9-1-2021

Multilevel Modeling of Interval-Contingent Data In Neuropsychology Research Using the *lmerTest* Package In R

Richard S. Pond Jr.
University of North Carolina Wilmington, pondr@uncw.edu

Matison W. McCool
University of North Carolina Wilmington

Brian A. Bulla
University of Southern Mississippi, brian.bulla@usm.edu

Multilevel Modeling of Interval-Contingent Data in Neuropsychology Research Using the lmerTest Package in R

Richard S. Pond, Jr.[1]

Matison W. McCool[1]

Brian A. Bulla[2]

[1]University of North Carolina Wilmington, Department of Psychology, Wilmington, NC

[2]The University of Southern Mississippi, Department of Psychology, Hattiesburg, MS

in press at the *Journal of Pediatric Neuropsychology*'s section focusing on methodological and statistical advances (please see published version here: https://rdcu.be/cdNL8)

Word count: 7937

Please address correspondence to:

Richard S. Pond, Jr., Ph.D.
Department of Psychology
University of North Carolina Wilmington
Wilmington, NC 28403
Phone: 910.962.3372
Email: pondr@uncw.edu
ORCID: 0000-0001-5446-7456

ABSTRACT

Intensive longitudinal research designs are becoming more common in the field of neuropsychology. They are a powerful approach to studying development and change in naturally occurring phenomena. However, to fully capitalize on the wealth of data yielded by these designs, researchers have to understand the nature of multilevel data structures. The purpose of the present article is to describe some of the basic concepts and techniques involved in modeling multilevel data structures. In addition, this article serves as a step-by-step tutorial to demonstrate how neuropsychologists can implement basic multilevel modeling techniques with real data and the R package, *lmerTest.* R may be an ideal option for some empirical scientists, applied statisticians, and clinicians, because it is a free and open-source programming language for statistical computing and graphics that offers a flexible and powerful set of tools for analyzing data. All data and code described in the present article have been made publicly available.

Keywords: Statistical Methods, Multilevel Modeling, Hierarchical Linear Modeling, Linear Mixed-Effects Modeling, R

Multilevel Modeling of Interval-Contingent Data in Neuropsychology Research Using the lmerTest Package in R

Over the past three decades, researchers across the social and behavioral sciences have increasingly become interested in using intensive longitudinal designs to study naturally occurring phenomena. This is particularly the case for pediatric neuropsychologists, who may be interested in the developmental trajectories of neurological disorders and behavior problems (e.g., Infante, Nguyen-Louie, Worley, Courtney, Coronado, & Jacobus, 2020; Sullivan, Brumback, Tapert, Brown, Baker, Colrain, Prouty, De Bellis, Clark, Nagel, Pohl, & Pfefferbaum, 2019) or the long-term consequences of traumatic brain injury (e.g., Séguin, Dégeilh, Bernier, El-Jalbout, & Beauchamp, 2020). For instance, a recent Google Scholar search of the terms "developmental trajectory" and "neuropsychology" yielded over 6,000 hits just for the year 2020. Neuropsychologists can employ such designs to examine microlevel processes and investigate how spontaneous experiences in their natural environment interact with stable characteristics to influence behavior, without the limitations imposed on by retrospective reporting (Bolger, Davis, & Rafaeli, 2003; Wheeler & Reis, 1991). However, the complexity of the data yielded in such designs often present the researcher with analytic challenges. The goal of the present article is to introduce multilevel modeling (MLM; sometimes referred to as hierarchical linear modeling or linear mixed-models), which is thought to be one of the best methods for analyzing data generated from intensive repeated assessment (Nezlek, 2011a). A second goal is to illustrate how to perform MLM analyses in R (i.e., a free and open-source programming language for statistical computing and graphics; see Weston & Yee, 2017) with the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) and real data. This article is intended for researchers with primary experience in ordinary least squares (OLS)-based techniques, such as regression and analysis of variance (ANOVA). For those readers who are interested in a more thorough explication of MLM, there are a number of excellent, authoritative introductions to the technique, including: Raudenbush and Bryk (2002), Kreft and de Leeuw (1998), and Snijders and Bosker (2012). A particularly user-friendly resource for the non-expert is Nezlek's (2011b) brief introduction to MLM.

Conceptual Background

Multilevel, or hierarchical, data structures are clustered data structures (sometimes referred to as *nested* data), such that observations at one level of analysis (e.g., ratings of daily stressors each night for a 30-day period) are clustered (or *nested*) within observations at another level of analysis (e.g., individual clients). Nesting can occur within participants, as in the case of studies that employ an intensive longitudinal or repeated-measures design. However, nesting can also occur between groups. Indeed, the fields of education (where students are nested within classrooms or schools) and organizational psychology and economics (where workers are nested within companies or industries) each have long, rich histories of studying and modeling multilevel data structures (e.g., Bryk & Raudenbush, 1992; Hoffman, 1997; Kreft & de Leeuw, 1994; Raudenbush & Bryk, 1986, 1988). This clustering or nesting creates dependency in the data, such that subsets of cases or responses are more similar to each other and, thus, more highly correlated with each other than with cases or responses in other subsets. The current article focuses on data structures typical in studying repeated assessments of naturally occurring phenomena; however, the principles and techniques discussed below are also relevant to grouped or organizational data structures. Figure 1 illustrates a three-level hierarchical data structure potentially relevant to intensive longitudinal research designs in pediatric neuropsychology.

