessary to believe that keeping up with future seasons of MLMs will be worth their while!

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## LITERATURE CITED


Aguinis H, Gottfredson RK, Culpepper SA. 2013. Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *J. Manag.* 39(6):1490–528

Algina J, Swaminathan H. 2011. Centering in two-level nested designs. In *Handbook of Advanced Multilevel Analysis*, ed. J Hox, JK Roberts, pp. 285–312. New York: Taylor & Francis

Antonakis J, Bastardoz N, Rönkkö M. 2021. On ignoring the random effects assumption in multilevel models: review, critique, and recommendations. *Organ. Res. Methods* 24(4):443–83

Asparouhov T, Muthén B. 2019. Latent variable centering of predictors and mediators in multilevel and time-series models. *Struct. Equ. Model.* 26(1):119–42

Barr DJ, Levy R, Scheepers C, Tily HJ. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Memory Lang.* 68(3):255–78

Bauer DJ. 2003. Estimating multilevel linear models as structural equation models. *J. Educ. Behav. Stat.* 28(2):135–67

Bell A, Fairbrother M, Jones K. 2019. Fixed and random effects models: making an informed choice. *Qual. Quant.* 53:1051–74

Berkhof J, Kampen JK. 2004. Asymptotic effect of misspecification in the random part of the multilevel model. *J. Educ. Behav. Stat.* 29(2):201–18

Berry D, Willoughby M. 2017. On the practical interpretability of cross-lagged panel models: rethinking a developmental workhorse. *Child Dev.* 88(4):1186–206

Blalock HM. 1984. Contextual-effects models: theoretical and methodological issues. *Annu. Rev. Sociol.* 10:353–72

Bliese PD, Maltarich MA, Hendricks JL. 2018. Back to basics with mixed-effects models: nine take-away points. *J. Bus. Psychol.* 33:1–23

Blozis SA, McTernan M, Harring JR, Zheng Q. 2020. Two-part mixed-effects location scale models. *Behav. Res. Methods* 52:1836–47

Brauer M, Curtin JJ. 2018. Linear mixed-effects models and the analysis of nonindependent data: a unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychol. Methods* 23(3):389–411

Brincks AM, Enders CK, Llabre MM, Bulotsky-Shearer RJ, Prado G, Feaster DJ. 2017. Centering predictor variables in three-level contextual models. *Multivar. Behav. Res.* 52(2):149–63

Brunton-Smith I, Sturgis P, Leckie G. 2017. Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location-scale model. *J. R. Stat. Soc. Ser. A* 180:551–68

Bürkner PC. 2018. Advanced Bayesian multilevel modeling with the R package brms. *R J.* 10(1):395–411

Burstein L. 1980. The analysis of multilevel data in educational research and evaluation. *Rev. Res. Educ.* 8:158–233

Cleveland W, Denby L, Liu C. 2000. Random location and scale effects: model building methods for a general class of models. *Comput. Sci. Stat.* 32:3–10

Courvoisier D, Walls TA, Cheval B, Hedeker D. 2019. A mixed-effects location scale model for time-to-event data: a smoking behavior application. *Addict. Behav.* 94:42–49

Culpepper SA. 2010. Studying individual differences in predictability with gamma regression and nonlinear multilevel models. *Multivar. Behav. Res.* 45(1):153–85

Curran PJ, Bauer DJ. 2011. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu. Rev. Psychol.* 62:583–619

Curran PJ, Howard AL, Bainter SA, Lane ST, McGinley JS. 2014. The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. *J. Consult. Clin. Psychol.* 82(5):879–94

Curran PJ, Lee T, Howard AL, Lane S, MacCallum R. 2012. Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In *Advances in Longitudinal Methods in the Social and Behavioral Sciences*, ed. JR Harring, GR Hancock, pp. 217–53. Charlotte, NC: Inf. Age Publ.

Cushing CC, Walters RW, Hoffman L. 2014. Aggregated N-of-1 randomized controlled trials: modern data analytics applied to a clinically valid method of intervention effectiveness. *J. Pediatr. Psychol.* 39(2):138–50

Eckardt R, Yammarino FJ, Dionne SD, Spain SM. 2021. Multilevel methods and statistics: the next frontier. *Organ. Res. Methods* 24(4):187–218

Enders CK. 2013. Centering predictors and contextual effects. See Scott et al. 2013, pp. 89–107

Enders CK, Tofighi D. 2007. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol. Methods* 12(2):121–38

Estrada E, Hamagami F, Ferrer E. 2020. Estimating age-based developmental trajectories using latent change score models based on measurement occasion. *Multivar. Behav. Res.* 55(3):454–77

Foulley JL, Quaas RL. 1995. Heterogeneous variances in Gaussian linear mixed models. *Genet. Sel. Evol.* 27:211–28

Grund S, Lüdtke O, Robitzsch A. 2019. Missing data in multilevel research. See Humphrey & LeBreton 2019, pp. 365–86

Hamaker EL, Klugkist I. 2011. Bayesian estimation of multilevel models. In *European Association for Methodology Series. Handbook for Advanced Multilevel Analysis*, ed. JJ Hox, JK Roberts, pp. 137–61. Oxfordshire, UK: Routledge/Taylor & Francis Group

Hamaker EL, Kuiper RM, Grasman RPPP. 2015. A critique of the cross-lagged panel model. *Psychol. Methods* 20(1):102–16

Hamaker EL, Muthén B. 2020. The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychol. Methods* 25(3):365–79

Hedeker D, Gibbons RD. 2006. *Longitudinal Data Analysis*. Hoboken, NJ: Wiley-Interscience

Hedeker D, Mermelstein RJ, Berbaum ML, Campbell RT. 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

