Journal of Pediatric Psychology, 2024, **00**, 1–24
https://doi.org/10.1093/jpepsy/jsae070
Original Research Article

Handling missing data in longitudinal clinical trials: three examples from the pediatric psychology literature

James Peugh [1,2,*] **and Constance Mara** [1,2]

[1]Behavioral Medicine Clinical Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[2]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, United States

*Corresponding author: Behavioral Medicine Clinical Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, United States. Email: James.Peugh@cchmc.org

Abstract

Researchers by default tend to choose complex models when analyzing nonindependent response variable data, this may be particularly applicable in the analysis of longitudinal trial data, possibly due to the ability of such models to easily address missing data by default. Both maximum-likelihood (ML) estimation and multiple imputation (MI) are well-known to be acceptable methods for handling missing data, but much of the recently published quantitative literature has addressed questions regarding the research designs and circumstances under which one should be chosen over the other. The purpose of this article is threefold. First, to clearly define the assumptions underlying three common longitudinal trial data analysis models for continuous dependent variable data: repeated measures analysis of covariance (RM-ANCOVA), generalized estimating equation (GEE), and a longitudinal linear mixed model (LLMM). Second, to clarify when ML or MI should be chosen, and to introduce researchers to an easy-to-use, empirically well-validated, and freely available missing data multiple imputation program: *BLIMP*. Third, to show how missing longitudinal trial data can be handled in the three data analysis models using three popular statistical analysis software packages (*SPSS*, *Stata*, and *R*) while keeping the published quantitative research in mind.

Keywords: clinical trial, longitudinal research, statistical approach, randomized controlled trial.

Options for handling missing responses and analyzing longitudinal trial data

> For scientific discoveries to be valid—whether in theory or empirically—a phenomenon must be accurately described: The scientist must use appropriate counterfactuals and eliminate competing explanations. Empirical work must also use an appropriate design and method, and empirical claims made about the phenomenon must be correctly characterized. (Wulff et al., 2023, p. 1)

Researchers ready to analyze a sample of longitudinal trial data have a variety of options to choose from, but each option has different assumptions that must be met to avoid inefficient or inappropriate analyses, biased treatment effect results, and incorrect conclusions (Locascio & Atri, 2011). However, correctly choosing and correctly implementing the longitudinal analysis method needed to answer the research question can often present challenges. First, research has shown that investigators tend to use more complicated longitudinal data analytic techniques than are necessary because such techniques are popular and widely taught, but often at the expense of less familiar, rarely taught, but potentially more parsimonious and effective alternatives (Bauer & Sterba, 2011; McNeish et al., 2017). Second, prior to analysis, longitudinal trial researchers can expect to be confronted with missing data, which must be handled correctly to obtain unbiased treatment effect estimates (e.g., Gomer & Yuan, 2021).

No missing data handling technique is completely foolproof, but maximum-likelihood (ML) estimation and multiple imputation (MI) have been well-known for two or more decades to be acceptable missing data handling practices because both have been shown to best minimize parameter estimate bias due to missing data (e.g., Carpenter et al. 2013; Enders, 2010, 2022; Graham, 2012; Larsen, 2011; Little & Rubin, 2002). One possible reason for the popularity of more complex longitudinal data analytic techniques, as stated above, is that within some statistical analysis software programs (e.g., *Mplus*, *Stata*, *R*, but not all, e.g., *SPSS*, *SAS*), missing data can be handled by the default maximum-likelihood (ML) estimation algorithm. However, many recent quantitative research studies have addressed questions regarding whether both ML and MI are equally suited to handle missing data and minimize parameter estimate bias across all design and analysis scenarios, especially for complex analysis models often used to answer longitudinal clinical trial research questions (Enders, 2022, 2023a, 2023b; Enders et al., 2016, 2018, 2020; Goldstein et al., 2014; Grund et al., 2018, 2019; 2021a,b; Keller & Enders, 2023). Further complicating the issue is the fact that, for many popular statistical analysis software programs (e.g., *SPSS, SAS*), their MI algorithms are not able to correctly impute missing values from correlated responses collected longitudinally.

This article can be considered an updated extension of a previous *Journal of Pediatric Psychology* publication on the topic (Little et al. 2014), and the goals of this article are threefold: (1) to clearly describe the assumptions underlying three possible longitudinal trial data analysis models: a repeated measures analysis of covariance (RM-ANCOVA), a generalized estimating equation (GEE), and a longitudinal linear mixed model (LLMM), (2) to show how missing longitudinal trial data can be handled in three popular statistical

Received: December 7, 2023. **Revised:** August 12, 2024. **Accepted:** August 12, 2024

analysis software packages (*SPSS*, *Stata*, and *R*) based on the relevant published quantitative literature (Garcia & Marder, 2017), and (3) to introduce readers to an empirically well-validated standalone Bayesian multiple imputation internet freeware software package (*BLIMP*[1]; Enders et al. 2018, 2020; Keller & Enders, 2022) if multiple imputation for missing data handling is needed. ML estimation as a missing data handling technique needs little, if any, explanation. Contingent on verifying that the ML estimation algorithm for a given statistical analysis software package is capable of producing valid parameter estimates from incomplete data, handling missing data with ML estimation simply involves fitting the analysis model to the sample data (e.g., Enders et al., 2020; but see also auxiliary correlate variable identification in Enders, 2022). Much more attention is given in this article to MI missing data handling in general and the MI possibilities offered in *BLIMP* specifically. For readers unfamiliar with *BLIMP*, Table 1 shows a list of basic commands along with options and specifications. This is not an exhaustive list; *BLIMP* is capable of much more functionality and interested readers can find additional information in the User's Guide (Keller & Enders, 2023).

Before proceeding further, several qualifying caveats warrant mention: (1) This article proceeds from a missing at random (MAR) assumption for all analyses presented. Additional missing data mechanisms (MCAR or MNAR) are not discussed here, but we refer the readers to extensive treatments of these issues in longitudinal trial data published elsewhere (e.g., Ben et al., 2023; Carpenter & Kenward, 2007; DeSouza et al., 2009; Fiero et al., 2017; Peugh et al., 2023), (2) This article proceeds from a longitudinal mixed linear, not longitudinal structural equation, modeling approach. Readers interested in either imputing missing values for (see Keller & Enders, 2022, pp. 171–173) or analyzing (e.g., Grimm et al., 2017; Little, 2013) longitudinal trial data with structural equation models can consult several published sources. (3) The example analyses shown here assume dependent variable data measured on a continuous scale. *BLIMP* can impute several types of missing data, including binary, nominal, ordinal, and count (but only if the count variable with missing data is the dependent variable currently) variable measurement scales. Censored continuous data is also common in pediatric research but is not addressed here. *BLIMP* is currently unable to impute missing censored data (but that ability may be available soon). (4) Assuming model-based (discussed below) multiple imputations, it is notable to mention that *BLIMP* is capable of imputing, analyzing, and outputting Bayesian parameter estimates. (5) Although multiply imputed data was generated using *BLIMP* and output for analysis using *SPSS*, *Stata*, and *R*, it is also important to note that *BLIMP* can output imputed data for use in many other statistical analysis software packages (e.g., *SAS* and *Mplus*; assuming those packages can perform the analyses shown here). Finally, it is important to also note that missing data in the examples shown here is addressed at the total score or scale score level, but published tutorials for addressing item-level missing data with multiple imputations are available (see Alacam et al., 2023).