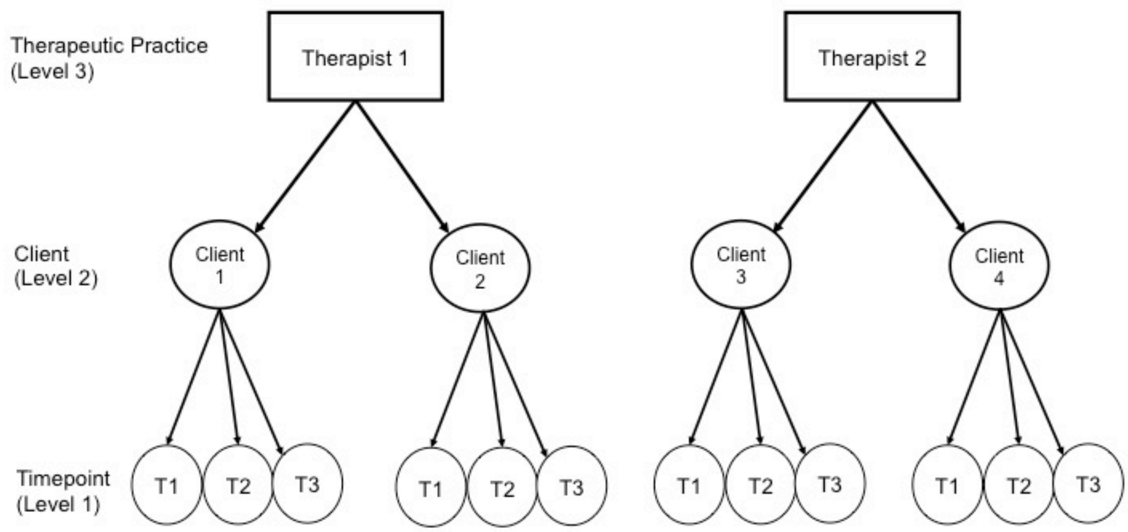

Figure 1. Example illustration of a three-level data structure, such that responses or assessments over time are nested within clients, and clients are nested within therapeutic practices.

Data from intensive repeated assessments can be further described as either *interval-contingent* or *event-contingent* (Nezlek, 2001, 2011a; Wheeler & Reis, 1991). Interval-contingent data occurs when participants provide data at a certain interval. For example, in diary-style research, participants typically report on their daily mood and relationships once per day (often at night), and daily responses are treated as being nested within participants (e.g., Gilbert, Pond, Haak, DeWall, & Keller, 2015; Pond, Kashdan, DeWall, Savostyanova, Lambert, & Fincham, 2012). Event-contingent data occur when a specific type of event triggers data collection efforts. For instance, in experience-sampling research, it's common to have participants make reports each time they engage in a social interaction (e.g., DeWall, Lambert, Pond, Kashdan, & Fincham, 2012; Nezlek, 1995; Nezlek, Hampton, & Shean, 2000), and interactions are treated as being nested within participants. Although the present article focuses on interval-contingent data, the principles and techniques discussed below also apply to event-contingent data.

Limitations of OLS

The dependency of nested data inherently violates the assumption of independence in OLS regression. The consequence of this dependency is that the standard errors produced by OLS are too small (Cohen, Cohen, Aiken, & West, 2003). This means that any confidence intervals or significance tests generated with those standard errors will be biased towards over-estimating significance (i.e., Type I error inflation). The more similar scores are within clusters, the more serious the problem. Furthermore, dependency is a particular concern for data that are nested as a result of repeated-assessments over time, which may be more relevant to the research designs of pediatric neuropsychologists (e.g., Infante et al., 2020), as opposed to nesting due to group membership (e.g., students nested within classrooms or people nested within families). The dependency that occurs in longitudinal designs deserves special attention here because of a phenomenon known as autocorrelation. Autocorrelation is present when adjacent observations are more strongly related to each other than nonadjacent observations (West & Hepworth, 1991). In longitudinal designs, observations that are closer in time tend to be more strongly related to each other than observations spaced further apart (e.g., the stress ratings of participants on day 1 will tend to be strongly correlated with their stress ratings on day 2 than on day 25 in a month-long diary study). Autocorrelation reduces estimates of within-person variability, biases standard errors, and inflates the Type I error rate of significance tests (Bolger et al., 2003).

OLS-based analytic strategies for handling multilevel data structures generally fall into two categories (Cohen et al., 2003; Nezlek, 2001). The first is to completely ignore the nesting and analyze responses as if there were no hierarchical structure (*disaggregated analysis*). This method is most at risk of Type I error inflation; however, dummy codes for cluster membership can be entered as predictors to partial out the average influence of cluster. This modified approach is limited in that: 1. it treats the cluster effect as *fixed*, meaning that one can only generalize to the specific clusters (e.g., specific persons or groups) that were represented in the sample and not a larger population from which one's clusters may have been randomly sampled (Cohen et al., 2003), and 2. it ignores between-cluster variation (i.e., the associations among modeled variables may be different for some clusters compared to others) (Nezlek, 2001).

The second OLS-based approach to handling multilevel data is to take the cluster average for each predictor and outcome and treat cluster as the unit of analysis in one's regression (*aggregated analysis*). Aggregated analysis leads to a loss of information, and, thus, power, and the reliability of the observations that go into each cluster mean is not taken into account (Nezlek, 2001). Perhaps most concerning is that the results of aggregated analyses are limited to the cluster level; results cannot be generalized to level 1 units (Cohen et al., 2003). This can promote misleading conclusions. For instance, does it make sense to suggest someone suffers from chronic stress based on this person's average level of stress reported across a 7-day period? Perhaps, this person just had a stressful few days.

Advantages of Multilevel Modeling

Multilevel modeling is an appropriate technique for modeling clustered data, because it was designed specifically to account for the cluster-level dependency that violates the assumption of independence in OLS regression. That is, MLM accounts for cluster-level dependency because it models multiple error terms across each level of analysis. It accomplishes this by estimating a regression equation for each unit of analysis at the lowest level, and then those coefficients (i.e., intercepts and slopes) are used as dependent variables in regressions at the next level of analysis (Nezlek, 2011b). Furthermore, despite separate equations for each level of analysis, all coefficients are estimated with maximum likelihood procedures simultaneously, and lower-level coefficients are estimated with two parameters: a *fixed* component and a *random* error term. In the context of a repeated-measures study, the fixed component would represent the average value for a parameter of interest (i.e., intercept or slope) across all subjects, whereas the random error would represent how much those parameters tend to vary across subjects. Thus, when one models repeated-measures with MLM, the model yields the average intercept and slope(s) across subjects (fixed components); however, the presence of significant random variance indicates that the intercepts and slope(s) significantly vary across each subject (as illustrated in Figure 2).

The advantage of modeling random error this way is that we don't need to assume that associations between our predictor variables and outcome are the same across each subject, which we would need to whenever using OLS (i.e., the homoscedasticity assumption in OLS regression). Instead, we can explicitly model this variability by examining between-person differences. Furthermore, modeling individual differences in intercepts and slopes can eliminate the problem of autocorrelation for repeated-measures (Bolger et al., 2003; Singer, 1998).