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(2):627–34
Hedeker D, Mermelstein RJ, Demirtas H, Berbaum ML. 2016. A mixed-effects location-scale model for ordinal questionnaire data. *Health Serv. Outcomes Res. Methodol.* 16(3):117–31
Hedeker D, Nordgren R. 2013. MIXREGLS: a program for mixed-effects location scale analysis. *J. Stat. Softw.* 52(12):1–38
Heisig JP, Schaeffer M. 2019. Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *Eur. Sociol. Rev.* 35(2):258–79
Hoffman L. 2012. Considering alternative metrics of time: Does anybody really know what "time" is? In *Advances in Longitudinal Methods in the Social and Behavioral Sciences*, ed. JR Harring, GR Hancock, pp. 255–87. Charlotte, NC: Inf. Age Publ.
Hoffman L. 2015. *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change*. Oxfordshire, UK: Routledge/Taylor & Francis Group
Hoffman L. 2019. On the interpretation of parameters in multivariate multilevel models across different combinations of model specification and estimation. *Adv. Methods Pract. Psychol. Sci.* 2(3):288–311
Hoffman L. 2021. Disaggregating associations of between-person differences in change over time from associations of within-person fluctuation in longitudinal data. PsyArXiv. **https://doi.org/10.31234/osf.io/qtc7r**
Hoffman L, Stawski RS. 2009. Persons as contexts: evaluating between-person and within-person effects in longitudinal analysis. *Res. Hum. Dev.* 6(2–3):97–210
Hofmann DA, Gavin MB. 1998. Centering decisions in hierarchical linear models: implications for research in organizations. *J. Manag.* 24(5):623–41
Hox J. 2010. *Multilevel Analysis: Techniques and Applications*. New York: Routledge Academic. 2nd ed.
Hox J, Roberts K. 2011. Multilevel analysis: where we were and where we are. In *Handbook of Advanced Multilevel Analysis*, ed. J Hox, JK Roberts, pp. 3–14. New York: Taylor & Francis
Humphrey SE, LeBreton JM, eds. 2019. *The Handbook of Multilevel Theory, Measurement, and Analysis*. Washington, DC: Am. Psychol. Assoc.
Judd CM, Westfall J, Kenny DA. 2017. Experiments with more than one random factor: designs, analytic models, and statistical power. *Annu. Rev. Psychol.* 68:601–25
Ke Z, Wang L. 2015. Detecting individual differences in change: methods and comparisons. *Struct. Equ. Model.* 22(3):382–400
Kreft IGG, de Leeuw J, Aiken LS. 1995. The effect of different forms of centering in hierarchical linear models. *Multivar. Behav. Res.* 30(1):1–21
LaHuis DM, Ferguson MW. 2009. The accuracy of significance tests for slope variance components in multilevel random coefficient models. *Organ. Res. Methods* 12(3):418–35
LaHuis DM, Jenkins DR, Hartman MJ, Hakoyama S, Clark PC. 2020. The effects of misspecifying the random part of multilevel models. *Methodology* 16(3):224–40
Leckie G, French R, Charlton C, Browne W. 2014. Modeling heterogeneous variance-covariance components in two-level models. *J. Educ. Behav. Stat.* 39(5):307–32
Lee Y, Nelder JA. 2006. Double hierarchical generalized linear models. *Appl. Stat.* 55(2):139–85
Lee Y, Noh M. 2012. Modelling random effect variance with double hierarchical generalized linear models. *Stat. Model.* 12(6):487–502
Lester HF, Cullen-Lester KL, Walters RW. 2021. From nuisance to novel research questions: using multilevel models to predict heterogeneous variances. *Organ. Res. Methods* 24(4):342–88
Li X, Hedeker D. 2012. A three-level mixed-effects location scale model with an application to ecological momentary assessment (EMA) data. *Stat. Med.* 31(26):3192–210
Lin X, Mermelstein RJ, Hedeker D. 2018. A three-level Bayesian mixed effects location scale model with an application to ecological momentary assessment data. *Stat. Med.* 37(13):2108–19
Liu Y, West SG. 2015. Weekly cycles in daily report data: an overlooked issue. *J. Personal.* 84(5):560–79
Loeys T, Josephy H, Dewitte M. 2018. More precise estimation of lower-level interaction effects in multilevel models. *Multivar. Behav. Res.* 53(3):335–47
Lu T. 2018. Mixed-effects location and scale Tobit joint models for heterogeneous longitudinal data with skewness, detection limits, and measurement errors. *Stat. Methods Med. Res.* 27(12):3525–43

Lüdtke O, Marsh HW, Robitzsch A, Trautwein U. 2011. A 2 × 2 taxonomy of multilevel latent contextual models: accuracy-bias trade-offs in full and partial error correction models. *Psychol. Methods* 16(4):444–67
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. *Psychol. Methods* 13(3):203–29
Luo W, Kwok O. 2009. The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivar. Behav. Res.* 44(2):182–212
Ma Q, Dunton GF, Hedeker D. 2021. A mixed effect location-scale model with mixture distributed scale random effects to analyze (near) identical entries in ecological momentary assessments. *Multivar. Behav. Res.* 56(1):160
Matuschek H, Kliegl R, Vasishth S, Baayen H, Bates D. 2017. Balancing type I error and power in linear mixed models. *J. Memory Lang.* 94:305–15
McNeish D. 2017a. Multilevel mediation with small samples: a cautionary note on the multilevel structural equation modeling framework. *Struct. Equ. Model.* 24(4):609–25
McNeish D. 2017b. Small sample methods for multilevel modeling: a colloquial elucidation of REML and the Kenward-Roger correction. *Multivar. Behav. Res.* 52(5):661–70
McNeish D. 2021. Specifying location-scale models for heterogeneous variances as multilevel SEMs. *Organ. Res. Methods* 24(3):630–53
McNeish D, Hamaker EL. 2020. A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychol. Methods* 25(5):610–35
McNeish D, Kelley K. 2019. Fixed effects models versus mixed effects models for clustered data: reviewing the approaches, disentangling the differences, and making recommendations. *Psychol. Methods* 24(1):20–35
McNeish D, Matta T. 2018. Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behav. Res. Methods* 50:1398–414
Mehta PD, Neale MC. 2005. People are variables too: multilevel structural equations modeling. *Psychol. Methods* 10(3):259–84
Meuleman B, Billiet J. 2009. A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Surv. Res. Methods* 3(1):45–58
Meyers J, Beretvas S. 2006. The impact of inappropriate modeling of cross-classified data structures. *Multivar. Behav. Res.* 41:473–97
Moerbeek M. 2004. The consequence of ignoring a level of nesting in multilevel analysis. *Multivar. Behav. Res.* 39(1):129–49
O'Keefe P, Rodgers JL. 2017. Double decomposition of level-1 variables in multilevel models: an analysis of the Flynn effect in the NSLY data. *Multivar. Behav. Res.* 52(5):630–47
Park J, Cardwell R, Yu H-T. 2020. Specifying the random effect structure in linear mixed effect models for analyzing psycholinguistic data. *Methodology* 16(2):92–111
Preacher KJ, Zhang Z, Zyphur MJ. 2011. Alternative methods for assessing mediation in multilevel data: the advantages of multilevel SEM. *Struct. Equ. Model.* 18(2):161–82
Preacher KJ, Zhang Z, Zyphur MJ. 2016. Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychol. Methods* 21(2):189–205
Preacher KJ, Zyphur MJ, Zhang Z. 2010. A general multilevel SEM framework for assessing multilevel mediation. *Psychol. Methods* 15(3):209–33
Rast P, Ferrer E. 2018. A mixed-effects location scale model for dyadic interactions. *Multivar. Behav. Res.* 53(5):756–75
Rast P, Hofer SM, 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. *Multivar. Behav. Res.* 47(2):177–200
Raudenbush SW, Bryk AS. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage. 2nd ed.
Rights JD, Preacher KJ, Cole DA. 2020. The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects. *Br. J. Math. Stat. Psychol.* 73:194–211
Rights JD, Sterba SK. 2019. Quantifying explained variance in multilevel models: an integrative framework for defining R-squared measures. *Psychol. Methods* 24(3):309–38