[1] Available at: Blimp (appliedmissingdata.com).

Handling missing data with multiple imputation

Researchers interested in more information regarding multiple imputations can consult several published sources (Carpenter et al. 2023; Enders, 2022; Kleinke et al. 2020; van Buuren, 2021). To aid in the example analyses presented here, we offer a brief overview of the two general types of imputation models available and the key steps involved in conducting multiple imputations. The two types of multiple imputation models available are described first.

Alternative hypothesis model (H_A:) and null hypothesis (H_0:) imputation models

There are two types (e.g., see Muthén & Muthén, 1998–2017, p. 576) of multiple imputation models: an alternative hypothesis imputation model (H_A:, or fully conditional specification [FCS], e.g., Enders et al., 2018; van Buuren et al., 2006; van Buuren, 2021) and a null hypothesis imputation model (H_0:, or model-based, e.g., Enders et al., 2020). To assist in describing both, let's assume three hypothetical continuous variables ($M_1 - M_3$) that each show varying amounts of missing data (M).

Alternative hypothesis model (H_A:) imputation: multivariate or univariate

Alternative hypothesis (H_A:) model imputation can be performed one of two ways: multivariate (or joint imputation; see Quartagno & Carpenter, 2019; Schafer, 1997, 1999, 2003; Schafer & Olsen, 1998) or univariate (factored regression or sequential specification; Lüdtke et al., 2020). Multivariate alternative hypothesis model imputation involves the use of all information available from all analysis variables to draw imputed values for missing data. For the hypothetical three variable ($M_1 - M_3$) example, all available information constitutes the means ($\bar{M}_1$–$\bar{M}3$), variances ($\sigma^2_{M_1}$–$\sigma^2_{M_3}$), and covariances (σ_{M_2,M_1}, σ_{M_3,M_1}, σ_{M_3,M_2}) (e.g., Enders, 2022, pp. 264–265) as shown:

$$\text{Means: } \begin{bmatrix} \bar{M}_1 & \bar{M}_2 & \bar{M}_3 \end{bmatrix}$$

$$\text{Covariance matrix: } \begin{bmatrix} \sigma^2_{M_1} & \sigma_{M_1,M_2} & \sigma_{M_1,M_3} \\ \sigma_{M_2,M_1} & \sigma^2_{M_2} & \sigma_{M_2,M_3} \\ \sigma_{M_3,M_1} & \sigma_{M_3,M_2} & \sigma^2_{M_3} \end{bmatrix}$$

Multivariate alternative hypothesis model imputation is presented here only as an illustrative segue into univariate alternative hypothesis model imputation and is not considered further.[2]

[2] Multivariate or joint alternative model imputation is still available in many commercially available statistical analysis software packages (e.g., in *R* if the *Jomo* package used, or in *Mplus*, if the 'TYPE=BASIC;' analysis specification is used with the 'DATA IMPUTATION:' command, for example). Multivariate or joint alternative hypothesis model imputation has been shown in some instances to produce inaccurate results (Bartlett et al., 2015; Liu et al., 2016, as cited in Keller & Enders, 2022, pp. 14–15). Observing inappropriate imputed values (e.g., imputed values other than 0 or 1 for a binary variable, or unrealistic imputed values for a continuous variable; see Enders, 2022, p. 272) are a common indication of a multivariate distribution being inappropriate for a given imputation situation.

Table 1. The basics of BLIMP syntax (from Keller & Enders, 2023)

Command	Options	Specifications
DATA:	Requires: (1) file path and (2) file type	Accepts: .csv, .dat, or .txt files e.g., C : \data.csv;
VARIABLES:		Listed in the order they appear in the data file
ORDINAL:	Categorical variables	Ideal if the binary variables listed here
NOMINAL:	Categorical variables	Ideal if ordered categorical variables listed here
COUNT:	Count variables	Count variables listed here
FIXED:		Variables with no missing values are listed here
CLUSTERID:	Clustering variable	For nested cross-sectional or longitudinal imputation
CENTER:	'groupmean' or 'grandmean'; see Enders and Tofighi, 2007	If 'CLUSTERID:' is specified, both can be used
MISSING:	One value allowed	e.g., Missing: -99; or Missing = NA;
FCS: OR MODEL: (never both)		Use FCS: for alternative hypothesis model (H_A:) imputation Use MODEL: for null hypothesis model (H_0:) imputation
SEED:	Must be less than 10 digits	Starts the random number generator needed for imputation
NIMPS:	e.g., see Graham et al. (2007)	Request the number of imputed datasets
BURN:	BURN: 5000 is recommended	Number of iterations before the first imputed dataset is saved
ITERATIONS:	ITERATIONS: 10 000 is recommended	Number of iterations between successive imputed datasets
CHAINS:	PROCESSORS, e.g., CHAINS 10 processors 4	Number of algorithms used to impute missing data
OPTIONS:	estimates, latent, manifest, or PSR	Potential scale reduction (PSR) requested here
SAVE:	For analysis in *SPSS*, *SAS*, or *R* For analysis in *Stata* For analysis in *Mplus*	e.g., SAVE: stacked = C : \imps.dat. e.g., SAVE: stacked0 = C : \imps.dat. e.g., SAVE: separate = C : \imps*.dat.

Notes: (1) If variable names appear as column headers in the DATA: file, the VARIABLES: command has to be omitted.
(2) A DATA: command file path is not needed if the data file and BLIMP input file are in the same location.
(3) If the number of CHAINS: and the number of NIMPS: are equal, additional commands (THIN:) are not needed.

Univariate alternative hypothesis model imputation (e.g., Bartlett et al., 2015) assumes the benefits of multivariate imputation can still be had in a series of more flexible univariate models. An example of a univariate alternative hypothesis imputation model would be an "all-in-turn" sequence of prediction models in which each analysis variable predicts, and is predicted by, all other analysis variables "in turn" for missing data imputation purposes as shown below (e.g., Enders, 2022, pp. 272–273; Keller & Enders, 2022, pp. 74–77):

$$\widehat{M}_1 = \beta_1 + \beta_{11}M_2 + \beta_{12}M_3 + \varepsilon_{1i}$$

$$\widehat{M}_2 = \beta_2 + \beta_{21}M_1 + \beta_{22}M_3 + \varepsilon_{2i}$$

$$\widehat{M}_3 = \beta_3 + \beta_{31}M_1 + \beta_{32}M_2 + \varepsilon_{3i}$$

It is important to note that univariate alternative hypothesis imputation is the default in *BLIMP*, and both the multivariate and univariate alternative hypothesis imputation models shown above are saturated models with zero degrees of freedom.