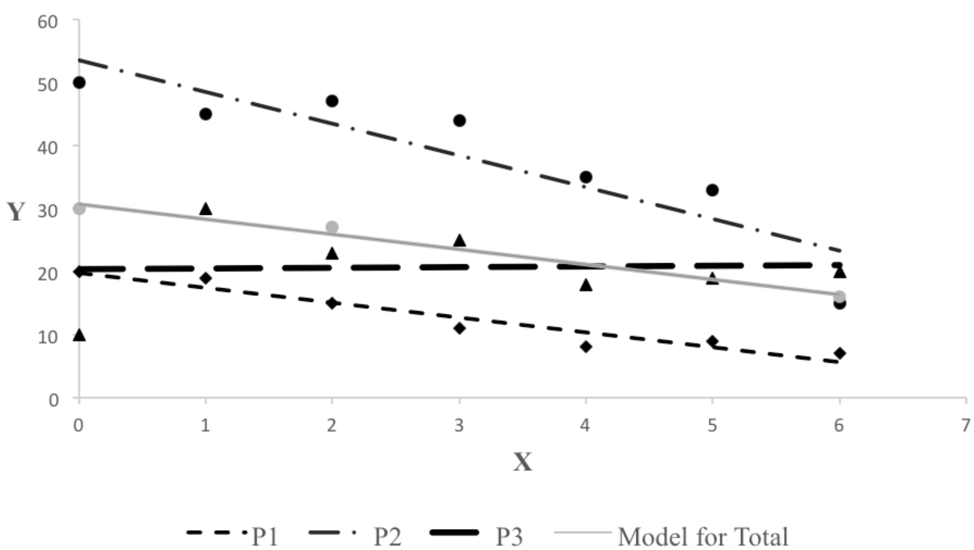

Figure 2. Example illustration of a model with a random intercept (different starting points for participants) and a random slope (association between X and Y varies across participants). The overall model fitted line is also plotted (fixed components for intercept and slope).

MLM not only accounts for the nested structure of data, but it also uses algorithms that apply Bayes shrinkage to weight observations by their reliabilities. This method moves outliers, or less reliable observations, closer to the mean (Nezlek, 2011b). By using Bayes shrinkage, more accurate estimates, relative to population measures, are achieved when compared to procedures that do not employ Bayes shrinkage (Littell et al., 1996; Raudenbush & Bryk, 2002). A final advantage to using MLM is that multilevel models do not require complete data sets. OLS-based procedures (e.g., regression, ANOVA) perform best with balanced designs and the absence of missing data. Oftentimes, in repeated-measures designs, cases with missing data at one time-point need to be deleted. For MLM, when data are missing on an outcome at a certain time-point, whole cases do not need to be deleted, nor do missing values on *y* need to be imputed. Instead, parameters are estimated with all available data using maximum likelihood procedures.

An Illustrative Example: Cerebral Asymmetry and Aggressive Tendencies

In the remainder of this article, we will demonstrate multilevel modeling with real interval-contingent data collected from a sample of young adults (M_{age}=18.70) across a 21-day period. We will use these data to test the hypothesis that within-person differences in cerebral asymmetry are associated with daily aggressive tendencies. The research literature shows that approach motivation, as well as traits and emotions linked to approach motivation, are positively associated with aggressive behavior (e.g., Carver, 2004; Carver & Harmon-Jones, 2009; Harmon-Jones, 2003; Harmon-Jones, Harmon-Jones, Abramson, & Peterson, 2009; Harmon-Jones & Peterson, 2008, 2009; Wilkowski & Meier, 2010). On the other hand, traits and emotions related to behavioral avoidance are typically inversely associated with aggression (e.g., Harmon-Jones, 2003; Pond, DeWall, Lambert, Deckman, Bonser, & Fincham, 2012; Smits & Kuppens, 2005).

Asymmetrical activation in the frontal cortices is closely linked to both approach and avoidance motivation. Indeed, greater left frontal activity tends to be associated with approach-related emotions (e.g., anger, joy; Drake & Myers, 2006), whereas right frontal activity tends to be associated with avoidance-related emotions (e.g., fear, disgust; Davidson, Ekman, Saron, Senulis, & Friesen, 1990; Harmon-Jones & Allen, 1998). Thus, one may predict that days in which people exhibit greater left frontal activity (vs right frontal activity) would be associated with greater aggressive tendencies. We can test this hypothesis by modeling daily aggressive tendencies as a function of a level-1 predictor (daily left vs right frontal activity). To illustrate the influence of level-2 predictors, on both level-1 and level-2, we will include gender and trait agreeableness into our model, as males tend to be more physically aggressive than females in general (Bettencourt & Miller, 1996) and agreeableness is negatively associated with aggressiveness (Meier & Robinson, 2004; Meier, Robinson, & Wilkowski, 2006; Ode, Robinson, & Wilkowski, 2008; Wilkowski, Robinson, & Meier, 2006). One may predict that the association between daily left frontal activity and daily aggressive tendencies will be strongest among people who are male and among people who are low in agreeableness.

We will perform our MLM analyses using the *lmerTest* package (Version 3.1-2; Kuznetsova et al., 2017) in R (Version 4.0.2). The *lmerTest* package requires the *lme4* package to also be installed (Bates, Mächler, Bolker, & Walker, 2015). R is a free and open-source programming language with a powerful set of tools for statistical computing and graphics. Although MLM analyses with the *lmerTest* package are relatively straightforward, it will require some basic knowledge of coding in R. All data and code for the following analyses can be found on the Open Science Framework (at: https://osf.io/vprwb/?view_only=473085ce5bbd420d868b216e63a64fd0) for later reference.

METHOD

Participants

Participants were 106 (76 females) undergraduates from a public university in the Southeastern United States. Of the 106 participants, the mean age was 18.70 years (*SD*=1.53). 7.5% of the participants identified as more than one race, 4.6% were African American, 3.7% were Asian, 79.5% were Caucasian, and 4.6% did not report their race. Students who participated in the experiment were given partial research credit needed for their Introduction to Psychology course and were paid $20.

Measures

Trait agreeableness. Participants completed the agreeableness subscale of the Big Five Inventory (John, Donahue, & Kentle, 1991; John, Naumann, & Soto, 2008). Participants rated their agreement (1=s*trongly disagree, 5=strongly agree*) according to how much they believed each statement (e.g., “Is helpful and unselfish with others”, “Is generally trusting”) applied to them. Items were summed to create a composite score, where higher totals reflected greater levels of trait agreeableness (*M*=34.41, *SD*=5.82; Cronbach’s α=0.80).

