

9-1-2021

Multilevel Modeling of Interval-Contingent Data In Neuropsychology Research Using the *ImerTest* 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

Follow this and additional works at: https://aquila.usm.edu/fac_pubs



Part of the [Psychology Commons](#)

Recommended Citation

Pond, R. S., McCool, M. W., Bulla, B. A. (2021). Multilevel Modeling of Interval-Contingent Data In Neuropsychology Research Using the *ImerTest* Package In R. *Journal of Pediatric Neuropsychology*, 7, 102-112.

Available at: https://aquila.usm.edu/fac_pubs/20685

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Faculty Publications by an authorized administrator of The Aquila Digital Community. For more information, please contact aquilastaff@usm.edu.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

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

Richard S. Pond, Jr.¹

Matison W. McCool¹

Brian A. Bulla²

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](https://orcid.org/0000-0001-5446-7456)

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

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

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

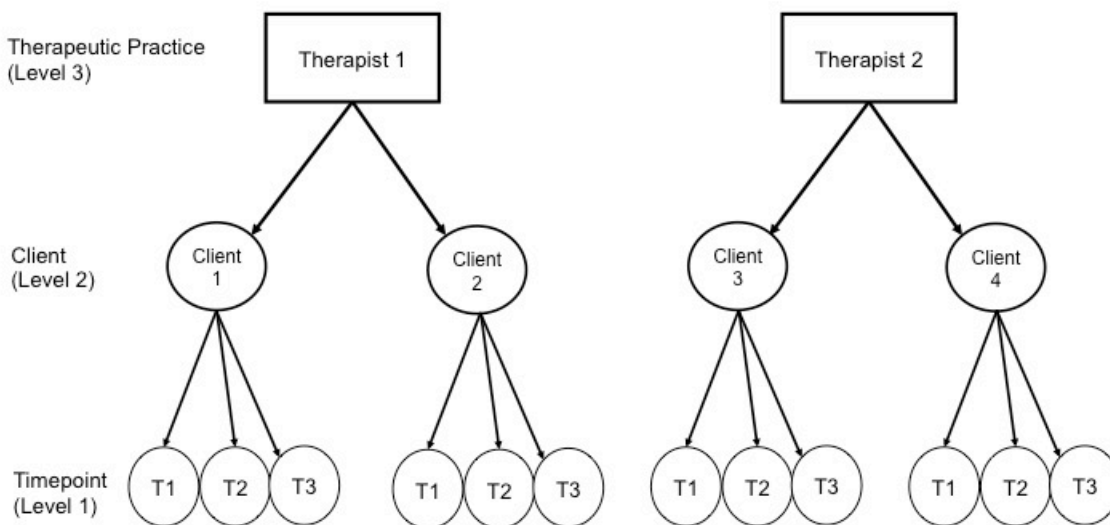
54 Over the past three decades, researchers across the social and behavioral
55 sciences have increasingly become interested in using intensive longitudinal designs to
56 study naturally occurring phenomena. This is particularly the case for pediatric
57 neuropsychologists, who may be interested in the developmental trajectories of
58 neurological disorders and behavior problems (e.g., Infante, Nguyen-Louie, Worley,
59 Courtney, Coronado, & Jacobus, 2020; Sullivan, Brumback, Tapert, Brown, Baker,
60 Colrain, Prouty, De Bellis, Clark, Nagel, Pohl, & Pfefferbaum, 2019) or the long-term
61 consequences of traumatic brain injury (e.g., Séguin, Dégeilh, Bernier, El-Jalbout, &
62 Beauchamp, 2020). For instance, a recent Google Scholar search of the terms
63 “developmental trajectory” and “neuropsychology” yielded over 6,000 hits just for the
64 year 2020. Neuropsychologists can employ such designs to examine microlevel
65 processes and investigate how spontaneous experiences in their natural environment
66 interact with stable characteristics to influence behavior, without the limitations imposed
67 on by retrospective reporting (Bolger, Davis, & Rafaeli, 2003; Wheeler & Reis, 1991).
68 However, the complexity of the data yielded in such designs often present the
69 researcher with analytic challenges. The goal of the present article is to introduce
70 multilevel modeling (MLM; sometimes referred to as hierarchical linear modeling or
71 linear mixed-models), which is thought to be one of the best methods for analyzing data
72 generated from intensive repeated assessment (Nezlek, 2011a). A second goal is to
73 illustrate how to perform MLM analyses in R (i.e., a free and open-source programming
74 language for statistical computing and graphics; see Weston & Yee, 2017) with the

75 *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) and real data. This
76 article is intended for researchers with primary experience in ordinary least squares
77 (OLS)-based techniques, such as regression and analysis of variance (ANOVA). For
78 those readers who are interested in a more thorough explication of MLM, there are a
79 number of excellent, authoritative introductions to the technique, including: Raudenbush
80 and Bryk (2002), Kreft and de Leeuw (1998), and Snijders and Bosker (2012). A
81 particularly user-friendly resource for the non-expert is Nezlek's (2011b) brief
82 introduction to MLM.