Null hypothesis model (H_0:) imputation: unfactored or factored predictors

Null hypothesis (H_0:) or model-based multiple imputation is just that: the model used to impute missing data is an exact match to the data analysis model needed to answer the research question. If the previous hypothetical three variables are predictors of a continuous outcome (O) in a multiple regression, imputation would be based on the following analysis model:

$$\widehat{O}_i = \beta_0 + \beta_1 M_{1i} + \beta_2 M_{2i} + \beta_3 M_{3i} + \varepsilon_i$$

In *BLIMP*, additional specification considerations are given in model-based imputation scenarios involving predictor variable (or, "x-side") missing data. By default, *BLIMP* will treat predictor variables with missing data as unfactored (Keller & Enders, 2022, pp. 14–21), meaning missing values for the three predictor variables will be imputed using both the outcome (O) prediction model above and the three "all-in-turn" univariate alternative hypothesis prediction imputation models shown previously. A second factored regression or sequential specification for predictor variable missing data (Keller, 2021; Keller & Enders, 2023) is also available in *BLIMP*. To illustrate, assume the imputation model shown above still involves the prediction of a continuous outcome (O), but is now predicted by: (1) B; a binary predictor with missing values, (2) Ct; a count predictor with missing values, and (3) $\vec{X}$; a significantly skewed continuous predictor with missing data (more on this in the last analysis example below). A factored regression or sequential specification approach would impute missing values based both the analysis model

$$\widehat{O}_i = \beta_0 + \beta_1 B_i + \beta_2 Ct_i + \beta_3 \vec{X}_i + \varepsilon_{1i},$$

as well as a factored regression or sequential specification as shown:

$$\widehat{B}_i = \gamma_{20} + \varepsilon_{Xi}.$$

$$\widehat{Ct}_i = \gamma_{10} + \gamma_{11} X_i + \varepsilon_{Cti}$$

$$\vec{X} = \gamma_{00} + \gamma_{01} Ct_i + \gamma_{02} X_i + \varepsilon_{Bi}$$

If factored regression or sequential specification is used to impute missing data in predictor variables, categorical variables with missing data are listed first, continuous variables with missing data are listed last (Lüdtke et al., 2020). Both unfactored and factored regression or sequential specification are mathematically equal in most situations (Keller & Enders, 2022, pp. 44–48). Model-based imputation scenarios are not saturated models but have positive degrees of freedom, the significance of which is elaborated upon (see Graham's [2012] "golden rule" of imputation) in the section below.

At this point, researchers could be asking two important clarification questions. First, how to know whether null hypothesis model (H_0: or model-based) or alternative hypothesis model (H_A: or FCS) imputation is needed? If the answer to a longitudinal trial research question involves a nonlinear estimate, such as a polynomial ($X^2 \rightarrow Y$) effect, a moderation ($X * Z \rightarrow Y$) effect, or a random effect (random slope; e.g., the $X \rightarrow Y$ effect shows non-zero variance across i individual participants, or c clustering units, such as schools or hospitals), null model (or model-based) imputation is needed to avoid biased conclusions (Enders et al., 2018, 2020; Keller & Enders, 2022, pp. 44–46). Second, if null model (or model-based) imputation is needed, how to know if unfactored or factored regression/sequential specification is needed? A factored regression or sequential specification is needed if predictors with missing data are nonlinear (e.g., X^2), count variables, or non-normally distributed continuous variables (Du et al., 2022; Keller & Enders, 2022, pp. 18, 44–47).

To summarize, multiple imputation should always be guided by the "golden rule" put forth by Graham (2012, p. 62, *italics added*): "The imputation model must be *at least* as complex as the analysis model." As such, and before discussing the key steps involved in imputing missing data, three "take-home" tips for researchers can be offered:

1) When quantitative research supports its use as a missing data handling approach (described below and in Appendix A), ML should be used based on parsimony. However, if MI is needed:
2) Researchers cannot go wrong using alternative hypothesis (H_A:) model imputation for item-level missing data (but see also Alacam et al., 2023).
3) Researchers cannot go wrong using null (H_0:) model imputation assuming the model used to impute missing values and data analysis model needed to answer the research question are an exact match (Enders, 2023b).

Key steps in multiple imputation

Multiple imputation is a repetitive mathematical process by which possible missing data values are computed, parameter estimates updated, and possible missing data values recomputed until the precision of imputed data values cannot be improved. This repetitive or iterative process is conducted by multiple mathematical computational algorithms, or chains. Chains are said to have converged when the precision of imputed values cannot be improved. The potential scale reduction (PSR) factor indicates when multiple imputation convergence is achieved (shown below). Specifically, PSR values less than 1.05 for imputing continuous variables, or less than 1.10 for imputing binary or multinomial variables (Enders, 2022; Muthén, 2010) indicate convergence. Importantly, some multiple imputation programs (e.g., *Mplus*) will not conduct analyses or output imputed datasets until convergence is achieved by default; other programs (e.g., *BLIMP*) require the researcher to examine the PSR values to confirm convergence. As such, a fourth "take-home" tip for researchers can be offered:

4) If MI is needed for missing data handling, researchers should always first check PSR values and verify that convergence has been achieved before imputing missing values, especially as missing data rates increase for categorical variables.

The analysis and pooling phases of multiple imputations first involve conducting the needed data analysis on all imputed datasets, then pooling the parameter estimates from all imputations into a single result to answer the research question. Again, it is important to note that some statistical analysis software programs (e.g., *Mplus*; assuming "Type=imputation;" is specified) analyze and pool by default in a single step. Other programs require additional syntax commands either prior to (e.g., *SPSS*) or following (e.g., *SAS*, *R*) the analysis phase to obtain pooled estimates.

Three examples of handling missing responses in longitudinal trial data will be presented below via modified reanalyses of previously published datasets: (1) a repeated measures analysis of covariance (RM-ANCOVA), (2) a generalized estimating equation (GEE), and (3) a longitudinal linear mixed model (LLMM).[3] Within each of the three analysis examples, a description of the example research study, the assumptions underlying the needed data analysis model, and data re-analysis results and discussion will all be offered. At the conclusion of each of the three analysis examples, a "Statistical considerations" section is also included that both summarizes the relevant missing data literature concerning the data analysis model and provides practical guidance for proper missing data handling. The needed missing data handling syntax scripts for the three analysis examples are provided for *SPSS*, *Stata*, and *R*.

[3] A "long" or "stacked" database, rather than a "wide" or "multivariate" database, is required to conduct all three (RM-ANCOVA, GEE, and longitudinal mixed linear modeling) analyses using mixed linear modeling. Specifically, RM-ANCOVA must be conducted using mixed linear modeling following imputation because most (if not all) statistical analysis software packages cannot pool general linear model analytic results obtained from imputed data analyses (see van Ginkel & Kroonenberg, 2014). Examples of how to create a "long" or "stacked" database can be found in Singer and Willett (2003, pp. 16–25) and Peugh & Enders (2005, pp. 718–720).

RM-ANCOVA example: Zeidan et al. (2015)[4]

A complete description of the Zeidan et al. trial (2015) is available from the source publication but is summarized briefly here. The authors applied an intense heat stimulus to ($N = 80$) participants in three distinct research sessions: baseline, pretest, and posttest, then examined the effects of unpleasantness perceptions rated on a 15-cm 0–10 visual analog scale (VAS; Price et al., 1994). Between the baseline and pretest sessions, participants were randomly assigned to four independent variable (IV) conditions and asked to practice those conditions for 4 days: (1) mindful meditation (value = 1), (2) "sham" meditation (value = 2), (3) book listening (value = 3), or (4) placebo anesthetic (petroleum jelly) (value = 4; reference group). Between the pretest and posttest sessions, participants were asked to again practice their IV condition during a 10-min break. The research question under investigation was: are there significant pretest/posttest VAS score differences across the IV groups after controlling for baseline VAS scores, biological sex, and age?