Rights JD, Sterba SK. 2020. New recommendations on the use of R-squared differences in multilevel model comparisons. *Multivar. Behav. Res.* 55(4):568–99

Rights JD, Sterba SK. 2021. R-squared measures for multilevel models with three or more levels. *Multivariate Behav. Res.* In press

Schmidt-Catran AW, Fairbrother M. 2016. The random effects in multilevel models: getting them wrong and getting them right. *Eur. Sociol. Rev.* 32(1):23–38

Scott MA, Simonof JS, Marx BP, eds. 2013. *The SAGE Handbook of Multilevel Modeling*. London: Sage

Shi Y, Leite W, Algina J. 2010. The impact of omitting the interaction between crossed factors in cross-classified random effects modelling. *Br. J. Math. Stat. Psychol.* 63(Part 1):1–15

Singer J, Willett J. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford, UK: Oxford Univ. Press

Sliwinski MJ, Hoffman L, Hofer SM. 2010. Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. *Res. Hum. Dev.* 7(1):45–60

Smid SC, McNeish D, Miočević M, van de Schoot R. 2020. Bayesian versus frequentist estimation for structural equation models in small sample contexts: a systematic review. *Struct. Equ. Model.* 27(1):131–61

Snijders TAB, Bosker RJ. 2012. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage. 2nd ed.

Townsend Z, Buckley J, Harada M, Scott MA. 2013. The choice between fixed and random effects. See Scott et al. 2013, pp. 73–88

Van Landeghem G, De Fraine B, Van Damme J. 2005. The consequence of ignoring a level of nesting in multilevel analysis: a comment. *Multivar. Behav. Res.* 40(4):423–34

Vandenberg RJ, Richardson HA. 2019. A primer on multilevel structural modeling: user-friendly guidelines. See Humphrey & LeBreton 2019, pp. 449–72

Walters RW. 2015. *Mixed-effects location-scale models for conditionally normally distributed repeated-measures data.* PhD Diss., Univ. Nebraska-Lincoln

Walters RW, Hoffman L, Templin J. 2018. The power to detect and predict individual differences in intra-individual variability using the mixed-effects location-scale model. *Multivar. Behav. Res.* 53(3):360–74

Wang L, Hamaker E, Bergeman CS. 2012. Investigating inter-individual differences in short-term intra-individual variability. *Psychol. Methods* 17(4):567–81

Wang L, Maxwell SE. 2015. On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychol. Methods* 20(1):63–83

Williams DR, Zimprich DR, Rast P. 2019. A Bayesian nonlinear mixed-effects location scale model for learning. *Behav. Res. Methods* 51:1968–86

Yaremych HE, Preacher KJ, Hedeker D. 2021. Centering categorical predictors in multilevel models: best practices and interpretation. *Psychol. Methods*. In press

Ye F, Daniel L. 2017. The impact of inappropriate modeling of cross-classified data structures on random-slope models. *J. Mod. Appl. Stat. Methods* 16(2):458–84

Zigler CK, Ye F. 2019. A comparison of multilevel mediation modeling methods: recommendations for applied researchers. *Multivar. Behav. Res.* 54(3):338–59

Zitzmann S, Lüdtke O, Robitzsch A, Hecht M. 2020. On the performance of Bayesian approaches in small samples: a comment on Smid, McNeish, Miocevic, and van de Schoot. *Struct. Equ. Model.* 28(1):40–50

Annual Review of Psychology

Volume 73, 2022

# Contents

Recollecting What We Once Knew: My Life in Psycholinguistics
*Lila R. Gleitman and Claire Gleitman* ..... 1

Memory and Reward-Based Learning: A Value-Directed Remembering Perspective
*Barbara J. Knowlton and Alan D. Castel* ..... 25

Normative Principles for Decision-Making in Natural Environments
*Christopher Summerfield and Paula Parpart* ..... 53

Speech Computations of the Human Superior Temporal Gyrus
*Ilina Bhaya-Grossman and Edward F. Chang* ..... 79

Cognitive, Systems, and Computational Neurosciences of the Self in Motion
*Jean-Paul Noel and Dora E. Angelaki* ..... 103

Exploring Cognition with Brain–Machine Interfaces
*Richard A. Andersen, Tyson Aflalo, Luke Bashford, David Bjånes, and Spencer Kellis* ..... 131

Brain Mechanisms Underlying the Subjective Experience of Remembering
*Jon S. Simons, Maureen Ritchey, and Charles Fernyhough* ..... 159

Neurophysiology of Remembering
*György Buzsáki, Sam McKenzie, and Lila Davachi* ..... 187

The Basis of Navigation Across Species
*Cody A. Freas and Ken Cheng* ..... 217

Computational Psychiatry Needs Time and Context
*Peter F. Hitchcock, Eiko I. Fried, and Michael J. Frank* ..... 243

Persistence and Disengagement in Personal Goal Pursuit
*Veronika Brandstätter and Katharina Bernecker* ..... 271

Social Motivation at Work: The Organizational Psychology of Effort for, Against, and with Others
*Adam M. Grant and Marissa S. Shandell* ..... 301