Daily visual field bias. Each day participants completed a modified version of the line bisection task (Milner, Brechmann, & Pagliarini, 1991). A line with a perpendicular marker was placed on the screen. One line had the marker placed 5mm to the left of center, one had the marker in the center, and another had the marker 5mm to the right of center. Participants were asked to indicate whether the marker was to the left of center, to the right of center, or directly in the center. Items were scored such that inaccurate left responses yielded negative values and inaccurate right responses yielded positive values. Inaccurately choosing "right of center" means that one perceives the *left* side of the line as longer than it really is. Thus, after summing across items, higher values indicated left visual field bias (*M*=0.07, *SD*=0.56). Left visual field bias is a correlate of *right* frontal activation (Nash, McGregor, & Inzlicht, 2010).

Daily aggressive tendencies. Participants completed an abbreviated form of the physical (e.g., "Given enough provocation today, I might hit another person") and verbal (e.g., "If people were annoying me today, I would tell them what I think of them) aggression subscales of the Aggression Questionnaire (AQ; Buss & Perry, 1992), which was modified for daily use. This measure was used because self-reported propensities toward physical and verbal aggression are strongly related to behavioral aggression (Giancola & Parrott, 2008). Items were summed to form a composite score, such that higher numbers indicated greater levels of daily aggressive tendencies (*M*=7.98, *SD*=5.08; person-level α=0.96[1]).

Procedure

The study was conducted with the institutional internal review board's approval. Informed consent was obtained during the initial laboratory visit and then participants completed demographic questions and several surveys, including the Big Five Inventory. Following the initial visit, participants were given a URL code specific to them of a secure server to complete their daily logs. The participants were instructed to access this page each night for the following 21 days. Email reminders were sent to each participant nightly to remind them to fill out their log. Overall, 106 participants completed 1617 logs out of a possible 2226 days, indicating a 73% response rate. Participants averaged 15.25 days of responding (ranging from 1 to 21 days).

Results

Preliminary Analysis: Degree of Dependency

Computing the *intraclass correlation coefficient* (ICC) for our dependent variable is an important first step in any multilevel modeling analysis. The ICC indicates the degree of dependency in our observations. That is, it tells us what proportion of variance in our outcome is accounted for by the clustering of observations (i.e., daily logs *nested within* participants; Shrout & Fleiss, 1979). The ICC ranges from 0 to 1, with larger values indicating greater dependency in the data, and, thus, greater inflation of the Type I error rate if that dependency is not accounted for (Cohen et al., 2003). The formula for the ICC is:

$$\frac{\sigma_\tau^2}{(\sigma_\tau^2 + \sigma_W^2)}$$

such that σ_τ^2 represents the amount of variance in an outcome due to differences between people (in a repeated-measures design) and σ_W^2 represents the pooled within-person error variance. Thus, $\sigma_\tau^2 + \sigma_W^2$ represents the total observed variance in our outcome.

One obtains these parameters, σ_τ^2 and σ_W^2, by estimating an *unconditional (null) model* for the dependent variable. An unconditional model is a multilevel model with no predictors, just an intercept. The basic two-level model is as follows:

$$y_{ij} = \beta_{0j} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

such that y_{ij} represents daily reports of aggressive tendencies (with *i* reports for *j* individuals); β_{0j} is the mean aggressive tendencies for each person aggregated across days; e_{ij} is the day-level, within-person variance around that mean (which is our estimate of σ_W^2); γ_{00} is the grand mean for aggressive tendencies; and u_{0j} is the between-person variance around the grand mean (which is our estimate of σ_τ^2). Thus, the first equation is the day-level or within-person part of the model (level-1) and the second equation is the between-person part of the model (level-2). The multilevel model could then be reduced by replacing terms to the following form:

$$y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

We can use the *lmerTest* package in R to run our unconditional model for daily aggressive tendencies using the *lmer* (for *linear mixed-effects regression*) function:

Model.Null <- lmer(VerbPhysAggSum ~ 1 + (1 | ID), data = LineBisectionStudy)

With this code, we created the object “Model.Null”, which is the name of our unconditional model (object names are user-defined and can be whatever you like). Within the *lmer*() function, we insert the formula for our regression. The variable to the left of the “~” is our outcome or dependent variable (we named it “*VerbPhysAggSum*”) and everything to the right are the predictors. Because this is an unconditional model, there are no predictors. There is only an intercept, which we model with a “1”.

Furthermore, everything to the left of our inner parentheses are the fixed components of our estimates and everything within those inner parentheses are the random components. We specify the level-1 variables for which we want a random error term to the left of the “|” and the grouping (clustering) variable to the right. Because daily reports are nested with participants, participant ID is our grouping variable. Thus, we are predicting daily reports of aggressive tendencies with an intercept and a random error term around that intercept. Finally, we have to specify which data set (or dataframe in R terminology) to work from. After running the code for the model, we can use the *summary*() function to get our output, as in:

summary(Model.Null)

Within the random effects part of the output, one should see a variance for the intercept of 18.69. This is our estimate of σ_{τ}^{2} (the between-person variance). We also have a residual variance of 7.76. This is our estimate of σ_{W}^{2} (the within-person variance). Using the formula for the ICC, we get 0.71, which indicates that 29% of the variability in daily aggressive tendencies was within-person and 71% of the variability was due to between-person differences.

Model Building

Level-1 predictor. Our ICC indicates that there is a great amount of dependency in our daily reports of aggressive tendencies as a function of person-level clustering. Thus, use of MLM to test our main hypotheses is warranted. We predicted that days in which people exhibit greater left frontal activity (vs right frontal activity) would be associated with greater aggressive tendencies. We can modify our unconditional model to test these hypotheses by adding in a level-1 predictor for daily visual field bias (VFB). Because it is a level-1 predictor, we can also estimate a random error term for its slope. Level-1 coefficients with a random error term added are said to be *randomly varying*, whereas level-1 coefficients without the random error term are said to be *fixed* (which shouldn't be confused with the estimate of the fixed component itself—the average coefficient for that predictor). Nezlek (2011a) argues that all level-1 coefficients should be modeled as randomly varying, so long as those random error terms can be estimated reliably. Sometimes random error terms cannot be estimated reliably and their presence may cause failure in model convergence. In such cases, removal of a random error term may be warranted.