83 **Conceptual Background**

84 Multilevel, or hierarchical, data structures are clustered data structures
85 (sometimes referred to as *nested* data), such that observations at one level of analysis
86 (e.g., ratings of daily stressors each night for a 30-day period) are clustered (or *nested*)
87 within observations at another level of analysis (e.g., individual clients). Nesting can
88 occur within participants, as in the case of studies that employ an intensive longitudinal
89 or repeated-measures design. However, nesting can also occur between groups.
90 Indeed, the fields of education (where students are nested within classrooms or
91 schools) and organizational psychology and economics (where workers are nested
92 within companies or industries) each have long, rich histories of studying and modeling
93 multilevel data structures (e.g., Bryk & Raudenbush, 1992; Hoffman, 1997; Kreft & de
94 Leeuw, 1994; Raudenbush & Bryk, 1986, 1988). This clustering or nesting creates
95 dependency in the data, such that subsets of cases or responses are more similar to
96 each other and, thus, more highly correlated with each other than with cases or
97 responses in other subsets. The current article focuses on data structures typical in

98 studying repeated assessments of naturally occurring phenomena; however, the
99 principles and techniques discussed below are also relevant to grouped or
100 organizational data structures. Figure 1 illustrates a three-level hierarchical data
101 structure potentially relevant to intensive longitudinal research designs in pediatric
102 neuropsychology.
103



104
105 *Figure 1. Example illustration of a three-level data structure, such that responses or*
106 *assessments over time are nested within clients, and clients are nested within*
107 *therapeutic practices.*

108
109 Data from intensive repeated assessments can be further described as either
110 *interval-contingent* or *event-contingent* (Nezlek, 2001, 2011a; Wheeler & Reis, 1991).
111 Interval-contingent data occurs when participants provide data at a certain interval. For
112 example, in diary-style research, participants typically report on their daily mood and

113 relationships once per day (often at night), and daily responses are treated as being
114 nested within participants (e.g., Gilbert, Pond, Haak, DeWall, & Keller, 2015; Pond,
115 Kashdan, DeWall, Savostyanova, Lambert, & Fincham, 2012). Event-contingent data
116 occur when a specific type of event triggers data collection efforts. For instance, in
117 experience-sampling research, it's common to have participants make reports each time
118 they engage in a social interaction (e.g., DeWall, Lambert, Pond, Kashdan, & Fincham,
119 2012; Nezlek, 1995; Nezlek, Hampton, & Shean, 2000), and interactions are treated as
120 being nested within participants. Although the present article focuses on interval-
121 contingent data, the principles and techniques discussed below also apply to event-
122 contingent data.

123 **Limitations of OLS**

124 The dependency of nested data inherently violates the assumption of
125 independence in OLS regression. The consequence of this dependency is that the
126 standard errors produced by OLS are too small (Cohen, Cohen, Aiken, & West, 2003).
127 This means that any confidence intervals or significance tests generated with those
128 standard errors will be biased towards over-estimating significance (i.e., Type I error
129 inflation). The more similar scores are within clusters, the more serious the problem.
130 Furthermore, dependency is a particular concern for data that are nested as a result of
131 repeated-assessments over time, which may be more relevant to the research designs
132 of pediatric neuropsychologists (e.g., Infante et al., 2020), as opposed to nesting due to
133 group membership (e.g., students nested within classrooms or people nested within
134 families). The dependency that occurs in longitudinal designs deserves special attention
135 here because of a phenomenon known as autocorrelation. Autocorrelation is present

136 when adjacent observations are more strongly related to each other than nonadjacent
137 observations (West & Hepworth, 1991). In longitudinal designs, observations that are
138 closer in time tend to be more strongly related to each other than observations spaced
139 further apart (e.g., the stress ratings of participants on day 1 will tend to be strongly
140 correlated with their stress ratings on day 2 than on day 25 in a month-long diary study).
141 Autocorrelation reduces estimates of within-person variability, biases standard errors,
142 and inflates the Type I error rate of significance tests (Bolger et al., 2003).

143 OLS-based analytic strategies for handling multilevel data structures generally
144 fall into two categories (Cohen et al., 2003; Nezlek, 2001). The first is to completely
145 ignore the nesting and analyze responses as if there were no hierarchical structure
146 (*disaggregated analysis*). This method is most at risk of Type I error inflation; however,
147 dummy codes for cluster membership can be entered as predictors to partial out the
148 average influence of cluster. This modified approach is limited in that: 1. it treats the
149 cluster effect as *fixed*, meaning that one can only generalize to the specific clusters
150 (e.g., specific persons or groups) that were represented in the sample and not a larger
151 population from which one's clusters may have been randomly sampled (Cohen et al.,
152 2003), and 2. it ignores between-cluster variation (i.e., the associations among modeled
153 variables may be different for some clusters compared to others) (Nezlek, 2001).

154 The second OLS-based approach to handling multilevel data is to take the
155 cluster average for each predictor and outcome and treat cluster as the unit of analysis
156 in one's regression (*aggregated analysis*). Aggregated analysis leads to a loss of
157 information, and, thus, power, and the reliability of the observations that go into each
158 cluster mean is not taken into account (Nezlek, 2001). Perhaps most concerning is that