Attitudes, Habits, and Behavior Change
*Bas Verplanken and Sheina Orbell* ..... 327

Childhood Antisocial Behavior: A Neurodevelopmental Problem
*Stephanie H.M. van Goozen, Kate Langley, and Christopher W. Hobson* ..... 353

Human Cooperation and the Crises of Climate Change, COVID-19, and Misinformation
*Paul A.M. Van Lange and David G. Rand* ..... 379

Diversity Training Goals, Limitations, and Promise: A Review of the Multidisciplinary Literature
*Patricia G. Devine and Tory L. Ash* ..... 403

Psychology and Indigenous People
*Roberto González, Héctor Carvacho, and Gloria Jiménez-Moya* ..... 431

Psychology Within and Without the State
*H. Clark Barrett* ..... 461

Personality Psychology
*Brent W. Roberts and Hee J. Yoon* ..... 489

Personal Values Across Cultures
*Lilach Sagiv and Shalom H. Schwartz* ..... 517

Educational Psychology Is Evolving to Accommodate Technology, Multiple Disciplines, and Twenty-First-Century Skills
*Arthur C. Graesser, John P. Sabatini, and Haiying Li* ..... 547

Cultivating Resilience During the COVID-19 Pandemic: A Socioecological Perspective
*Ning Zhang, Shujuan Yang, and Peng Jia* ..... 575

What Are the Health Consequences of Upward Mobility?
*Edith Chen, Gene H. Brody, and Gregory E. Miller* ..... 599

The Social Effects of Emotions
*Gerben A. van Kleef and Stéphane Côté* ..... 629

Catching Up on Multilevel Modeling
*Lesa Hoffman and Ryan W. Walters* ..... 659

Optimizing Research Output: How Can Psychological Research Methods Be Improved?
*Jeff Miller and Rolf Ulrich* ..... 691

Replicability, Robustness, and Reproducibility in Psychological Science
*Brian A. Nosek, Tom E. Hardwicke, Hannah Moshontz, Aurélien Allard, Katherine S. Corker, Anna Dreber, Fiona Fidler, Joe Hilgard, Melissa Kline Struhl, Michèle B. Nuijten, Julia M. Rohrer, Felipe Romero, Anne M. Scheel, Laura D. Scherer, Felix D. Schönbrodt, and Simine Vazire* ..... 719

Quantum Cognition
*Emmanuel M. Pothos and Jerome R. Busemeyer* ..... 749

## Indexes

Cumulative Index of Contributing Authors, Volumes 63–73 ..... 779

Cumulative Index of Article Titles, Volumes 63–73 ..... 784

## Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at http://www.annualreviews.org/errata/psych

# Related Articles

From the ***Annual Review of Anthropology***, Volume 50 (2021)

Doing Fieldwork Without Knowing It
*Renato Rosaldo*

Slippage: An Anthropology of Shamanism
*Bruce Grant*

Music, Language, Aurality: Latin American and Caribbean Resoundings
*Amanda Minks and Ana María Ochoa Gautier*

Syndemics: A Cross-Disciplinary Approach to Complex Epidemic Events Like COVID-19
*Merrill Singer, Nicola Bulled, Bayla Ostrach, and Shir Lerman Ginzburg*

Desiring Bureaucracy
*Tess Lea*

Food Insecurity, Nutritional Inequality, and Maternal–Child Health: A Role for Biocultural Scholarship in Filling Knowledge Gaps
*Barbara A. Piperata and Darna L. Dufour*

Transgressing Time: Archaeological Evidence in/of the Anthropocene
*Matt Edgeworth*

Political Theology/Theopolitics: The Thresholds and Vulnerabilities of Sovereignty
*Carlota McAllister and Valentina Napolitano*

The Earliest South African Hominids
*Ronald J. Clarke, Travis Rayne Pickering, Jason L. Heaton, and Kathleen Kuman*

Human Evolution in Asia: Taking Stock and Looking Forward
*Sang-Hee Lee and Autumn Hudock*

Rethinking the Landscape: Emerging Approaches to Archaeological Remote Sensing
*Jesse Casana*

Peirce and Archaeology: Recent Approaches
*Joanne P. Baron*

Touch and Social Interaction
*Asta Cekaite and Marjorie H. Goodwin*

Conversation and Culture
*Simeon Floyd*

Language and the Military: Necropolitical Legitimation, Embodied Semiotics, and Ineffable Suffering
*Janet McIntosh*

The Human Sleep Paradox: The Unexpected Sleeping Habits of *Homo sapiens*
*David R. Samson*

Intersectional Ecologies: Reimagining Anthropology and Environment
*Sarah E. Vaughn, Bridget Guarasci, and Amelia Moore*

Postcolonial Semiotics
*Angela Reyes*

The Political Economy of Attention
*Morten Axel Pedersen, Kristoffer Albris, and Nick Seaver*

Constructed Languages
*Christine Schreyer*

Feminism in the House of Anthropology
*Lilith Mahmud*

Pidgins and Creoles: Debates and Issues
*Christine Jourdan*

Archaeoprimatology: The *Longue Durée* Interface Between Humans and Nonhuman Primates
*Bernardo Urbani*

Recent Research on the Archaeology of War and Violence
*Andrew K. Scherer*

The Evolution of Human Infancy: Why It Helps to Be Helpless
*Karen R. Rosenberg*

From the ***Annual Review of Clinical Psychology***, Volume 17 (2021)

Smoking Treatment: A Report Card on Progress and Challenges
*Timothy B. Baker and Danielle E. McCarthy*

Network Analysis of Psychopathology: Controversies and Challenges
*Richard J. McNally*

Developing and Validating Clinical Questionnaires
*Anthony J. Rosellini and Timothy A. Brown*

The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence
*Roman Kotov, Robert F. Krueger, David Watson, David C. Cicero, Christopher C. Conway, Colin G. DeYoung, Nicholas R. Eaton, Miriam K. Forbes, Michael N. Hallquist, Robert D. Latzman, Stephanie N. Mullins-Sweatt, Camilo J. Ruggero, Leonard J. Simms, Irwin D. Waldman, Monika A. Waszczuk, and Aidan G.C. Wright*

History and Status of Prolonged Grief Disorder as a Psychiatric Diagnosis
*Sophia Kakarala, Holly G. Prigerson, James Gang, and Paul K. Maciejewski*