RM-ANCOVA is a well-known fixed effects model that can accommodate correlated repeated dependent variable measurements collected over time (Edwards, 2000; Girden, 1992; Grady & Helms, 1995; Hedeker & Gibbons, 2006; Kleinbaum et al., 1998; Myers, 1979), but also has several strict assumptions that must be met to avoid inferential (Type-I or Type-II) errors (Muth et al., 2016). First, RM-ANCOVA assumes dependent variable data are normally distributed, which is often unlikely in practice (e.g., Micceri, 1989). Second, RM-ANCOVA assumes a temporally structured and balanced data collection schedule (Edwards, 2000; Peugh & Heck, 2017), defined as all participants providing all necessary repeated dependent variable measurements exactly on the timetable dictated by the longitudinal design (Helms, 1992). Longitudinal data collection efforts often do not proceed on schedule, which creates an irregularly timed, unstructured, and unbalanced dataset that violates RM-ANCOVA assumptions (de Melo et al., 2022; Helms, 1992; Krueger & Tian, 2004).

Third, two of the most well-known RM-ANCOVA assumptions are homogeneity of all possible pairwise dependent variable correlations (or covariances, regardless of separation distance in time) and homogeneity of all dependent variable variances across independent variable (IV) conditions (referred to as homoscedasticity or sphericity; Keselman et al., 2001). Homogeneous dependent variable correlations/covariances and homogeneous dependent variable variances are together referred to as compound symmetry (de Melo et al., 2022; Everitt, 1998; Howell, 2007; Keselman et al., 2001; Locascio & Atri, 2011). Compound symmetry is widely viewed as unrealistic (Howell, 2007; Locascio & Atri, 2011) and correction factors aimed at reducing inferential errors when compound symmetry is violated (e.g., Greenhouse-Geisser, Hyun-Feldt, lower bound, etc.) are crude corrections at best (Rubin et al., 2007). Finally, RM-ANCOVA requires complete data and will assume missing completely at random (MCAR) for any missing data (de Melo et al., 2022; Garcia & Marder, 2017; Muth et al., 2016). In the unlikely event that MCAR is plausible, listwise deletion cannot be considered an acceptable missing data handling practice in most cases due to the resulting bias in estimates due to loss of information and a decrease in statistical power (e.g., see Abraham & Russell, 2008; Enders, 2022). However, if assumptions are met and missing data is handled via multiple imputation, RM-ANCOVA is an ideal analytic choice if pretest-posttest differences are of interest (Locascio & Atri, 2011; Omar et al., 1999).

Results from both ML and MI missing data handling showed a significant between subjects (IV group) by within-subjects (pretest-posttest) interaction (-0.67, $p < 0.001$). Post-hoc follow-up analyses, at posttest, showed both the meditation ($\bar{X}_{\text{Difference}} = -2.70$, $t_{76} = -5.83$; $p < .001$) and "sham" meditation ($\bar{X}_{\text{Difference}} = -1.83$, $t_{76} = -4.02$; $p < .001$) had significantly lower VAS scores compared to the book listening and placebo control conditions. Additional post-hoc pairwise comparisons showed, at posttest, the mindfulness meditation condition had significantly lower VAS unpleasantness scores than both the book listening ($\bar{X}_{\text{Difference}} = -1.44$, $t_{38} = 2.35$, $p < .05$, $d = .74$) and placebo control ($\bar{X}_{\text{Difference}} = -2.34$, $t_{38} = 3.23$, $p < .01$, $d = 1.02$). However, post-hoc pairwise comparisons also showed, at posttest, the "sham" meditation condition also had significantly lower VAS unpleasantness scores than both the book listening ($\bar{X}_{\text{Difference}} = -1.36$, $t_{38} = 2.47$, $p < .05$, $d = .78$) and placebo control ($\bar{X}_{\text{Difference}} = -2.25$, $t_{38} = 3.37$, $p < .01$, $d = 1.06$). There were no significant differences between the mindfulness meditation and "sham" meditation conditions at the posttest.[5] A summary of RM-ANCOVA analysis results is shown in Figure 2. *SPSS*, *Stata*, and *R* data analysis syntax scripts are also shown in Appendix A.

Statistical considerations: RM-ANCOVA

In the Zeidan et al. (2015) example, and although not readily apparent, all the predictors of VAS response variable scores (the IV group indicator, the pretest/posttest indicator, and the baseline VAS, biological sex, and age control covariates) have complete data. Missing data is restricted to the pretest/posttest VAS response variable scores. Published research (Little, 1992; Little & Rubin, 2002; von Hippel, 2007 as cited in Enders et al., 2020; Grund et al., 2018; van Buuren, 2021) has shown ML is an acceptable missing data handling technique if missing data is restricted to the response variable only (see also White & Carlin, 2010). RM-ANCOVA analysis syntax scripts for handling missing data with ML estimation are offered for *SPSS*, *Stata*, and *R* in Appendix A. However, *BLIMP* imputation syntax and RM-ANCOVA analysis scripts for analyzing imputed data are also offered in Appendix A for *SPSS*, *Stata*, and *R* for researchers facing the more realistic RM-ANCOVA scenario of missing data on more variables than just the response.

GEE example: Epstein et al. (2022)[6]

A complete discussion of the Epstein et al. (2022) trial design specifics can be found in the original publication but is briefly summarized here. The authors tested whether Enhanced Focused Concentration and Attention Learning (FOCAL+; treatment) decreased long off-roadway glances and improved driving performance versus a modified version of driver

[4] This work was supported by the National Center for Complementary and Integrative Health (Grants R21-AT007247, F32-AT006949, K99-AT008238), the National Institutes of Health–National Institute of Neurological Disorders and Stroke (Grant NS239426).

[5] Additional sensitivity analyses showed RM-ANCOVA results did not change whether a compound symmetry (highly restrictive) or unstructured (least restrictive) covariance matrix was specified.

[6] Funded by the National Institutes of Health; ClinicalTrials.gov number, NCT02848092.

BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

NOTE: Split chain PSR is being used. This splits each chain's iterations to create twice as many chains.

Comparing iterations across 2 chains	Highest PSR	Parameter #
51 to 100	1.055	1
101 to 200	1.047	2
151 to 300	1.022	2
201 to 400	1.014	1
251 to 500	1.016	1
301 to 600	1.006	15
351 to 700	1.002	1
401 to 800	1.003	1
451 to 900	1.009	7
501 to 1000	1.008	1

Figure 1. Potential scale reduction (PSR) diagnostic output for the Zeidan et al. (2015) trial data if imputation was needed for an RM-ANCOVA model.