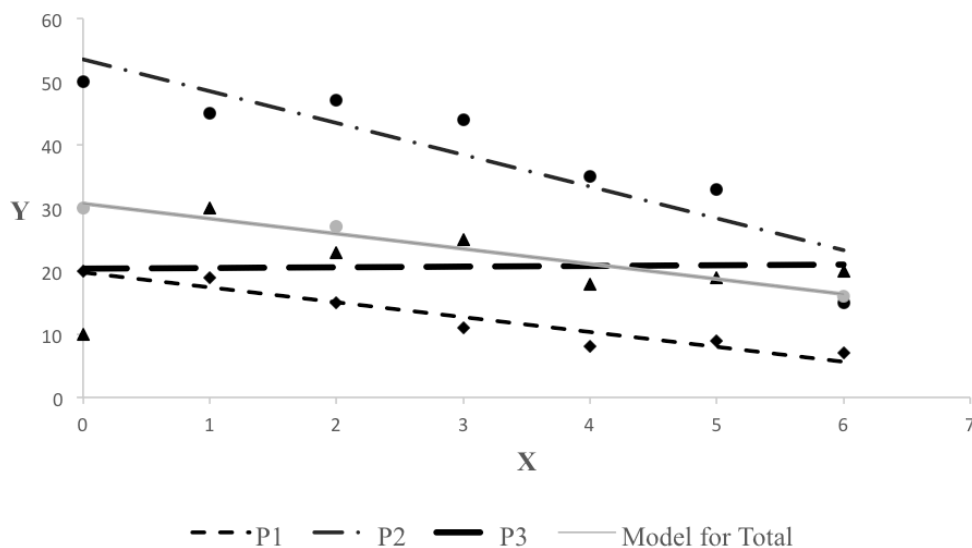
159 the results of aggregated analyses are limited to the cluster level; results cannot be
160 generalized to level 1 units (Cohen et al., 2003). This can promote misleading
161 conclusions. For instance, does it make sense to suggest someone suffers from chronic
162 stress based on this person's average level of stress reported across a 7-day period?
163 Perhaps, this person just had a stressful few days.

164 **Advantages of Multilevel Modeling**

165 Multilevel modeling is an appropriate technique for modeling clustered data,
166 because it was designed specifically to account for the cluster-level dependency that
167 violates the assumption of independence in OLS regression. That is, MLM accounts for
168 cluster-level dependency because it models multiple error terms across each level of
169 analysis. It accomplishes this by estimating a regression equation for each unit of
170 analysis at the lowest level, and then those coefficients (i.e., intercepts and slopes) are
171 used as dependent variables in regressions at the next level of analysis (Nezlek,
172 2011b). Furthermore, despite separate equations for each level of analysis, all
173 coefficients are estimated with maximum likelihood procedures simultaneously, and
174 lower-level coefficients are estimated with two parameters: a *fixed* component and a
175 *random* error term. In the context of a repeated-measures study, the fixed component
176 would represent the average value for a parameter of interest (i.e., intercept or slope)
177 across all subjects, whereas the random error would represent how much those
178 parameters tend to vary across subjects. Thus, when one models repeated-measures
179 with MLM, the model yields the average intercept and slope(s) across subjects (fixed
180 components); however, the presence of significant random variance indicates that the
181 intercepts and slope(s) significantly vary across each subject (as illustrated in Figure 2).

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

189



190

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

195

196 MLM not only accounts for the nested structure of data, but it also uses
 197 algorithms that apply Bayes shrinkage to weight observations by their reliabilities. This

198 method moves outliers, or less reliable observations, closer to the mean (Nezlek,
199 2011b). By using Bayes shrinkage, more accurate estimates, relative to population
200 measures, are achieved when compared to procedures that do not employ Bayes
201 shrinkage (Littell et al., 1996; Raudenbush & Bryk, 2002). A final advantage to using
202 MLM is that multilevel models do not require complete data sets. OLS-based
203 procedures (e.g., regression, ANOVA) perform best with balanced designs and the
204 absence of missing data. Oftentimes, in repeated-measures designs, cases with
205 missing data at one time-point need to be deleted. For MLM, when data are missing on
206 an outcome at a certain time-point, whole cases do not need to be deleted, nor do
207 missing values on y need to be imputed. Instead, parameters are estimated with all
208 available data using maximum likelihood procedures.

209 **An Illustrative Example: Cerebral Asymmetry and Aggressive Tendencies**

210 In the remainder of this article, we will demonstrate multilevel modeling with real
211 interval-contingent data collected from a sample of young adults ($M_{\text{age}}=18.70$) across a
212 21-day period. We will use these data to test the hypothesis that within-person
213 differences in cerebral asymmetry are associated with daily aggressive tendencies. The
214 research literature shows that approach motivation, as well as traits and emotions linked
215 to approach motivation, are positively associated with aggressive behavior (e.g., Carver,
216 2004; Carver & Harmon-Jones, 2009; Harmon-Jones, 2003; Harmon-Jones, Harmon-
217 Jones, Abramson, & Peterson, 2009; Harmon-Jones & Peterson, 2008, 2009; Wilkowski
218 & Meier, 2010). On the other hand, traits and emotions related to behavioral avoidance
219 are typically inversely associated with aggression (e.g., Harmon-Jones, 2003; Pond,
220 DeWall, Lambert, Deckman, Bonser, & Fincham, 2012; Smits & Kuppens, 2005).

221 Asymmetrical activation in the frontal cortices is closely linked to both approach and
222 avoidance motivation. Indeed, greater left frontal activity tends to be associated with
223 approach-related emotions (e.g., anger, joy; Drake & Myers, 2006), whereas right
224 frontal activity tends to be associated with avoidance-related emotions (e.g., fear,
225 disgust; Davidson, Ekman, Saron, Senulis, & Friesen, 1990; Harmon-Jones & Allen,
226 1998). Thus, one may predict that days in which people exhibit greater left frontal
227 activity (vs right frontal activity) would be associated with greater aggressive tendencies.
228 We can test this hypothesis by modeling daily aggressive tendencies as a function of a
229 level-1 predictor (daily left vs right frontal activity). To illustrate the influence of level-2
230 predictors, on both level-1 and level-2, we will include gender and trait agreeableness
231 into our model, as males tend to be more physically aggressive than females in general
232 (Bettencourt & Miller, 1996) and agreeableness is negatively associated with
233 aggressiveness (Meier & Robinson, 2004; Meier, Robinson, & Wilkowski, 2006; Ode,
234 Robinson, & Wilkowski, 2008; Wilkowski, Robinson, & Meier, 2006). One may predict
235 that the association between daily left frontal activity and daily aggressive tendencies
236 will be strongest among people who are male and among people who are low in
237 agreeableness.