Violence, Place, and Strengthened Space: A Review of Immigration Stress, Violence Exposure, and Intervention for Immigrant Latinx Youth and Families
*Sarah A. Jolie, Ogechi Cynthia Onyeka, Stephanie Torres, Cara DiClemente, Maryse Richards, and Catherine DeCarlo Santiago*

Social Behavior as a Transdiagnostic Marker of Resilience
*Ruth Feldman*

Mental Health and Wealth: Depression, Gender, Poverty, and Parenting
*Megan V. Smith and Carolyn M. Mazure*

Ketamine and the Future of Rapid-Acting Antidepressants
*Lace M. Riggs and Todd D. Gould*

Intimate Relationships and Depression: Searching for Causation in the Sea of Association
*Mark A. Whisman, David A. Sbarra, and Steven R.H. Beach*

Saving Lives: Recognizing and Intervening with Youth at Risk for Suicide
*Alejandra Arango, Polly Y. Gipson, Jennifer G. Votta, and Cheryl A. King*

Early Environmental Upheaval and the Risk for Schizophrenia
*Vincent Paquin, Mylène Lapierre, Franz Veru, and Suzanne King*

DSM-5 Level of Personality Functioning: Refocusing Personality Disorder on What It Means to Be Human
*Carla Sharp and Kiana Wall*

Developmental Perspectives on the Study of Persons with Intellectual Disability
*Jacob A. Burack, David W. Evans, Natalie Russo, Jenilee-Sarah Napoleon, Karen J. Goldman, and Grace Iarocci*

Clinical and Translational Implications of an Emerging Developmental Substructure for Autism
*John N. Constantino, Tony Charman, and Emily J.H. Jones*

Conduct Disorders and Empathy Development
*Paul J. Frick and Emily C. Kemp*

Cognitive Behavioral Therapy for the Eating Disorders
*W. Stewart Agras and Cara Bohon*

Child Sexual Abuse as a Unique Risk Factor for the Development of Psychopathology: The Compounded Convergence of Mechanisms
*Jennie G. Noll*

Clinical Neuroscience of Addiction: What Clinical Psychologists Need to Know and Why
*Lara A. Ray and Erica N. Grodin*

Virtual Reality Therapy in Mental Health
*Paul M.G. Emmelkamp and Katharina Meyerbröker*

Resilience in Development and Psychopathology: Multisystem Perspectives
*Ann S. Masten, Cara M. Lucke, Kayla M. Nelson, and Isabella C. Stallworthy*

Designing Evidence-Based Preventive Interventions That Reach More People, Faster, and with More Impact in Global Contexts
*Mary Jane Rotheram-Borus*

Pathology in Relationships
*Susan C. South*

From the ***Annual Review of Developmental Psychology***, Volume 3 (2021)

A Conversation with Michael Rutter
*Michael Rutter and Janet F. Werker*

Clinical Staging for Youth Mental Disorders: Progress in Reforming Diagnosis and Clinical Care
*Patrick D. McGorry and Cristina Mei*

Neurodevelopmental Preparedness for Language in the Neonatal Brain
*Caroline Nallet and Judit Gervain*

Interactive Development of Adaptive Learning and Memory
*Catherine A. Hartley, Kate Nussenbaum, and Alexandra O. Cohen*

Achievement Motivation: What We Know and Where We Are Going
*Allan Wigfield, Katherine Muenks, and Jacquelynne S. Eccles*

Personality Assessment of Children and Adolescents
*Rebecca L. Shiner, Christopher J. Soto, and Filip De Fruyt*

Executive Functions in Social Context: Implications for Conceptualizing, Measuring, and Supporting Developmental Trajectories
*Yuko Munakata and Laura E. Michaelson*

Young Children's Interactions with Objects: Play as Practice and Practice as Play
*Jeffrey J. Lockman and Catherine S. Tamis-LeMonda*

Contributions of the Fragile Families and Child Wellbeing Study to Child Development
*Sarah James, Sara McLanahan, and Jeanne Brooks-Gunn*

Early Childhood Obesity: A Developmental Perspective
*Megan H. Pesch and Julie C. Lumeng*

Asthma as a Developmental Disorder
*Fernando D. Martinez*

Developmental Cognitive Neuroscience in the Era of Networks and Big Data: Strengths, Weaknesses, Opportunities, and Threats
*Damien A. Fair, Nico U.F. Dosenbach, Amy H. Moore, Theodore Satterthwaite, and Michael P. Milham*

From the ***Annual Review of Law and Social Science***, Volume 17 (2021)

Law and/or/as Civility
*Keith J. Bybee*

Social Theory and Legal Theory: Contemporary Interactions
*Roger Cotterrell*

Hobbling: The Effects of Proactive Policing and Mass Imprisonment on Children's Education
*Benjamin Justice*

Governance by Data
*Fleur Johns*

Truth Commission Impact on Policy, Courts, and Society
*Onur Bakiner*

Legal Responsibility Among the Young and the Elderly
*Eve M. Brank and Lindsey E. Wylie*

Gender Inequalities in Markets
*Tamar Kricheli-Katz*

The Reasonable Person Standard: Psychological and Legal Perspectives
*Mark D. Alicke and Stephanie H. Weigel*

Business and Human Rights: Alternative Approaches to Transnational Regulation
*Surya Deva*

Protecting Basic Legal Freedoms: International Legal Complexes, Accountability Devices, and the Deviant Case of China
*Terence C. Halliday, Shira Zilberstein, and Wendy Espeland*

The Impact of Experienced and Expressed Emotion on Legal Factfinding
*Jessica M. Salerno*

Street-Level Meta-Strategies: Evidence on Restorative Justice and Responsive Regulation
*John Braithwaite*

Transitional Justice and Property: Inextricably Linked
*Helena Alviar García*

Replicability in Empirical Legal Research
*Jason M. Chin and Kathryn Zeiler*

Water Security and International Law
*Philippe Cullet, Lovleen Bhullar, and Sujith Koonan*

What Is Cultural Cognition, and Why Does It Matter?
*Jeffrey J. Rachlinski*

Contract Schemas
*Roseanna Sommers*

Algorithms and Decision-Making in the Public Sector
*Karen Levy, Kyla E. Chasalow, and Sarah Riley*