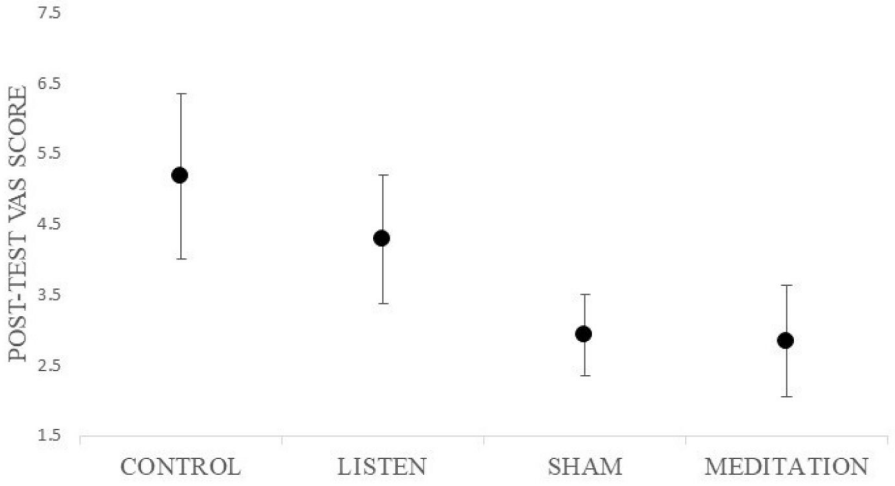

Figure 2. 95% CI error bar results at post-test for the Zeidan et al. (2015) trial data.

education and training (ADTSEA; control) in a driving simulator task. Participants diagnosed with ADHD ($N = 152$) completed three, 30-min driving simulator tasks at baseline, 1-month posttreatment, and 6-months posttreatment. During the driving simulation, participants were required to complete a distraction task that forced them to look at a location on the screen other than the forward roadway. All participants wore special eye gaze tracking goggles to capture the number and duration of off-road glances; long glances were defined as looking away from the forward roadway for more than 2 s.

Between the baseline and 1-month simulated drives, all participants were randomly assigned to treatment or control and received five additional 90-min driver training sessions presented in two distinct stages. These training sessions differed in that participants randomly assigned to treatment were trained on simulated drives until they had no long off-road glances and provided 50% correct answers to the distraction task. Control participants viewed driver's education slides and videos that contained comprehension tests at the end of each. All participants performed additional driving simulator training drives wearing eye-tracking goggles, but only treatment participants heard an auditory alarm if any off-road glance exceeded 2 s, control participants did not hear an alarm.

At 1-month and 6-month intervals following the training sessions, participants returned to complete the same driving simulation assessments that they performed at baseline, except no alarms were sounded for excessive off-road gazes. The response variable (DV) of interest was the participants' standard deviation of lane position (SDLP), defined as the average distance in feet of deviation from the center of the street assessed every 17 ms. The research questions of interest here for the Epstein et al. (2022) trial were: After controlling for baseline SDLP scores, participant age at licensure, and average speed during the simulation drives, were there

significant SDLP differences between the FOCAL+ (treatment) and ADTSEA (control) groups at the 1-month (Is the treatment effective?) and 6-month (Are significant treatment gains maintained?) assessments? Participant attrition resulted in 10.5% missing SDLP scores.

GEEs are an extension of fixed effect general linear model analyses, such as RM-ANCOVA, to not only allow for the analysis of correlated repeated measures count, multinomial, or binary dependent variable data (Liang & Zeger, 1986; Zeger & Liang, 1986; Zeger et al., 1988; cited in Hedeker & Gibbons, 2006) but to also allow the analysis of correlated repeated measures on a continuous scale that are not distributed normally (Locascio & Atri, 2011; Wang et al., 2016). Specifically, unlike fixed effect linear model analyses that assume multivariate normality of repeated dependent measures, GEEs: (1) assume marginal univariate normally distributed repeated responses (Hedeker & Gibbons, 2006), (2) fit a fixed effect model like an RM-ANCOVA to the repeated dependent variable data's marginal distribution (Liang & Zeger, 1986; Park, 1993), and (3) estimate the effect of the IV on the population response variable means (Albert, 1999; Edwards, 2000).

A key GEE assumption is that the fixed effect estimates that quantify IV influences on the repeated measures responses are of interest, but neither the structure of the correlated repeated dependent variables nor the residual variance/covariance estimates are needed (e.g., Burton et al., 1998; Diggle et al., 1994; Hardin & Hilbe, 2003; Ziegler, 2011). Specifically, although considered a nuisance in GEEs, the repeated measures correlation matrix must be specified in the service of fixed effect and fixed effect standard error estimation precision (Edwards, 2000; Locascio & Atri, 2011; Zeger & Liang, 1986). Correlation matrix specification options include: (1) an independence matrix that assumes uncorrelated responses, (2) an exchangeable matrix that assumes all possible pairwise dependent variable correlations are equivalent, (3) an autoregressive (or band-diagonal) matrix that assumes repeated dependent variables collected more closely in time are more highly correlated than dependent variables separated further in time, and (4) an unstructured matrix that estimates unique correlations for all possible pairwise repeated dependent variable measures (Garcia & Marder, 2017; Hedeker & Gibbons, 2006; Schluchter, 1988; Schober & Vetter, 2018)[7].

Just like RM-ANCOVA, GEEs also assume a temporally structured and balanced database with all dependent variable assessments collected consistently over time (Locascio & Atri, 2011). Although research has shown that fixed effect standard errors can be negatively biased under conditions of both a small sample size (Wang et al., 2016) and a misspecified repeated measures correlation matrix (Garcia & Marder, 2017), GEEs also have several well-known and often cited advantages. GEEs are very easy to estimate (Garcia & Marder, 2017) and have statistical power advantages under small sample size conditions (Ma et al., 2012; McNeish et al., 2017; Muth et al., 2016). Most importantly, GEEs produce unbiased fixed effect estimates that are robust to an incorrectly specified repeated measures correlation matrix (Ballinger, 2004; Diggle et al., 1994; Edwards, 2000; Garcia & Marder, 2017; Ghisletta & Spini, 2004; Hedeker & Gibbons, 2006; Wang et al., 2016; Zeger et al., 1988) under most conditions. Further, if assumptions are met and missing data is handled with multiple imputations (de Melo et al., 2022; Muth et al., 2016), GEEs are ideal under circumstances (e.g., research questions involving both treatment efficacy and maintenance) where fixed effect estimates alone are sufficient to answer the research question.

Returning to the Epstein et al. (2022) trial example, an explanatory paragraph linking Epstein et al. (2022) data specifics is offered before the diagnostic phase *BLIMP* syntax is shown in Appendix B. Diagnostic multiple imputation results showed an acceptable and stable PSR < 1.05 convergence value was observed conservatively after 500 burn-in iterations (i.e., consistent and stable PSR decreases were observed only after 500 iterations), as shown in Figure 3. *BLIMP* imputation syntax for a GEE is also shown in Appendix B. Pooled GEE results from analyzing the 100 imputed datasets showed, after controlling for age at licensure and average driving simulation speed, the FOCAL+ group had significantly lower SDLP versus the control group at both the 1-month ($\bar{X}_{\text{Difference}} = -.165$, $p = .001$; $d = .80$) and 6-month ($\bar{X}_{\text{Difference}} = -.188$, $p < .001$; $d = .83$) assessments, indicating both treatment efficacy and maintenance[8]. A summary of the GEE analysis results is shown in Figure 4. *SPSS*, *Stata*, and *R* data analysis syntax scripts are also shown in Appendix B.