238 We will perform our MLM analyses using the *lmerTest* package (Version 3.1-2;
239 Kuznetsova et al., 2017) in R (Version 4.0.2). The *lmerTest* package requires the *lme4*
240 package to also be installed (Bates, Mächler, Bolker, & Walker, 2015). R is a free and
241 open-source programming language with a powerful set of tools for statistical computing
242 and graphics. Although MLM analyses with the *lmerTest* package are relatively
243 straightforward, it will require some basic knowledge of coding in R. All data and code

244 for the following analyses can be found on the Open Science Framework (at:
245 https://osf.io/vprwb/?view_only=473085ce5bbd420d868b216e63a64fd0) for later
246 reference.

247 **METHOD**

248 **Participants**

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

255 **Measures**

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

263 *Daily visual field bias.* Each day participants completed a modified version of the
264 line bisection task (Milner, Brechmann, & Pagliarini, 1991). A line with a perpendicular
265 marker was placed on the screen. One line had the marker placed 5mm to the left of
266 center, one had the marker in the center, and another had the marker 5mm to the right

267 of center. Participants were asked to indicate whether the marker was to the left of
268 center, to the right of center, or directly in the center. Items were scored such that
269 inaccurate left responses yielded negative values and inaccurate right responses
270 yielded positive values. Inaccurately choosing “right of center” means that one
271 perceives the *left* side of the line as longer than it really is. Thus, after summing across
272 items, higher values indicated left visual field bias ($M=0.07$, $SD=0.56$). Left visual field
273 bias is a correlate of *right* frontal activation (Nash, McGregor, & Inzlicht, 2010).

274 *Daily aggressive tendencies.* Participants completed an abbreviated form of the
275 physical (e.g., “Given enough provocation today, I might hit another person”) and verbal
276 (e.g., “If people were annoying me today, I would tell them what I think of them”) aggression
277 subscales of the Aggression Questionnaire (AQ; Buss & Perry, 1992), which
278 was modified for daily use. This measure was used because self-reported propensities
279 toward physical and verbal aggression are strongly related to behavioral aggression
280 (Giancola & Parrott, 2008). Items were summed to form a composite score, such that
281 higher numbers indicated greater levels of daily aggressive tendencies ($M=7.98$,
282 $SD=5.08$; person-level $\alpha=0.96^1$).

283 **Procedure**

284 The study was conducted with the institutional internal review board’s approval.
285 Informed consent was obtained during the initial laboratory visit and then participants
286 completed demographic questions and several surveys, including the Big Five
287 Inventory. Following the initial visit, participants were given a URL code specific to them
288 of a secure server to complete their daily logs. The participants were instructed to
289 access this page each night for the following 21 days. Email reminders were sent to

290 each participant nightly to remind them to fill out their log. Overall, 106 participants
291 completed 1617 logs out of a possible 2226 days, indicating a 73% response rate.
292 Participants averaged 15.25 days of responding (ranging from 1 to 21 days).

293 **Results**

294 **Preliminary Analysis: Degree of Dependency**

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

$$303 \frac{\sigma_{\tau}^2}{(\sigma_{\tau}^2 + \sigma_W^2)}$$

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

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

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

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

311
 312 such that y_{ij} represents daily reports of aggressive tendencies (with i reports for j
 313 individuals); β_{0j} is the mean aggressive tendencies for each person aggregated across
 314 days; e_{ij} is the day-level, within-person variance around that mean (which is our
 315 estimate of σ_w^2); γ_{00} is the grand mean for aggressive tendencies; and u_{0j} is the
 316 between-person variance around the grand mean (which is our estimate of σ_τ^2). Thus,
 317 the first equation is the day-level or within-person part of the model (level-1) and the
 318 second equation is the between-person part of the model (level-2). The multilevel model
 319 could then be reduced by replacing terms to the following form:

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

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

323

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

325

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

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

341

```
342 summary(Model.Null)
```

343

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

350 **Model Building**

351 *Level-1 predictor.* Our ICC indicates that there is a great amount of dependency
352 in our daily reports of aggressive tendencies as a function of person-level clustering.
353 Thus, use of MLM to test our main hypotheses is warranted. We predicted that days in
354 which people exhibit greater left frontal activity (vs right frontal activity) would be

355 associated with greater aggressive tendencies. We can modify our unconditional model
 356 to test these hypotheses by adding in a level-1 predictor for daily visual field bias (VFB).
 357 Because it is a level-1 predictor, we can also estimate a random error term for its slope.
 358 Level-1 coefficients with a random error term added are said to be *randomly varying*,
 359 whereas level-1 coefficients without the random error term are said to be *fixed* (which
 360 shouldn't be confused with the estimate of the fixed component itself—the average
 361 coefficient for that predictor). Nezlek (2011a) argues that all level-1 coefficients should
 362 be modeled as randomly varying, so long as those random error terms can be
 363 estimated reliably. Sometimes random error terms cannot be estimated reliably and
 364 their presence may cause failure in model convergence. In such cases, removal of a
 365 random error term may be warranted.