The updated level-1 equation would then be:

$$y_{ij} = \beta_{0j} + \beta_{1j}(VFB_{ij}) + e_{ij}$$

and the updated level-2 equations are:

$$\text{Intercept: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\text{VFB: } \beta_{1j} = \gamma_{10} + u_{1j}$$

such that γ_{10} is the average slope across participants describing the relationship between daily visual field bias and daily aggressive tendencies; and u_{1j} is the between-person random variance around that average slope. Replacing terms, the reduced form equation would then be:

$$y_{ij} = \gamma_{00} + \gamma_{10}(VFB_{ij}) + u_{0j} + u_{1j}(VFB_{ij}) + e_{ij}$$

However, prior to adding level-1 predictors to one's unconditional model, researchers should first consider centering options for their continuous variables. *Centering* refers to the process of transforming a variable into deviations around a fixed reference point. Centering is not as straightforward for MLM as it is for OLS-based regression, and there is no universal consensus among modelers on what type of centering is best. For a more thorough discussion of centering in the context of MLM, please see Enders and Tofighi (2007), Kreft and de Leeuw (1998), Kreft, de Leeuw, and Aiken (1995), and Bryk and Raudenbush (1992). Furthermore, centering is not a trivial matter in MLM. It's not employed just for facilitating the interpretation of model parameters. Rather, differences in centering options can actually change the models and hypotheses being tested.

Level-1 predictors can be entered either *uncentered*, *grand mean-centered* (i.e., centering each score around the variable's mean for all cases in the sample), or *group mean-centered* (i.e., centering each score around the cluster mean in which that particular case occurs; referred to as *person-centering* in repeated-measures designs or sometimes *centering within context*). Models with level-1 predictors entered either uncentered or grand mean-centered are statistically equivalent. However, differences arise with group mean-centering. To summarize the general arguments on centering in the literature cited above, when level-1 predictors are entered grand mean-centered, between-cluster differences actually influence level-1 parameter estimates. Thus, level-1 effects are confounded by level-2 differences. For instance, in the context of the present study, suppose that daily reports of visual field bias were associated with aggressive tendencies. If the predictor was grand mean-centered, then we wouldn't be able to separate daily fluctuations in visual field bias from general person-level individual differences. Perhaps more aggressive people tend to be biased toward the right visual field, in general, and less aggressive people tend to be biased toward the left? In this case, the level-1 test would say little about the impact of daily fluctuations in visual field bias on aggressive tendencies. For researchers interested in development or within-person change, those daily fluctuations are of interest.

When level-1 predictors are entered group mean-centered, level-2 differences are removed from level-1 parameter estimates (because cluster means are subtracted from each score within that cluster). Thus, instead of contributing to level-1 effects, those level-2 differences are *controlled*. In the context of our study, if visual field bias is person-centered, then the slope would represent predicted aggressiveness as a function of daily, within-person change in left versus right field bias compared to what is typical for each separate participant. Most authors tend to recommend group mean-centering level-1 predictors in MLM; however, there is disagreement about how level-2 differences should be addressed. Normally with group mean-centering, level-2 differences are removed. That removal treats the analysis as though all clusters have exactly the same mean, which would be erroneous if the researcher was actually interested in between-cluster differences (Kreft & de Leeuw, 1998). This error can be corrected by re-introducing the cluster means as a level-2 predictor of the level-1 intercept (however, see Hox et al., 2017 for a counter-argument against re-introducing cluster means).

In the following analysis, daily visual field bias was person-centered and the person-level means were entered as a level-2 predictor of the intercept (to re-introduce those level-2 differences). The level-1 equation becomes:

$$y_{ij} = \beta_{0j} + \beta_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}$$

and the level-2 equations become:

$$\text{Intercept: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\bar{X}_{VFB.j}) + u_{0j}$$

$$\text{VFB: } \beta_{1j} = \gamma_{10} + u_{1j}$$

such that γ_{01} is the average slope between individual differences in typical visual field bias and average daily aggressive tendencies across participants. Replacing terms, the reduced form equation would then be:

$$y_{ij} = \gamma_{00} + \gamma_{10}(VFB_{ij} - \bar{X}_{VFB.j}) + \gamma_{01}(\bar{X}_{VFB.j}) + u_{0j} + u_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}$$

To model our level-1 predictor, we can modify our code for the *lmer()* function:

```
MLMLevel1model <- lmer(VerbPhysAggSum ~ 1 + GRCbias + GRPMnbias + (1
+ GRCbias | ID), data = LineBisectionStudy)
```

We created our new object (MLMLevel1model) by adding two fixed components and one random error term to our unconditional model. One fixed component represents the average slope for daily visual field bias, person-centered (GRCbias), and, because it is a level-1 variable, we can model a random error term for it. The next fixed component represents the slope for person-level individual differences in visual field bias, on average, across the study period (GRPMnbias), which is a level-2 predictor and, thus, has no random error term (this is a two-level model, so there are no level-3 units for this slope to vary across).

Using the same *summary()* function as before (with the new model name entered), we obtain our results. Again, we predicted that days in which people exhibited greater right visual field bias (as opposed to left visual field bias) would be associated with greater aggressive tendencies, given the literature linking right visual field bias, left frontal activity, behavioral approach, and aggression (e.g., Harmon-Jones & Peterson, 2008; Nash et al., 2010). Contrary to hypotheses, daily increases in *left* visual field bias (an indicator of *right* frontal activity) was positively associated with aggressive tendencies, b=0.76, t(33.35)=2.28, p=0.029.[2] Furthermore, people who tended to exhibit greater left visual field bias, in general, over the study period also tended to report greater aggressive tendencies compared to those who generally exhibited right visual field bias, b=2.18, t(111.48)=2.13, p=0.035.

Level-2 predictors and cross-level interactions. We will now demonstrate how to explore traditional level-2 predictors and potential cross-level interactions. Cross-level interactions allow us to examine whether level-1 slopes vary as a function of level-2 predictors (sometimes referred to as *slopes-as-outcomes* analysis; Nezlek, 2001). In the present study, we included the level-2 (person-level) variables of gender and trait agreeableness, because males tend to be more physically aggressive than females (e.g., Bettencourt & Miller, 1996) and trait agreeableness tends to be negatively associated with aggressiveness (e.g., Ode, Robinson, & Wilkowski, 2008). We will modify our level-2 equations to include main effects (slopes) for gender and agreeableness, as well as one gender x visual field bias interaction and one agreeableness x visual field bias interaction. The level-1 equation remains:

$$y_{ij} = \beta_{0j} + \beta_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}$$

however, the level-2 equations change to:

$$\text{Intercept: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\bar{X}_{VFB.j}) + \gamma_{02}(Gender_{.j}) + \gamma_{03}(Agreeableness_{.j}) + u_{0j}$$