Statistical considerations: GEE

Handling missing data with ML is not available in GEE because the estimation algorithm used to obtain parameter estimates is not ML in nature and requires complete data (see Footnote 7). Listwise deletion is the default in any statistical analysis software program capable of performing GEE analyses. Missing data in GEE analyses must be handled with MI. Syntax scripts in *BLIMP* to impute missing data, and in *SPSS*, *Stata*, and *R* to analyze and pool GEE analysis results from imputed data in are given in Appendix B.

LLMM example: Kashikar-Zuck et al. (2012)[9]

Specific design information for the Kashikar-Zuck et al. (2012) trial can be found in the original publication but is briefly summarized here. Functional Disability Inventory (FDI; Walker & Greene, 1991) data were collected from a sample of ($N = 114$) pediatric patients diagnosed with juvenile fibromyalgia syndrome (FMS) at three data collection sites. FDI scores were obtained at a baseline assessment, after which all participants were randomly assigned to receive either cognitive-behavioral therapy (CBT; treatment) or FMS

[7] As an aside, it is worth noting for researchers familiar with maximum-likelihood estimation that GEE estimation does not involve maximum likelihood, but instead proceeds in five iterative steps: (1) fixed effects quantifying the influence of the IV on response variable means are estimated first assuming the response variables are uncorrelated, (2) response variable residuals, defined as the difference between the observed and model-predicted values and based on both the preliminary fixed effect estimates and the user-specified correlation matrix, are computed, (3) the residuals matrix is used to modify and update both the fixed effect and fixed effect standard error estimates, (4) residuals are re-calculated, a new response variable correlation matrix is estimated, and fixed effect and fixed effect standard error estimates are again modified, and (5) the process repeats until a criterion for model convergence is met (i.e., fixed effect estimates and fixed effect standard error estimates can no longer be improved upon) (McNeish et al., 2017; Muth et al., 2016; Liang & Zeger, 1986). This estimation algorithm is not maximum likelihood in nature, requires complete data, and will listwise delete any missing data by default.

[8] Additional sensitivity analyses showed results did not change whether an independence, first-order autoregressive, or unstructured correlation matrix was specified.

[9] Funded by National Institutes of Health grant # R01-AR-050028 (PI: Kashikar-Zuck), ClinicalTrials.gov number NCT00086047.

BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

NOTE: Split chain PSR is being used. This splits each chain's iterations to create twice as many chains.

Comparing iterations across 2 chains	Highest PSR	Parameter #
51 to 100	1.111	24
101 to 200	1.039	11
151 to 300	1.034	25
201 to 400	1.046	25
251 to 500	1.013	13

Figure 3. Potential scale reduction (PSR) diagnostic output for the Epstein et al. (2022) trial data assuming a GEE imputation model.

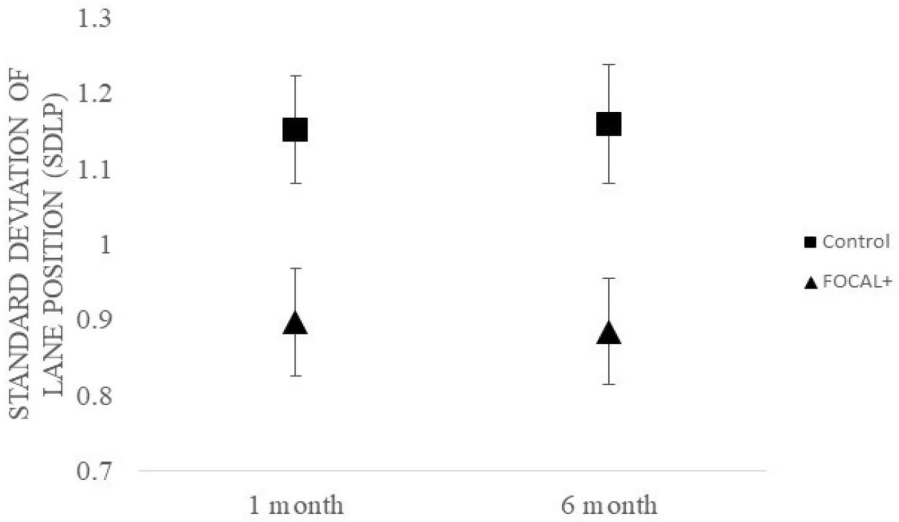

Figure 4. 95% CI error bar results post-baseline for the Epstein et al. (2022) trial data.

education (FM education; control). All participants received eight, 45-min intervention sessions, one per week. Control (FM education) participants engaged with the therapist in discussions of healthy lifestyle habits. Treatment condition (CBT) participants received muscle relaxation training, cognitive distraction and problem-solving skills, and behavioral activity pacing and relapse prevention techniques. Participants provided FDI scores at the end of the 8-week active treatment phase, and again at a 6-month follow-up. The research questions for the Kashikar-Zuck et al. (2012) trial are: (1) Does CBT result in greater FDI score reductions over time versus FM education after controlling for data collection site and tender point sensitivity average scores (based on an 18-point dolorimetry examination at each time point; see Kashikar-Zuck et al., 2012, p. 4)? If so, (2) Are significant CBT treatment effects over time moderated by depression (as measured by Child Depression Inventory T-scores [CDI_T]; Kovacs, 1992)?

Longitudinal linear mixed models (LLMM) further expand upon both RM-ANCOVA and GEE analyses under the assumption that both dependent variable average change over time (fixed effects), and variation in response change over time (random effects) across participants, are of interest (Bauer & Sterba, 2011; de Melo et al., 2022; Diggle et al., 2002; Edwards, 2000; Fitzmaurice et al., 2004; Harville, 1977; Laird & Ware, 1982; Lininger et al., 2015; Longford, 1993; Muth et al., 2016; Omar et al., 1999; Schober & Vetter, 2018; Snijders & Bosker, 2012; Ziegler, 2011). Unlike GEEs, LLMMs model both the joint relationship between predictors and repeated measures as well as the correlation between both the repeated dependent variables and dependent variable residual variances (Garcia & Marder, 2017; Muth et al., 2016). If the sample size is large (recall GEEs have an advantage at smaller sample sizes; see also Bell et al., 2008), explaining variation in response variable change over time is of interest, a repeated measures fixed effect is suspected to vary across participants, and the number of repeated measures contributed by each participant varies, LLMM (not GEE) is needed (Burton et al., 1998; Ma et al., 2012). However, compared to GEEs, LLMMs require more assumptions to be met (McNeish et al., 2017), small sample sizes are problematic for standard error estimation (Bell et al., 2008; Verbeke & Lesaffre, 1997), and model misspecifications can result in Type-1 inferential errors (Agresti et al., 2004; McNeish et al., 2017; Schober & Vetter, 2018). However, LLMMs afford greater flexibility in research design (unlike RM-ANCOVA and GEE, LLMM data need not be structured or balanced), types of predictors (time-

invariant or time-varying; Edwards, 2000; Raudenbush & Bryk, 2002), and missing data handling options (Garcia & Marder, 2017; but see below). LLMMs are ideal for research questions involving the explanation of variance in dependent variable changes over time (McNeish et al., 2017).