366 The updated level-1 equation would then be:

$$367 \quad y_{ij} = \beta_{0j} + \beta_{1j}(VFB_{ij}) + e_{ij}$$

368 and the updated level-2 equations are:

$$369 \quad \text{Intercept: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$370 \quad \text{VFB: } \beta_{1j} = \gamma_{10} + u_{1j}$$

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

$$375 \quad y_{ij} = \gamma_{00} + \gamma_{10}(VFB_{ij}) + u_{0j} + u_{1j}(VFB_{ij}) + e_{ij}$$

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

387 Level-1 predictors can be entered either *uncentered*, *grand mean-centered* (i.e.,
388 centering each score around the variable's mean for all cases in the sample), or *group*
389 *mean-centered* (i.e., centering each score around the cluster mean in which that
390 particular case occurs; referred to as *person-centering* in repeated-measures designs or
391 sometimes *centering within context*). Models with level-1 predictors entered either
392 uncentered or grand mean-centered are statistically equivalent. However, differences
393 arise with group mean-centering. To summarize the general arguments on centering in
394 the literature cited above, when level-1 predictors are entered grand mean-centered,
395 between-cluster differences actually influence level-1 parameter estimates. Thus, level-
396 1 effects are confounded by level-2 differences. For instance, in the context of the
397 present study, suppose that daily reports of visual field bias were associated with
398 aggressive tendencies. If the predictor was grand mean-centered, then we wouldn't be

399 able to separate daily fluctuations in visual field bias from general person-level
400 individual differences. Perhaps more aggressive people tend to be biased toward the
401 right visual field, in general, and less aggressive people tend to be biased toward the
402 left? In this case, the level-1 test would say little about the impact of daily fluctuations in
403 visual field bias on aggressive tendencies. For researchers interested in development or
404 within-person change, those daily fluctuations are of interest.

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

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

$$423 \quad y_{ij} = \beta_{0j} + \beta_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}$$

424 and the level-2 equations become:

$$425 \quad \text{Intercept: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\bar{X}_{VFB.j}) + u_{0j}$$

$$426 \quad \text{VFB: } \beta_{1j} = \gamma_{10} + u_{1j}$$

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

$$430 \quad y_{ij} = \gamma_{00} + \gamma_{10}(VFB_{ij} - \bar{X}_{VFB.j}) + \gamma_{01}(\bar{X}_{VFB.j}) + u_{0j} + u_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}$$

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

432

```
433 MLMLevel1model <- lmer(VerbPhysAggSum ~ 1 + GRCbias + GRPMnbias + (1
```

```
434 + GRCbias | ID), data = LineBisectionStudy)
```

435

436 We created our new object (*MLMLevel1model*) by adding two fixed components
 437 and one random error term to our unconditional model. One fixed component
 438 represents the average slope for daily visual field bias, person-centered (*GRCbias*),
 439 and, because it is a level-1 variable, we can model a random error term for it. The next
 440 fixed component represents the slope for person-level individual differences in visual
 441 field bias, on average, across the study period (*GRPMnbias*), which is a level-2

442 predictor and, thus, has no random error term (this is a two-level model, so there are no
443 level-3 units for this slope to vary across).

444 Using the same *summary()* function as before (with the new model name
445 entered), we obtain our results. Again, we predicted that days in which people exhibited
446 greater right visual field bias (as opposed to left visual field bias) would be associated
447 with greater aggressive tendencies, given the literature linking right visual field bias, left
448 frontal activity, behavioral approach, and aggression (e.g., Harmon-Jones & Peterson,
449 2008; Nash et al., 2010). Contrary to hypotheses, daily increases in *left* visual field bias
450 (an indicator of *right* frontal activity) was positively associated with aggressive
451 tendencies, $b=0.76$, $t(33.35)=2.28$, $p=0.029$.² Furthermore, people who tended to exhibit
452 greater left visual field bias, in general, over the study period also tended to report
453 greater aggressive tendencies compared to those who generally exhibited right visual
454 field bias, $b=2.18$, $t(111.48)=2.13$, $p=0.035$.

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

465 agreeableness x visual field bias interaction. The level-1 equation remains:

$$466 \quad y_{ij} = \beta_{0j} + \beta_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}$$

467 however, the level-2 equations change to:

$$468 \quad \text{Intercept: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\bar{X}_{VFB.j}) + \gamma_{02}(Gender_{.j}) + \gamma_{03}(Agreeableness_{.j}) + u_{0j}$$

$$469 \quad \text{VFB: } \beta_{1j} = \gamma_{10} + \gamma_{11}(Gender_{.j}) + \gamma_{12}(Agreeableness_{.j}) + u_{1j}$$

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

$$478 \quad y_{ij} = \gamma_{00} + \gamma_{10}(VFB_{ij} - \bar{X}_{VFB.j}) + \gamma_{01}(\bar{X}_{VFB.j}) + \gamma_{02}(Gender_{.j}) \\ + \gamma_{03}(Agreeableness_{.j}) + \gamma_{11}((VFB_{ij} - \bar{X}_{VFB.j}) * Gender_{.j}) \\ + \gamma_{12}((VFB_{ij} - \bar{X}_{VFB.j}) * Agreeableness_{.j}) + u_{0j} + u_{1j}(VFB_{ij} - \bar{X}_{VFB.j}) + e_{ij}$$

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

485 et al., 2003; Nezlek, 2011a).

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

488

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

492

493 We added fixed components for gender (dummy-coded as 0=Female and
494 1=Male) and trait agreeableness (*GMCagreeable*), as well as the relevant cross-level
495 interactions (the *lmer()* function uses “:” to indicate interactions). Recall that level-2
496 predictors do not have random error terms in a two-level model, so no adjustment is
497 needed to the random part of the code.