$$\text{VFB: } \beta_{1j} = \gamma_{10} + \gamma_{11}(Gender_{.j}) + \gamma_{12}(Agreeableness_{.j}) + u_{1j}$$

such that γ_{02} represents the slope between gender and average daily aggressive tendencies across participants (i.e., the main effect for gender); γ_{03} represents the slope between trait agreeableness and average daily aggressive tendencies across participants (i.e., the main effect for agreeableness); γ_{11} represents the cross-level interaction between gender and visual field bias (i.e., we are predicting whether the slope between daily visual field bias and aggressive tendencies varies as a function of gender); and γ_{12} represents the cross-level interaction between trait agreeableness and daily visual field bias. Replacing terms, the reduced form equation would then be:

$$\begin{aligned}
y_{ij} &= \gamma_{00} + \gamma_{10}(VFB_{ij} - \bar{X}_{VFB.j}) + \gamma_{01}(\bar{X}_{VFB.j}) + \gamma_{02}(Gender_{.j}) \\
&+ \gamma_{03}(Agreeableness_{.j}) + \gamma_{11}((VFB_{ij} - \bar{X}_{VFB.j}) * Gender_{.j}) \\
&+ \gamma_{12}((VFB_{ij} - \bar{X}_{VFB.j}) * Agreeableness_{.j}) + u_{0j} + u_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}
\end{aligned}$$

In a two-level model, entering level-2 predictors is fairly straightforward. For continuous variables, predictors can be entered either uncentered or grand mean-centered. The choice will not change model tests or hypotheses, because the options are statistically equivalent. We tend to enter continuous level-2 predictors grand mean-centered to facilitate interpretation of model coefficients. Categorical predictors, on the other hand, should be entered uncentered and either dummy or contrast-coded (Cohen et al., 2003; Nezlek, 2011a).

To model our level-2 predictors, we can modify our code for the *lmer()* function as such:

```
MLMmodelCross <- lmer(VerbPhysAggSum ~ 1 + GRCbias + GRPMnbias +
GMCagreeable + Male + GRCbias:GMCagreeable + GRCbias:Male + (1 +
GRCbias | ID), data = LineBisectionStudy)
```

We added fixed components for gender (dummy-coded as 0=Female and 1=Male) and trait agreeableness (GMCagreeable), as well as the relevant cross-level interactions (the *lmer*() function uses ":" to indicate interactions). Recall that level-2 predictors do not have random error terms in a two-level model, so no adjustment is needed to the random part of the code.

Using the same *summary()* function as before (with the new model name entered), we obtain our results (see Table 1 for results of final model). As predicted, males tended to report more aggressive tendencies across the study period, and trait agreeableness was negatively associated with aggressive tendencies across the study period. Furthermore, the association between daily visual field bias and aggressive tendencies was qualified by a visual field bias x gender interaction. There was not a cross-level interaction with trait agreeableness.

Table 1

Daily Aggressive Tendencies as a Function of Visual Field Bias (VFB; Person-Centered), Gender, and Trait Agreeableness (Final Model)

Parameter	Variable	*Estimate*	*t*-value	*df*	*p*-value
gamma_00	Intercept	7.45	16.52	97.90	<.001
gamma_10	VFB (person-centered)	1.33	3.44	30.92	0.0017
gamma_01	VFB (person means)	1.18	1.21	114.06	0.23
gamma_02	Gender	2.63	2.92	106.05	0.004
gamma_03	Agreeableness	-0.27	-3.98	99.53	0.0001
gamma_11	VFB x Gender	-1.71	-2.54	31.88	0.016
gamma_12	VFB x Agreeableness	-0.07	-1.18	30.70	0.25

To evaluate the nature of significant interactions in MLM, researchers can use procedures recommended by Aiken and West (1991) and Preacher and colleagues (2006). Computational tools for probing simple slopes for interactions in MLM can be found here: http://quantpsy.org/interact/hlm2.htm. Although, daily visual field bias was generally not associated with aggressive tendencies for males, *b*=-0.38, *t*=-0.69, *p*=0.49, males who exhibited right visual field bias were much more aggressive than females who exhibited similar right visual field bias, *b*=3.36, *t*=3.51, *p*=0.0007, as may be anticipated. However, for females, days with greater left field visual bias (versus right field bias) were associated with more aggressive days, *b*=1.33, *t*=3.44, *p*=0.0008. Because we used the *lmerTest* package to model our data, we can easily use the *sjPlot* package (version 2.8.4; Lüdecke, 2020) to plot our interaction and simple slopes with the *plot_model()* function, as in:

plot_model(MLMmodelCross, type = "int", mdrt.values = "minmax", colors = "bw", title = "", axis.title = c("Visual Field Bias", "Aggressive Tendencies"))

such that the argument, *type* = *"int",* indicates that we want to plot the interaction for the specified model name, and the argument, *mdrt.values* = *"minmax"*, indicates what values of the moderator for which we want simple slopes. The minimum/maximum value was chosen because our moderator in the visual field bias x gender interaction (i.e., gender) is categorical. For continuous moderators, researchers can choose the mean of the moderator and +/-1 *SD* (*"meansd"*). We also removed a plot title and added titles to our axes. The argument, *colors* = *"bw",* indicates that the plot is to be in black and white, so that group trajectories can be distinguished by dashed versus solid lines; however, this option can be modified to allow colored lines. Please see Figure 3 for the resulting plot.

CONCLUSION

The present article described some of the basic concepts and techniques involved in modeling multilevel data structures. Furthermore, this article serves as a tutorial for implementing basic MLM techniques in the R package, *lmerTest* (version 3.1-2; Kuznetsova et al., 2017) with real, publicly available data. The study and analyses were presented within the context of intensive longitudinal repeated-measures designs.