Parole Board Decision Making and Constitutional Rights
*Amelia Courtney Hritz*

Infrastructures and Laws: Publics and Publicness
*Benedict Kingsbury and Nahuel Maisley*

Advancing Socioeconomic Rights Through Interdisciplinary Factfinding: Opportunities and Challenges
*Sarah Knuckey, Joshua D. Fisher, Amanda M. Klasing, Tess Russo, and Margaret L. Satterthwaite*

Philanthrocapitalism and the Separation of Powers
*Linsey McGoey*

Constitutional Dictatorships, from Colonialism to COVID-19
*Jens Meierhenrich*

Black Lives Matter in Historical Perspective
*Megan Ming Francis and Leah Wright-Rigueur*

Women's Rights After War: On Gender Interventions and Enduring Hierarchies
*Marie E. Berry and Milli Lake*

On the Interdependence of Liberal and Illiberal/Authoritarian Legal Forms in Racial Capitalist Regimes…The Case of the United States
*Michael McCann and Filiz Kahraman*

From the ***Annual Review of Neuroscience***, Volume 44 (2021)

Sources of Individual Differences in Pain
*Jeffrey S. Mogil*

Current Status of and Perspectives on the Application of Marmosets in Neurobiology
*Hideyuki Okano*

Astrocytes and Behavior
*Paulo Kofuji and Alfonso Araque*

A Common Space Approach to Comparative Neuroscience
*Rogier B. Mars, Saad Jbabdi, and Matthew F.S. Rushworth*

Parkinson's Disease Genetics and Pathophysiology
*Gabriel E. Vázquez-Vélez and Huda Y. Zoghbi*

Neural and Molecular Mechanisms of Biological Embedding of Social Interactions
*Ian M. Traniello and Gene E. Robinson*

Smartphones and the Neuroscience of Mental Health
*Claire M. Gillan and Robb B. Rutledge*

Integrated Patterning Programs During *Drosophila* Development Generate the Diversity of Neurons and Control Their Mature Properties
*Anthony M. Rossi, Shadi Jafari, and Claude Desplan*

Consolidating the Circuit Model for Addiction
*Christian Lüscher and Patricia H. Janak*

Ensheathment and Myelination of Axons: Evolution of Glial Functions
*Klaus-Armin Nave and Hauke B. Werner*

Neocortical Layer 1: An Elegant Solution to Top-Down and Bottom-Up Integration
*Benjamin Schuman, Shlomo Dellal, Alvar Prönneke, Robert Machold, and Bernardo Rudy*

Human Representation Learning
*Angela Radulescu, Yeon Soon Shin, and Yael Niv*

Dense Circuit Reconstruction to Understand Neuronal Computation: Focus on Zebrafish
*Rainer W. Friedrich and Adrian A. Wanner*

The Role of the Medial Prefrontal Cortex in Moderating Neural Representations of Self and Other in Primates
*Masaki Isoda*

Inferring Macroscale Brain Dynamics via Fusion of Simultaneous EEG-fMRI
*Marios G. Philiastides, Tao Tu, and Paul Sajda*

Ion Channel Degeneracy, Variability, and Covariation in Neuron and Circuit Resilience
*Jean-Marc Goaillard and Eve Marder*

Oxytocin, Neural Plasticity, and Social Behavior
*Robert C. Froemke and Larry J. Young*

Physiology and Pathophysiology of Mechanically Activated PIEZO Channels
*Ruhma Syeda*

The Geometry of Information Coding in Correlated Neural Populations
*Rava Azeredo da Silveira and Fred Rieke*

The Cortical Motor Areas and the Emergence of Motor Skills: A Neuroanatomical Perspective
*Peter L. Strick, Richard P. Dum, and Jean-Alban Rathelot*

Perceptual Inference, Learning, and Attention in a Multisensory World
*Uta Noppeney*

Adaptive Prediction for Social Contexts: The Cerebellar Contribution to Typical and Atypical Social Behaviors
*Catherine J. Stoodley and Peter T. Tsai*

Neurophysiology of Human Perceptual Decision-Making
*Redmond G. O'Connell and Simon P. Kelly*

How Cortical Circuits Implement Cortical Computations: Mouse Visual Cortex as a Model
*Cristopher M. Niell and Massimo Scanziani*

Spatial Transcriptomics: Molecular Maps of the Mammalian Brain
*Cantin Ortiz, Marie Carlén, and Konstantinos Meletis*

From the ***Annual Review of Organizational Psychology and Organizational Behavior***, Volume 8 (2021)

Reflections on a Career Studying Individual Differences in the Workplace
*Paul R. Sackett*

Workplace Envy
*Michelle K. Duffy, KiYoung Lee, and Elizabeth A. Adair*

The Science of Workplace Instruction: Learning and Development Applied to Work
*Kurt Kraiger and J. Kevin Ford*

Balancing the Scales: A Configurational Approach to Work-Life Balance
*Nancy P. Rothbard, Arianna M. Beetz, and Dana Harari*

The Lazy or Dishonest Respondent: Detection and Prevention
*Winfred Arthur Jr., Ellen Hagen, and Felix George Jr.*

Emotion Work: A Work Psychology Perspective
*Dieter Zapf, Marcel Kern, Franziska Tschan, David Holman, and Norbert K. Semmer*

Chief Executive Officer Succession and Board Decision Making: Review and Suggestions for Advancing Industrial and Organizational Psychology, Human Resources Management, and Organizational Behavior Research
*Anthony J. Nyberg, Ormonde R. Cragun, and Donald J. Schepker*

Trait Activation Theory: A Review of the Literature and Applications to Five Lines of Personality Dynamics Research
*Robert P. Tett, Margaret J. Toich, and S. Burak Ozkum*

Theory and Technology in Organizational Psychology: A Review of Technology Integration Paradigms and Their Effects on the Validity of Theory
*Richard N. Landers and Sebastian Marin*

Overqualification at Work: A Review and Synthesis of the Literature
*Berrin Erdogan and Talya N. Bauer*

Putting People Down and Pushing Them Out: Sexual Harassment in the Workplace
*Lilia M. Cortina and Maira A. Areguin*

The Science and Practice of Item Response Theory in Organizations
*Jonas W.B. Lang and Louis Tay*

From the ***Annual Review of Political Science***, Volume 24 (2021)