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

505

506

507

508 **Table 1**

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

511	512	513	514	515	516	517
Parameter	Variable	Estimate	t-value	df	p-value	
514	gamma_00	Intercept	7.45	16.52	97.90	<.001
515	gamma_10	VFB (person-centered)	1.33	3.44	30.92	0.0017
516	gamma_01	VFB (person means)	1.18	1.21	114.06	0.23
517	gamma_02	Gender	2.63	2.92	106.05	0.004
518	gamma_03	Agreeableness	-0.27	-3.98	99.53	0.0001
519	gamma_11	VFB x Gender	-1.71	-2.54	31.88	0.016
520	gamma_12	VFB x Agreeableness	-0.07	-1.18	30.70	0.25

521

522

523 To evaluate the nature of significant interactions in MLM, researchers can use
 524 procedures recommended by Aiken and West (1991) and Preacher and colleagues
 525 (2006). Computational tools for probing simple slopes for interactions in MLM can be
 526 found here: <http://quantpsy.org/interact/hlm2.htm>. Although, daily visual field bias was
 527 generally not associated with aggressive tendencies for males, $b=-0.38$, $t=-0.69$,
 528 $p=0.49$, males who exhibited right visual field bias were much more aggressive than
 529 females who exhibited similar right visual field bias, $b=3.36$, $t=3.51$, $p=0.0007$, as may
 530 be anticipated. However, for females, days with greater left field visual bias (versus right
 531 field bias) were associated with more aggressive days, $b=1.33$, $t=3.44$, $p=0.0008$.
 532 Because we used the *lmerTest* package to model our data, we can easily use the *sjPlot*
 533 package (version 2.8.4; Lüdtke, 2020) to plot our interaction and simple slopes with
 534 the *plot_model()* function, as in:

535

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

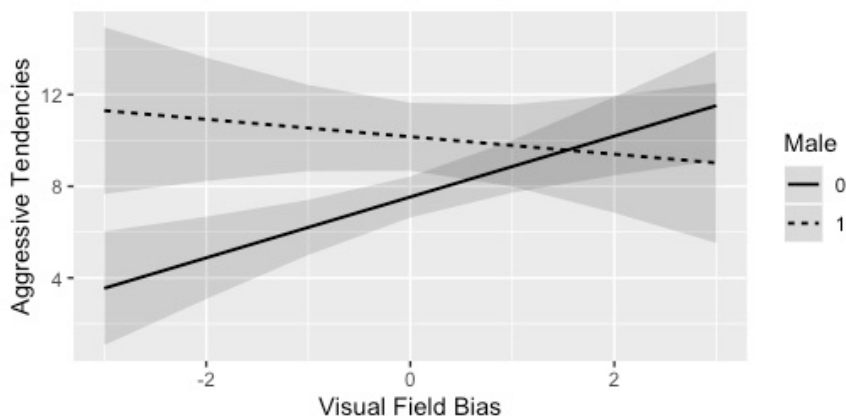
538

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

549

CONCLUSION

550 The present article described some of the basic concepts and techniques
551 involved in modeling multilevel data structures. Furthermore, this article serves as a
552 tutorial for implementing basic MLM techniques in the R package, *lmerTest* (version 3.1-
553 2; Kuznetsova et al., 2017) with real, publicly available data. The study and analyses
554 were presented within the context of intensive longitudinal repeated-measures designs.



555

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

559

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

573 The present paper was not meant to be exhaustive but rather demonstrate basic
574 analyses in MLM. Although our article presented the analysis of a two-level model with
575 interval-contingent data, the principles and techniques discussed can be generalized to
576 the analysis of event-contingent or group/organizational data sets. Furthermore, the
577 techniques discussed can be easily extended to accommodate three-level data
578 structures. There are also a number of advanced topics in MLM that go beyond the
579 scope of the present article. These include effect sizes, determining the statistical power
580 of multilevel designs, mediation analyses, nonlinear and categorical dependent
581 variables, and scale reliability. Interested readers are encouraged to read the work of
582 Kreft and de Leeuw (1998), Raudenbush and Bryk (2002), Hox, Moerbeek, and van de
583 Schoot (2017), Nezlek (2011b), Snijders and Bosker (2012), and Bauer and colleagues
584 (2006). With that stated, we hope that this article will serve as a resource for
585 researchers interested in implementing multilevel modeling in their own work.

586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605

606 **REFERENCES**

- 607 Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting*
608 *interactions*. London: Sage.
- 609 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
610 models using lme4. *Journal of Statistical Software*, 67, 1-48.
- 611 Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random
612 indirect effects and moderated mediation in multilevel models: New procedures
613 and recommendations. *Psychological Methods*, 11, 142-163.
- 614 Bettencourt, B., & Miller, N. (1996). Gender differences in aggression as a function of
615 provocation: A meta-analysis. *Psychological bulletin*, 119, 422-447.
- 616 Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived.
617 *Annual Review of Psychology*, 54, 579-616.
- 618 Bonito, J. A., Ruppel, E. K., & Keyton, J. (2012). Reliability estimates for multilevel
619 designs in group research. *Small Group Research*, 43, 443-467.
- 620 Buss, A. H., & Perry, M. P. (1992). The aggression questionnaire. *Journal of Personality*
621 *and Social Psychology*, 63, 452-459.
- 622 Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park,
623 CA: Sage.
- 624 Carver, C. S. (2004). Negative affects deriving from the behavioral approach system.
625 *Emotion*, 4, 3-22.
- 626 Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect:
627 Evidence and implications. *Psychological Bulletin*, 135, 183-204.
- 628