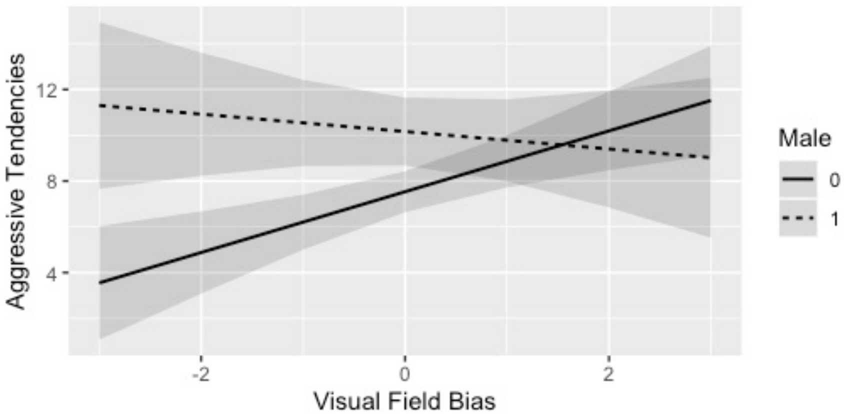

Figure 3. Daily aggressive tendencies as a function of daily fluctuations in visual field bias and gender. The shaded regions are confidence intervals around the slope lines, which can be removed with "ci.lvl=NA".

Specifically, the research design yielded interval-contingent data, where participants reported on their thoughts, feelings, and behaviors each evening for a 21-day period. This is a particularly powerful research design for studying growth and change in naturally occurring phenomena. Such designs appear to be on the rise among pediatric neuropsychologists, especially those interested in developmental trajectories of neurological disorders, behavior problems, and brain injury (e.g., Infante et al., 2020; Séguin et al., 2020). In such work, it is imperative that the researcher does not ignore the hierarchical structure of the data that were collected. Not only will ignoring cluster-level dependency inflate one's Type I error rate (i.e., increasing the risk of falsely declaring findings significant), but it also provides an inadequate and inaccurate representation of the sampled observations, as well as the populations they describe. Ultimately, this will contribute to the replicability and generalizability of one's research findings.

The present paper was not meant to be exhaustive but rather demonstrate basic analyses in MLM. Although our article presented the analysis of a two-level model with interval-contingent data, the principles and techniques discussed can be generalized to the analysis of event-contingent or group/organizational data sets. Furthermore, the techniques discussed can be easily extended to accommodate three-level data structures. There are also a number of advanced topics in MLM that go beyond the scope of the present article. These include effect sizes, determining the statistical power of multilevel designs, mediation analyses, nonlinear and categorical dependent variables, and scale reliability. Interested readers are encouraged to read the work of Kreft and de Leeuw (1998), Raudenbush and Bryk (2002), Hox, Moerbeek, and van de Schoot (2017), Nezlek (2011b), Snijders and Bosker (2012), and Bauer and colleagues (2006). With that stated, we hope that this article will serve as a resource for researchers interested in implementing multilevel modeling in their own work.

REFERENCES

Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions.* London: Sage.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11,* 142-163.

Bettencourt, B., & Miller, N. (1996). Gender differences in aggression as a function of provocation: A meta-analysis. *Psychological bulletin*, *119*, 422-447.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing live as it is lived. *Annual Review of Psychology, 54,* 579-616.

Bonito, J. A., Ruppel, E. K., & Keyton, J. (2012). Reliability estimates for multilevel designs in group research. *Small Group Research, 43*, 443-467.

Buss, A. H., & Perry, M. P. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63,* 452-459.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Carver, C. S. (2004). Negative affects deriving from the behavioral ap- proach system. *Emotion, 4,* 3-22.

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin, 135,* 183-204.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd Edition)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Davidson, R., Ekman, P., Saron, C., Senulis, J., & Friesen, W. (1990). Approach withdrawal and cerebral asymmetry: Emotional expression and brain physiology. *Journal of Personality and Social Psychology, 58*, 330-341.

DeWall, C. N., Lambert, N. M., Pond, R. S. Jr., Kashdan, T. B., & Fincham, F. D. (2012). A grateful heart is a nonviolent heart: Cross-sectional, experience sampling, longitudinal, and experimental evidence. *Social Psychological and Personality Science*, *3*, 232-240.

Drake, R. A., & Myers, L. R. (2006). Visual attention, emotion, and action tendency: Feeling active or passive. *Cognition and Emotion, 20*, 608-622.

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

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19,* 72-91.

Giancola, P. R., & Parrott, D. J. (2008). Further evidence for the construct validity of laboratory aggression paradigms. *Aggressive Behavior, 34,* 214-229.

Gilbert, L. R., Pond, R. S. Jr., Haak, E. A., DeWall, C. N., & Keller, P. S. (2015). Sleep problems exacerbate the emotional consequences of interpersonal rejection. *Journal of Social and Clinical Psychology*, *34*, 50-63.

Harmon-Jones, E. (2003). Anger and the behavioral approach system. *Personality and Individual Differences, 35,* 995-1005.

Harmon-Jones, E., & Allen, J. J. B. (1998). Anger and prefrontal brain activity: EEG asymmetry consistent with approach motivation despite negative affective valence. *Journal of Personality and Social Psychology, 74*, 1310-1316.

Harmon-Jones, E., Harmon-Jones, C., Abramson, L., & Peterson, C. K. (2009). PANAS positive activation is associated with anger. *Emotion, 9,* 183-196.

Harmon-Jones, E., & Peterson, C. K. (2008). Effect of trait and state approach motivation on aggression. *Journal of Research in Personality, 42,* 1381-1385.

Harmon-Jones, E., & Peterson, C. K. (2009). Supine body position reduces neural response to anger evocation. *Psychological Science, 20,* 1209-1210.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23,* 723-744.

Hox, J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). New York, NY: Routledge.

Hox, J. J., & Kleiboer, A. M. (2007). Retrospective questions or a diary method? A two level multitrait-multimethod analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 311-325.

Infante, M. A., Nguyen-Louie, T. T., Worley, M., Courtney, K. E., Coronado, C., & Jacobus, J. (2020). Neuropsychological trajectories associated with adolescent alcohol and cannabis use: A prospective 14-year study. *Journal of the International Neuropsychological Society, 26,* 480-491.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory--Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic Data Analysis.* New York, NY: Guildford Press.

Kreft, I. G., & de Leeuw, J. (1994). The gender gap in earnings: A two-way nested multiple regression analysis with random effects. *Sociological Methods & Research*, *22*, 319-341.

Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling.* Newbury Park, CA: Sage.

Kreft, I. G., de Leeuw, J., & Aiken, L. (1995). The effects of different forms of centering on hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1-22.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82,* 1-26.

Lai, M. H. C. (in press). Composite reliability of multilevel data: It's about observed scores and construct meanings. *Psychological Methods,*

Lüdecke, D. (2020). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.4, URL: https://CRAN.R-project.org/package=sjPlot.