Measuring Liberalism, Confronting Evil: A Retrospective
*Ira Katznelson*

Presidential Unilateral Power
*Kenneth Lowande and Jon C. Rogowski*

Violence Against Civilians During Armed Conflict: Moving Beyond the Macro- and Micro-Level Divide
*Laia Balcells and Jessica A. Stanton*

The Causes of Populism in the West
*Sheri Berman*

Networks of Conflict and Cooperation
*Jennifer M. Larson*

Nationalism: What We Know and What We Still Need to Know
*Harris Mylonas and Maya Tudor*

Party and Ideology in American Local Government: An Appraisal
*Sarah F. Anzia*

Social Protection and State–Society Relations in Environments of Low and Uneven State Capacity
*Arthur Alik-Lagrange, Sarah K. Dreier, Milli Lake, and Alesha Porisky*

The Continuing Dilemma of Race and Class in the Study of American Political Behavior
*Fredrick C. Harris and Viviana Rivera-Burgos*

Conflict-Related Sexual Violence
*Ragnhild Nordås and Dara Kay Cohen*

Secrecy in International Relations and Foreign Policy
*Allison Carnegie*

How Do Electoral Gender Quotas Affect Policy?
*Amanda Clayton*

Who Enters Politics and Why?
*Saad Gulzar*

Ethics of Field Experiments
*Trisha Phillips*

The Persistence of Racial Cues and Appeals in American Elections
*LaFleur Stephens-Dougan*

What Can We Learn from Written Constitutions?
*Zachary Elkins and Tom Ginsburg*

The Rise of Local Politics: A Global Review
*Patrick Le Galès*

External Validity
*Michael G. Findley, Kyosuke Kikuta, and Michael Denly*

Machine Learning for Social Science: An Agnostic Approach
*Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart*

The Backlash Against Globalization
*Stefanie Walter*

The Politics of the Black Power Movement
*James Lance Taylor*

Populism, Democracy, and Party System Change in Europe
*Milada Anna Vachudova*

From the ***Annual Review of Public Health***, Volume 42 (2021)

Climate Change Disinformation and How to Combat It
*Stephan Lewandowsky*

Considerations for Developing an Agenda for Gun Violence Prevention Research
*Mark Rosenberg*

Influenza Virus: Tracking, Predicting, and Forecasting
*Sheikh Taslim Ali and Benjamin J. Cowling*

Prediabetes and What It Means: The Epidemiological Evidence
*Justin B. Echouffo-Tcheugui and Elizabeth Selvin*

Reproducible Research: A Retrospective
*Roger D. Peng and Stephanie C. Hicks*

A Critical Review of the Social and Behavioral Contributions to the Overdose Epidemic
*Magdalena Cerdá, Noa Krawczyk, Leah Hamilton, Kara E. Rudolph, Samuel R. Friedman, and Katherine M. Keyes*

Early Childhood Adversity, Toxic Stress, and the Impacts of Racism on the Foundations of Health
*Jack P. Shonkoff, Natalie Slopen, and David R. Williams*

Expanding Implementation Research to Prevent Chronic Diseases in Community Settings
*Stephanie Mazzucca, Elva M. Arredondo, Deanna M. Hoelscher, Debra Haire-Joshu, Rachel G. Tabak, Shiriki K. Kumanyika, and Ross C. Brownson*

Understanding and Responding to Health Literacy as a Social Determinant of Health
*Don Nutbeam and Jane E. Lloyd*

Vaccine Hesitancy, Acceptance, and Anti-Vaccination: Trends and Future Prospects for Public Health
*Ève Dubé, Jeremy K. Ward, Pierre Verger, and Noni E. MacDonald*

Air Quality in Africa: Public Health Implications
*Asmamaw Abera, Johan Friberg, Christina Isaxon, Michael Jerrett, Ebba Malmqvist, Cheryl Sjöström, Tahir Taj, and Ana Maria Vargas*

At the Water's Edge: Coastal Settlement, Transformative Adaptation, and Well-Being in an Era of Dynamic Climate Risk
*William Solecki and Erin Friedman*

Climate Change, Food Supply, and Dietary Guidelines
*Colin W. Binns, Mi Kyung Lee, Bruce Maycock, Liv Elin Torheim, Keiko Nanishi, and Doan Thi Thuy Duong*

Environmental Health Threats to Latino Migrant Farmworkers
*Federico Castillo, Ana M. Mora, Georgia L. Kayser, Jennifer Vanos, Carly Hyland, Audrey R. Yang, and Brenda Eskenazi*

Environmental Influences on the Human Microbiome and Implications for Noncommunicable Disease
*Jiyoung Ahn and Richard B. Hayes*

Extreme Weather and Climate Change: Population Health and Health System Implications
*Kristie L. Ebi, Jennifer Vanos, Jane W. Baldwin, Jesse E. Bell, David M. Hondula, Nicole A. Errett, Katie Hayes, Colleen E. Reid, Shubhayu Saha, June Spector, and Peter Berry*

Green Infrastructure and Health
*Mark J. Nieuwenhuijsen*

Addressing Social Needs in Health Care Settings: Evidence, Challenges, and Opportunities for Public Health
*Matthew W. Kreuter, Tess Thompson, Amy McQueen, and Rachel Garg*

Benchmarking as a Public Health Strategy for Creating Healthy Food Environments: An Evaluation of the INFORMAS Initiative (2012–2020)
*Gary Sacks, Janelle Kwon, Stefanie Vandevijvere, and Boyd Swinburn*

Cash Transfers and Health
*Sicong Sun, Jin Huang, Darrell L. Hudson, and Michael Sherraden*

Declining Life Expectancy in the United States: Missing the Trees for the Forest
*Sam Harper, Corinne A. Riddell, and Nicholas B. King*

Enhancing Community Engagement by Schools and Programs of Public Health in the United States
*Mindi B. Levin, Janice V. Bowie, Steven K. Ragsdale, Amy L. Gawad, Lisa A. Cooper, and Joshua M. Sharfstein*

Progress in National Policies Supporting the Sustainable Development Goals: Policies that Matter to Income and Its Impact on Health
*Amy Raub and Jody Heymann*

Sugar-Sweetened Beverage Reduction Policies: Progress and Promise
*James Krieger, Sara N. Bleich, Stephanie Scarmo, and Shu Wen Ng*