Returning to the Kashikar-Zuck et al. (2012) trial, descriptive statistics showed 6.43% of FDI, 7.02% of CDI, and 7.89% of tender point examination data was missing. Further, a priori unconditional analyses showed a model with intercept and linear slope fixed and random effects (quadratic and higher fixed effect and random effect terms were all nonsignificant; Snijders & Bosker, 2012) best modeled FDI score changes during and after active treatment. Again, a paragraph linking the specifics of the Kashikar-Zuck et al. trial data to *BLIMP* syntax specifications is provided before the *BLIMP* diagnostic syntax is shown in Appendix C. Multiple imputation diagnostics showed an acceptable PSR (<1.05) convergence value was obtained after 2000 iterations as shown in Figure 5. The *BLIMP* imputation syntax is also shown in Appendix C. Pooled imputation longitudinal LLMM analysis results showed CBT participants had significantly lower FDI scores over time (`IV_GROUP * TIME` $= -1.79$, $t = -2.13$, $p < .05$), but this effect was not moderated by depression (`IV_GROUP * TIME * CDI_T` $= -0.05$, $t = -.31$, $p > .05$). Effect size computations showed that treatment group, time in assessment months, and the group by time interaction explained ($R^2_{\text{PSEUDO}} = .039$; Rights & Sterba, 2021) roughly 4% of FDI variance. Model-predicted LLMM analysis results are shown in Figure 6. *SPSS*, *Stata*, and *R* data analysis syntax scripts are also shown in Appendix C.

Statistical considerations: LLMM

MI is needed to handle missing data under several different data analysis scenarios when using LLMM analyses. Monte Carlo simulation research has shown that notable parameter estimate bias occurs in LLMM estimation scenarios involving either a random effect (i.e., if missing data is limited to the repeated response variable measures only) or a cross-level interaction (i.e., if one or more predictor, moderator, or control covariate variables also show missing data) if missing data is handled with ML (see Enders et al., 2018, 2020; see also Grund et al., 2021a; Keller & Enders, 2023). The Kashikar-Zuck et al. (2012) example has both a random effect (random slope; the effect of *TIME* as a predictor of changes in FDI scores over time is estimated as varying randomly across participants) and a cross-level interaction (the random linear slope variance is predicted by a binary indicator of treatment random assignment, but also by a moderator [CDI] and a control covariate [tender-point exam scores] that both show missing data). Said differently, and unlike RM-ANCOVA, even if the variable used as the metric of time and the variable indicating treatment random assignment both have complete data and missing is limited to the response variable only, notable bias will occur in both the random slope estimate and the cross-level interaction estimate likely needed to answer the longitudinal trial research question if ML is used to handle missing data. Additional research has also shown that the use of the Yeo–Johnson transform procedure, together with model-based imputation with factored regression, produced unbiased LLMM moderation effect estimates that involved nonnormal time-invariant (level 2) predictors (e.g., CDI scores; Keller & Enders, 2023). Syntax scripts in *BLIMP* to impute missing data, and syntax scripts for *SPSS*, *Stata*, and *R* to analyze and pool LLMM analysis results from multiple imputations in are given in Appendix C.

Summary

Methodologists have known for decades that, although no missing data handling technique can eliminate parameter estimate bias due to missing data, ML and MI are the preferred missing data handling techniques because they have been shown empirically to optimally minimize parameter estimate bias. However, recent methodological research has posed questions as to whether both equally minimize bias under all data analysis scenarios or are there analysis models for which one better minimizes bias versus the other. Further, McNeish et al. (2017) stated that researchers tended to use complicated data analysis models when simpler ones would more efficiently answer the research question because such methods were likely the only ones taught. These two facts beg an obvious question: Why? A reasonable answer might involve the default maximum likelihood of missing data handling that is unavailable for simpler models (GEE; leaving MI as the only option), but commonly available for more complex models (LLMM) even if more recent research shows its use results in biased parameter estimates under specific data analysis conditions (Enders et al., 2018, 2020). As shown in the example analyses, ML is not always the appropriate or even available option for handling missing data in longitudinal clinical trials. Few researchers have experience with multiple imputations, and fewer still have access to multiple imputation software packages capable of properly imputing missing