Meier, B. P., & Robinson, M. D. (2004). Does quick to blame mean quick to anger?: The role of agreeableness in dissociating blame and anger. *Personality and Social Psychology Bulletin, 30*, 856-867.

Meier, B. P., Robinson, M. D., & Wilkowski, B. M. (2006). Turning the other cheek: Agreeableness and the regulation of aggression-related primes. *Psychological Science, 17*,136-142.

Milner, A. D., Brechmann, M., & Pagliarini, L. (1992). To halve and to halve not: An analysis of line bisection in normal subjects. *Neuropsychologia, 30*, 515-526.

Nash, K., McGregor, I., & Inzlicht, M. (2010). Line bisection as a neural marker of approach motivation. *Psychophysiology*, *47*, 979-983.

Nezlek, J. B. (1995). Social construction, gender/sex similarity, and social interaction in close personal relationships. *Journal of Social and Personal Relationships*, *12*, 503-520.

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event and interval contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27*, 771-785.

Nezlek, J. B. (2007). Multilevel modeling in research on personality. In R. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 502-523). New York: Guilford.

Nezlek, J. B. (2011a). Multilevel modeling analyses of diary-style data. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 357-383). New York: Guilford.

Nezlek, J. B. (2011b). *Multilevel modeling for social and personality psychology.* Thousand Oaks, CA: Sage.

Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within person variability. *Journal of Research in Personality, 69,* 149-155.

Nezlek, J. B., Hampton, C., & Shean, G. D. (2000). Clinical depression and everyday social interaction in a community sample. *Journal of Abnormal Psychology, 107*, 11-19.

Ode, S., Robinson, M. D., & Wilkowski, B. M. (2008). Can one's temper be cooled?: A role for agreeableness in moderating neuroticism's influence on anger and aggression. *Journal of research in personality*, *42*, 295-311.

Pond, R. S. Jr., DeWall, C. N., Lambert, N. M., Deckman, T., Bonser, I. M., & Fincham, F. D. (2012). Repulsed by violence: Disgust sensitivity buffers trait, behavioral, and daily aggression. *Journal of Personality and Social Psychology, 102,* 175-188.

Pond, R. S. Jr., Kashdan, T. B., DeWall, C. N., Savostyanova, A., Lambert, N. M., & Fincham, F. D. (2012). Emotion differentiation moderates aggressive tendencies in angry people: A daily diary analysis. *Emotion*, *12*, 326-337.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437-448.

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 1-17.

Raudenbush, S. W., & Bryk, A. S. (1988). Chapter 10: Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, *15*, 423-475.

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

Séguin, M., Dégeilh, F., Bernier, A., El-Jalbout, R., & Beauchamp, M. H. (2020). It's a matter of surgency: Traumatic brain injury is associated with changes in preschoolers' temperament. *Neuropsychology, 34*(4), 375-387.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*, 420-428.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323-355.

Smits, D. J. M., & Kuppens, P. (2005). The relations between anger, coping with anger, and aggression, and the BIS/BAS system. *Personality and Individual Differences, 39,* 783-793.

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

Sullivan, E. V., Brumback, T., Tapert, S. F., Brown, S. A., Baker, F. C., Colrain, I. M., Prouty, D., De Bellis, M. D., Clark, D. B., Nagel, B. J., Pohl, K. M., & Pfefferbaum, A. (2019). Disturbed cerebellar growth trajectories in adolescents who initiate alcohol drinking. *Biological Psychiatry, 87,* 632-644.

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday events: Origins, types, and uses. *Journal of Personality, 59,* 339-354.

West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality, 59*, 609-662.

Weston, S. J., & Yee, D. (2017). Why you should become a UseR: A brief introduction to R. *APS Observer*, *30*.

Wilkowski, B. M., & Meier, B. P. (2010). Bring it on: Angry facial expressions potentiate approach-motivated motor behavior. *Journal of Personality and Social Psychology, 98,* 201-210.

Wilkowski, B. M., Robinson, M. D., & Meier, B. P. (2006). Agreeableness and the prolonged spatial processing of antisocial and prosocial information. *Journal of Research in Personality, 40*, 1152-1168.

FOOTNOTES

[1]Although it is beyond the scope of the present article, it should be noted that traditional measurements of reliability (e.g., Cronbach's α) do not correctly break down the total variance of multilevel data into the necessary error and systematic variances at each level (Hox & Kleiboer, 2007). Please see Nezlek (2007, 2011a, 2011b, 2017), Bonito and colleagues (2012), Geldhof and colleagues (2014), and Lai (in press) for recommendations on testing scale reliability with nested data. Furthermore, differences in variances of items can impact reliability estimates. Raudenbush and Bryk (2002) suggest rescaling items at level-1 to achieve relatively equal error variances across items, if needed.

[2] Multilevel modeling with the *lmerTest* package uses the Satterthwaithe approximation for obtaining degrees of freedom, which is a weighted average of the between and within degrees of freedom. Therefore, the Satterthwaithe approximation results in fractional degrees of freedom. Kenny, Kashy, and Cook (2006) recommend the Satterthwaithe estimate of degrees of freedom because it takes into account the mixture of between and within parts of the estimate.

Author note

Richard S. Pond, Jr., Matison W. McCool, Department of Psychology, University of North Carolina Wilmington; Brian A. Bulla, Department of Psychology, The University of Southern Mississippi.

This study was funded by a Charles L. Cahill Award for Faculty Research and Development awarded to Richard Pond from the University of North Carolina Wilmington.

Please address correspondence to Richard S. Pond, Jr., Teaching Laboratory 2080, Department of Psychology, University of North Carolina Wilmington, Wilmington, NC 28403. Email: pondr@uncw.edu. Phone: 910.962.3372. Fax: 910.962.7010.

Declarations

Funding: This study was funded by a Charles L. Cahill Award for Faculty Research and Development awarded to Richard Pond from the University of North Carolina Wilmington.

Conflicts of interests/Competing interests: Not applicable

Availability of data and materials: All data and code for the manuscript can be found on the Open Science Framework (at: https://osf.io/vprwb/?view_only=473085ce5bbd420d868b216e63a64fd0).

Code availability: All data and code for the manuscript can be found on the Open Science Framework (at: https://osf.io/vprwb/?view_only=473085ce5bbd420d868b216e63a64fd0).

Authors' contributions: Richard Pond contributed to the collection and analysis of the data, as well as manuscript preparation; Matison McCool contributed to data analysis and manuscript preparation; Brian Bulla contributed to data collection.