Improving Access to Care: Telemedicine Across Medical Domains
*William Barbosa, Kina Zhou, Emma Waddell, Taylor Myers, and E. Ray Dorsey*

The Effects of Home Care Provider Mix on the Care Recipient: An International, Systematic Review of Articles from 2000 to 2020
*Norma B. Coe, R. Tamara Konetzka, Melissa Berkowitz, Emily Blecker, and Courtney H. Van Houtven*

Trends in Abortion Policies in Low- and Middle-Income Countries
*Pascale Allotey, T.K. Sundari Ravindran, and Vithiya Sathivelu*

From the ***Annual Review of Sociology***, Volume 47 (2021)

La Sociología de las Emociones en América Latina
*Marina Ariza*

From Physics to Russian Studies and on into China Research: My Meandering Journey Toward Sociology
*Martin King Whyte*

Living Sociology: On Being in the World One Studies
*Michael Burawoy*

Ethnography, Data Transparency, and the Information Age
*Alexandra K. Murphy, Colin Jerolmack, and DeAnna Smith*

Rethinking Culture and Cognition
*Karen A. Cerulo, Vanina Leschziner, and Hana Shepherd*

The Influence of Simmel on American Sociology Since 1975
*Miloš Broćić and Daniel Silver*

Whatever Happened to Socialization?
*Jeffrey Guhin, Jessica McCrory Calarco, and Cynthia Miller-Idriss*

A Retrospective on Fundamental Cause Theory: State of the Literature and Goals for the Future
*Sean A.P. Clouston and Bruce G. Link*

The Sociology of Emotions in Latin America
*Marina Ariza*

Negative Social Ties: Prevalence and Consequences
*Shira Offer*

The (Un)Managed Heart: Racial Contours of Emotion Work in Gendered Occupations
*Adia Harvey Wingfield*

The Society of Algorithms
*Jenna Burrell and Marion Fourcade*

Trust in Social Relations
*Oliver Schilke, Martin Reimann, and Karen S. Cook*

New Directions in the Study of Institutional Logics: From Tools to Phenomena
*Michael Lounsbury, Christopher W.J. Steele, Milo Shaoqing Wang, and Madeline Toubiana*

The Civil Rights Revolution at Work: What Went Wrong
*Frank Dobbin and Alexandra Kalev*

University Governance in Meso and Macro Perspectives
*Christine Musselin*

Populism Studies: The Case for Theoretical and Comparative Reconstruction
*Cihan Tuğal*

Recent Trends in Global Economic Inequality
*Ho-fung Hung*

The Sharing Economy: Rhetoric and Reality
*Juliet B. Schor and Steven P. Vallas*

Comparative Perspectives on Racial Discrimination in Hiring: The Rise of Field Experiments
*Lincoln Quillian and Arnfinn H. Midtbøen*

Gender, Power, and Harassment: Sociology in the #MeToo Era
*Abigail C. Saguy and Mallory E. Rees*

Black Men and Black Masculinity
*Alford A. Young Jr.*

The "Burden" of Oppositional Culture Among Black Youth in America
*Karolyn Tyson and Amanda E. Lewis*

New Destinations and the Changing Geography of Immigrant Incorporation
*Chenoa A. Flippen and Dylan Farrell-Bryan*

Social Inequality and the Future of US Life Expectancy
*Iliya Gutin and Robert A. Hummer*

Markets Everywhere: The Washington Consensus and the Sociology of Global Institutional Change
*Sarah Babb and Alexander Kentikelenis*

Women's Health in the Era of Mass Incarceration
*Christopher Wildeman and Hedwig Lee*

Social Issues in Contemporary Russia: Women's Rights, Corruption, and Immigration Through Three Sociological Lenses
*Marina Zaloznaya and Theodore P. Gerber*

The Social and Sociological Consequences of China's One-Child Policy
*Yong Cai and Wang Feng*

From the ***Annual Review of Statistics and Its Application***, Volume 8 (2021)

Modeling Player and Team Performance in Basketball
*Zachary Terner and Alexander Franks*

Graduate Education in Statistics and Data Science: The Why, When, Where, Who, and What
*Marc Aerts, Geert Molenberghs, and Olivier Thas*

Statistical Evaluation of Medical Tests
*Vanda Inácio, María Xosé Rodríguez-Álvarez, and Pilar Gayoso-Diz*

Simulation and Analysis Methods for Stochastic Compartmental Epidemic Models
*Tapiwa Ganyani, Christel Faes, and Niel Hens*

Missing Data Assumptions
*Roderick J. Little*

Consequences of Asking Sensitive Questions in Surveys
*Ting Yan*

Synthetic Data
*Trivellore E. Raghunathan*

Algorithmic Fairness: Choices, Assumptions, and Definitions
*Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum*

Online Learning Algorithms
*Nicolò Cesa-Bianchi and Francesco Orabona*

Space-Time Covariance Structures and Models
*Wanfang Chen, Marc G. Genton, and Ying Sun*

Extreme Value Analysis for Financial Risk Management
*Natalia Nolde and Chen Zhou*

Sparse Structures for Multivariate Extremes
*Sebastian Engelke and Jevgenijs Ivanovs*

Compositional Data Analysis
*Michael Greenacre*

Distance-Based Statistical Inference
*Marianthi Markatou, Dimitris Karlis, and Yuxin Ding*

A Review of Empirical Likelihood
*Nicole A. Lazar*

Tensors in Statistics
*Xuan Bi, Xiwei Tang, Yubai Yuan, Yanqing Zhang, and Annie Qu*

Flexible Models for Complex Data with Applications
*Christophe Ley, Slađana Babić, and Domien Craens*

Adaptive Enrichment Designs in Clinical Trials
*Peter F. Thall*

Quantile Regression for Survival Data
*Limin Peng*

Statistical Applications in Educational Measurement
*Hua-Hua Chang, Chun Wang, and Susu Zhang*

Statistical Connectomics
*Jaewon Chung, Eric Bridgeford, Jesús Arroyo, Benjamin D. Pedigo, Ali Saad-Eldin, Vivek Gopalakrishnan, Liang Xiang, Carey E. Priebe, and Joshua T. Vogelstein*

Twenty-First-Century Statistical and Computational Challenges in Astrophysics
*Eric D. Feigelson, Rafael S. de Souza, Emille E.O. Ishida, and Gutti Jogesh Babu*