- 629 Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple*
630 *regression/correlation analysis for the behavioral sciences (3rd Edition)*.
631 Mahwah, NJ: Lawrence Erlbaum Associates.
- 632 Davidson, R., Ekman, P., Saron, C., Senulis, J., & Friesen, W. (1990). Approach
633 withdrawal and cerebral asymmetry: Emotional expression and brain physiology.
634 *Journal of Personality and Social Psychology, 58*, 330-341.
- 635 DeWall, C. N., Lambert, N. M., Pond, R. S. Jr., Kashdan, T. B., & Fincham, F. D. (2012).
636 A grateful heart is a nonviolent heart: Cross-sectional, experience sampling,
637 longitudinal, and experimental evidence. *Social Psychological and Personality*
638 *Science, 3*, 232-240.
- 639 Drake, R. A., & Myers, L. R. (2006). Visual attention, emotion, and action tendency:
640 Feeling active or passive. *Cognition and Emotion, 20*, 608-622.
- 641 Enders, C. K., & Tofghi, D. (2007). Centering predictor variables in cross-sectional
642 multilevel models: a new look at an old issue. *Psychological methods, 12*, 121-
643 138.
- 644 Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a
645 multilevel confirmatory factor analysis framework. *Psychological Methods, 19*,
646 72-91.
- 647 Giancola, P. R., & Parrott, D. J. (2008). Further evidence for the construct validity of
648 laboratory aggression paradigms. *Aggressive Behavior, 34*, 214-229.
- 649 Gilbert, L. R., Pond, R. S. Jr., Haak, E. A., DeWall, C. N., & Keller, P. S. (2015). Sleep
650 problems exacerbate the emotional consequences of interpersonal rejection.
651 *Journal of Social and Clinical Psychology, 34*, 50-63.

- 652 Harmon-Jones, E. (2003). Anger and the behavioral approach system. *Personality and*
653 *Individual Differences, 35*, 995-1005.
- 654 Harmon-Jones, E., & Allen, J. J. B. (1998). Anger and prefrontal brain activity: EEG
655 asymmetry consistent with approach motivation despite negative affective
656 valence. *Journal of Personality and Social Psychology, 74*, 1310-1316.
- 657 Harmon-Jones, E., Harmon-Jones, C., Abramson, L., & Peterson, C. K. (2009). PANAS
658 positive activation is associated with anger. *Emotion, 9*, 183-196.
- 659 Harmon-Jones, E., & Peterson, C. K. (2008). Effect of trait and state approach
660 motivation on aggression. *Journal of Research in Personality, 42*, 1381-1385.
- 661 Harmon-Jones, E., & Peterson, C. K. (2009). Supine body position reduces neural
662 response to anger evocation. *Psychological Science, 20*, 1209-1210.
- 663 Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear
664 models. *Journal of Management, 23*, 723-744.
- 665 Hox, J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and*
666 *applications* (3rd ed.). New York, NY: Routledge.
- 667 Hox, J. J., & Kleiboeer, A. M. (2007). Retrospective questions or a diary method? A two
668 level multitrait-multimethod analysis. *Structural Equation Modeling: A*
669 *Multidisciplinary Journal, 14*, 311-325.
- 670 Infante, M. A., Nguyen-Louie, T. T., Worley, M., Courtney, K. E., Coronado, C., &
671 Jacobus, J. (2020). Neuropsychological trajectories associated with adolescent
672 alcohol and cannabis use: A prospective 14-year study. *Journal of the*
673 *International Neuropsychological Society, 26*, 480-491.

- 674 John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory--Versions*
675 *4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of
676 Personality and Social Research.
- 677 John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big
678 Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John,
679 R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and*
680 *research* (pp. 114-158). New York, NY: Guilford Press.
- 681 Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic Data Analysis*. New
682 York, NY: Guildford Press.
- 683 Kreft, I. G., & de Leeuw, J. (1994). The gender gap in earnings: A two-way nested
684 multiple regression analysis with random effects. *Sociological Methods &*
685 *Research*, 22, 319-341.
- 686 Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park,
687 CA: Sage.
- 688 Kreft, I. G., de Leeuw, J., & Aiken, L. (1995). The effects of different forms of centering
689 on hierarchical linear models. *Multivariate Behavioral Research*, 30, 1-22.
- 690 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). ImerTest Package:
691 Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1-26.
- 692 Lai, M. H. C. (in press). Composite reliability of multilevel data: It's about observed
693 scores and construct meanings. *Psychological Methods*,
- 694 Lüdtke, D. (2020). *sjPlot: Data Visualization for Statistics in Social Science*. R
695 package version 2.8.4, URL: <https://CRAN.R-project.org/package=sjPlot>.
- 696

- 697 Meier, B. P., & Robinson, M. D. (2004). Does quick to blame mean quick to anger?: The
698 role of agreeableness in dissociating blame and anger. *Personality and Social*
699 *Psychology Bulletin, 30*, 856-867.
- 700 Meier, B. P., Robinson, M. D., & Wilkowski, B. M. (2006). Turning the other cheek:
701 Agreeableness and the regulation of aggression-related primes. *Psychological*
702 *Science, 17*, 136-142.
- 703 Milner, A. D., Brechmann, M., & Pagliarini, L. (1992). To halve and to halve not: An
704 analysis of line bisection in normal subjects. *Neuropsychologia, 30*, 515-526.
- 705 Nash, K., McGregor, I., & Inzlicht, M. (2010). Line bisection as a neural marker of
706 approach motivation. *Psychophysiology, 47*, 979-983.
- 707 Nezlek, J. B. (1995). Social construction, gender/sex similarity, and social interaction in
708 close personal relationships. *Journal of Social and Personal Relationships, 12*,
709 503-520.
- 710 Nezlek, J. B. (2001). Multilevel random coefficient analyses of event and interval
711 contingent data in social and personality psychology research. *Personality and*
712 *Social Psychology Bulletin, 27*, 771-785.
- 713 Nezlek, J. B. (2007). Multilevel modeling in research on personality. In R. Robins, R. C.
714 Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality*
715 *psychology* (pp. 502-523). New York: Guilford.
- 716 Nezlek, J. B. (2011a). Multilevel modeling analyses of diary-style data. In M. R. Mehl &
717 T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp.
718 357-383). New York: Guilford.
- 719

- 720 Nezlek, J. B. (2011b). *Multilevel modeling for social and personality psychology*.
721 Thousand Oaks, CA: Sage.
- 722 Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within
723 person variability. *Journal of Research in Personality*, 69, 149-155.
- 724 Nezlek, J. B., Hampton, C., & Shean, G. D. (2000). Clinical depression and everyday
725 social interaction in a community sample. *Journal of Abnormal Psychology*, 107,
726 11-19.
- 727 Ode, S., Robinson, M. D., & Wilkowski, B. M. (2008). Can one's temper be cooled?: A
728 role for agreeableness in moderating neuroticism's influence on anger and
729 aggression. *Journal of research in personality*, 42, 295-311.
- 730 Pond, R. S. Jr., DeWall, C. N., Lambert, N. M., Deckman, T., Bonser, I. M., & Fincham,
731 F. D. (2012). Repulsed by violence: Disgust sensitivity buffers trait, behavioral,
732 and daily aggression. *Journal of Personality and Social Psychology*, 102, 175-
733 188.
- 734 Pond, R. S. Jr., Kashdan, T. B., DeWall, C. N., Savostyanova, A., Lambert, N. M., &
735 Fincham, F. D. (2012). Emotion differentiation moderates aggressive tendencies
736 in angry people: A daily diary analysis. *Emotion*, 12, 326-337.
- 737 Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing
738 interactions in multiple linear regression, multilevel modeling, and latent curve
739 analysis. *Journal of Educational and Behavioral Statistics*, 31, 437-448.
- 740 Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects.
741 *Sociology of Education*, 1-17.
742

- 743 Raudenbush, S. W., & Bryk, A. S. (1988). Chapter 10: Methodological advances in
744 analyzing the effects of schools and classrooms on student learning. *Review of*
745 *Research in Education*, 15, 423-475.
- 746 Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and*
747 *data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- 748 Séguin, M., Dégeilh, F., Bernier, A., El-Jalbout, R., & Beauchamp, M. H. (2020). It's a
749 matter of surgency: Traumatic brain injury is associated with changes in
750 preschoolers' temperament. *Neuropsychology*, 34(4), 375-387.
- 751 Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater
752 reliability. *Psychological bulletin*, 86, 420-428.
- 753 Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical
754 models, and individual growth models. *Journal of Educational and Behavioral*
755 *Statistics*, 23, 323-355.
- 756 Smits, D. J. M., & Kuppens, P. (2005). The relations between anger, coping with anger,
757 and aggression, and the BIS/BAS system. *Personality and Individual Differences*,
758 39, 783-793.
- 759 Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and*
760 *advanced multilevel modeling* (2nd ed.). London: Sage.
- 761 Sullivan, E. V., Brumback, T., Tapert, S. F., Brown, S. A., Baker, F. C., Colrain, I. M.,
762 Prouty, D., De Bellis, M. D., Clark, D. B., Nagel, B. J., Pohl, K. M., & Pfefferbaum,
763 A. (2019). Disturbed cerebellar growth trajectories in adolescents who initiate
764 alcohol drinking. *Biological Psychiatry*, 87, 632-644.
- 765

- 766 Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday events: Origins, types, and
767 uses. *Journal of Personality, 59*, 339-354.
- 768 West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data:
769 Daily experiences. *Journal of Personality, 59*, 609-662.
- 770 Weston, S. J., & Yee, D. (2017). Why you should become a UseR: A brief introduction
771 to R. *APS Observer, 30*.
- 772 Wilkowski, B. M., & Meier, B. P. (2010). Bring it on: Angry facial expressions potentiate
773 approach-motivated motor behavior. *Journal of Personality and Social
774 Psychology, 98*, 201-210.
- 775 Wilkowski, B. M., Robinson, M. D., & Meier, B. P. (2006). Agreeableness and the
776 prolonged spatial processing of antisocial and prosocial information. *Journal of
777 Research in Personality, 40*, 1152-1168.
- 778
- 779
- 780
- 781
- 782
- 783
- 784
- 785
- 786
- 787
- 788

789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811

FOOTNOTES

¹Although 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 & Kleiboyer, 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 Satterthwaite approximation for obtaining degrees of freedom, which is a weighted average of the between and within degrees of freedom. Therefore, the Satterthwaite approximation results in fractional degrees of freedom. Kenny, Kashy, and Cook (2006) recommend the Satterthwaite estimate of degrees of freedom because it takes into account the mixture of between and within parts of the estimate.

812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844

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.

845 Declarations

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

849 **Conflicts of interests/Competing interests:** Not applicable

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

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

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

859
 860
 861
 862