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                         201 to 400            1.193            34
                         401 to 800            1.092             9
                         601 to 1200           1.079            41
                         801 to 1600           1.083             3
                        1001 to 2000           1.030             3
```

Figure 5. Potential scale reduction (PSR) diagnostic output for the Kashikar-Zuck et al. (2012) trial data assuming a LLMM imputation model.

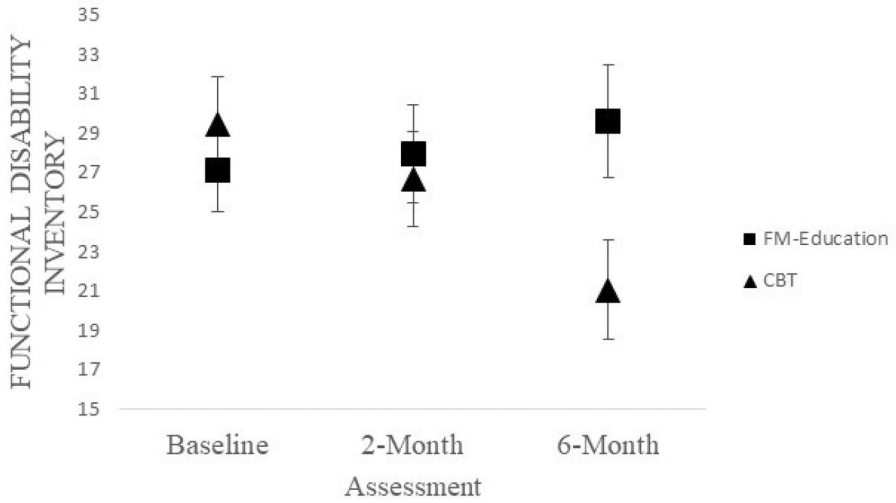

Figure 6. 95% CI error bar trajectory results for the Kashikar-Zuck et al. (2012) trial data.

Table 2. A longitudinal trial data analysis decision tree (from McNeish et al., 2017)

	RM-ANCOVA	GEE	LLMM
Advantages	Powerful if assumptions are met (normality, compound symmetry) and no missing data (if so, can be analyzed as a general linear model [GLM])	Very flexible: Can analyze longitudinal DV data on any measurement scale Continuous DVs need not be normally distributed Greater statistical power at smaller *N* vs. LLMM	Explicitly models the correlated nature of clustered/repeated measures data via level-1 and level-2 Missing data are often handled via the default maximum likelihood (ML) estimation algorithm
Disadvantages	Assumptions are rarely met Listwise deletion is often the default	The likelihood of a Type-1 error increases as *N* decreases and the discrepancy between the user-specified and population correlation matrix increases Listwise deletion is often the default	Both large *N*s and properly specified residual covariance matrices at level 1 and level 2 are needed for unbiased estimates ML is not appropriate for missing data handling for random slopes or cross-level interactions
Ideal	If only $T = 2$ timepoints (pretest/posttest) of data are available for analysis	If time-point-specific fixed effect estimates are sufficient to answer the research question (random effects are discarded post-estimation)	If overall DV trajectory differences by random assignment are needed to answer the research question

data in a longitudinal trial. Although multiple imputation missing data handling is available in several commercially available software packages, *BLIMP* is featured here for three reasons: (1) it is internet freeware, making it completely accessible, (2) it can handle missing data on any measurement scale, and (3) it has numerous publications documenting both its development and efficacy. Further, the example analyses shown here provide only a very small sample of the imputation and Bayesian analysis capabilities possible in *BLIMP*.

Table 2 offers a summary of the advantages and disadvantages of all three options to assist trialists in determining the optimal data analysis model for their longitudinal research design circumstances. A RM-ANCOVA was used to analyze data from the Zeidan et al. (2015) trial because: (1) the VAS scores were normally distributed; testing the null hypothesis of normally distributed data showed a nonsignificant result, (2) as a tightly controlled laboratory experiment, the assumption of structured and balanced data collection was met, and (3) analysis results were identical whether a compound symmetric or unstructured covariance matrix was specified. As such, an RM-ANCOVA was the most efficient method for answering the research question. A GEE was needed for the Epstein et al. (2022) trial data because variances, such as standard deviation of lane position, are well-known to be non-normally distributed (e.g., McNeish & Hamaker, 2020). Recall a GEE does not require continuous dependent variables to be normally distributed. Further, GEE rather than an LLMM was needed for the Epstein et al. (2022) trial data because evidence of treatment efficacy at the 1-month assessment and evidence of treatment maintenance at the 6-month assessment were needed to answer the research questions. Finally, GEE results for the Epstein et al. (2022) trial data were identical regardless of the correlation matrix structure specified. In the Kashikar-Zuck et al. (2012) trial data, trajectory differences between the independent variable groups assessed in an LLMM answered questions regarding CBT better-changing disability over time compared to an educational control. It is worth noting that RM-ANOVA could

have been used if analysis assumptions had been met, and GEE could have been used if research questions involved short-term treatment efficacy and longer-term treatment maintenance. In summary, McNeish et al.'s (2017) implicit recommendation is a prudent one: estimate only the parameters needed to answer the specific research question(s).

Finally, from a practical perspective, all analyses shown here were performed using *SPSS* and *Stata* due to their popularity and widespread use. Both statistical analysis software packages are commercially available but at a notable cost. Although not obvious from the examples shown here, *BLIMP* will provide Bayesian model parameter estimates by default if model-based (H_0:) imputation is used as was shown here. Researchers satisfied with Bayesian parameter estimates can handle missing data and answer their research questions in a single step using *BLIMP*. For researchers more comfortable with frequentist (e.g., ML) estimation, as shown in the examples here, *SPSS*, *Stata*, and *R* can analyze and pool imputed data from *BLIMP* using frequentist analyses. Perhaps most importantly, *BLIMP* and *R* both are empirically well-validated internet freeware programs. Assuming access to a stable internet connection, researchers at all levels of training anywhere in the world can handle missing values, analyze longitudinal trial data, and answer research questions completely free of cost.

Data availability

The data underlying this article were provided by the grant recipients (see footnotes) and manuscript publication authors by permission. Data will be shared on request to the corresponding author and only with the permission of the grant authors and funding sources if relevant.

Author contributions

James Peugh (Data curation [Lead], Formal analysis [Lead], Investigation [Equal], Methodology [Lead], Project administration [Equal], Software [Lead], Supervision [Equal], Validation [Equal], Writing—original draft [Lead], Writing—review & editing [Equal]) and Constance Mara (Conceptualization [Lead], Data curation [Supporting], Formal analysis [Supporting], Investigation [Supporting], Methodology [Supporting], Project administration [Supporting], Resources [Supporting], Software [Supporting], Supervision [Supporting], Validation [Supporting], Writing—original draft [Supporting], Writing—review & editing [Equal])

Funding

None declared

Conflicts of interest: None declared.

Acknowledgments

The authors would like to extend sincere thanks and gratitude to (in alphabetical order): Stephen Becker, Robert Coghill, Craig Enders, Jeffery Epstein, Susmita Kashikar-Zuck, Brian Keller, Kaylee Litson, and Fadel Zeidan for allowing the use of their trial data for the analytic illustrative examples and their helpful feedback on the manuscript.

References

Albert, P. S. (1999). Longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine*, *18*, 1707–1732. https://doi.org/10.1002/(SICI)1097-0258(19990715)18:13<1707::AID-SIM138>3.0.CO;2-H

Abraham, W. T., & Russell, D. W. (2008). Statistical power analysis in psychological research. *Social and Personality Psychology Compass*, *2*, 283–301. https://doi.org/10.1111/j.1751-9004.2007.00052.x

Agresti, A., Caffo, B., & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis*, *47*, 639–653. http://dx.https://doi.org/10.1016/j.csda.2003.12.009

Alacam, E., Enders, C. K., Du, H., & Keller, B. T. (2023). A factored regression model for composite scores with item-level missing data. *Psychological Methods*, 1–20. https://doi.org/10.1037/met0000584

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, *7*, 127–150. https://doi.org/10.1177/1094428104263672

Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R.; Alzheimer's Disease Neuroimaging Initiative. (2015). Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Statistical Methods in Medical Research*, *24*, 462–487. https://doi.org/10.1177/0962280214521348

Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: performance of alternative specifications and methods of estimation. *Psychological Methods*, *16*, 373–390. http://dx.https://doi.org/10.1037/a0025813

Bell, B. A., Ferron, J. M., & Kromrey, J. D. (2008). *Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two-level models*. Paper presented at the proceedings of the Section on Survey Research Methods, Joint Statistical Meetings. Alexandria, VA: American Statistical Association. Retrieved from https://www.amstat.org/sections/srms/Proceedings/y2008/Files/300933.pdf

Ben, Â. J., van Dongen, J. M., Alili, M. E., Heymans, M. W., Twisk, J. W. R., MacNeil-Vroomen, J. L., de Wit, M., van Dijk, S. E. M., Oosterhuis, T., & Bosmans, J. E. (2023). The handling of missing data in trial-based economic evaluations: should data be multiply imputed prior to longitudinal linear mixed-model analyses? *The European Journal of Health Economics: HEPAC: Health Economics in Prevention and Care*, *24*, 951–965. https://doi.org/10.1007/s10198-022-01525-y

Burton, P., Gurrin, L., & Sly, P. (1998). Tutorial in biostatistics. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modeling. *Statistics in Medicine*, *17*, 1261–1291. https://doi.org/10.1002/(sici)1097-0258(19980615)17:113.0.co;2-z

Carpenter, J. R., Bartlett, J., Morris, T., Wood, A., Quartagno, M., & Kenward, M. G. (2023). *Multiple imputation and its application* (2nd ed.). Wiley.

Carpenter, J. R., & Kenward, M. G. (2007). *Missing data in randomised controlled trials: A practical guide*. https://researchonline.lshtm.ac.uk/id/eprint/4018500

Carpenter, J. R., Roger, J. H., & Kenward, M. G. (2013). Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics*, *23*, 1352–1371.

de Melo, M. B., Daldegan-Bueno, D., Menezes Oliveira, M. G., & de Souza, A. L. (2022). Beyond ANOVA and MANOVA for repeated measures: advantages of generalized estimated equations and generalized linear mixed models and its use in neuroscience research. *The European Journal of Neuroscience*, *56*, 6089–6098. https://doi.org/10.1111/ejn.15858

DeSouza, C. M., Legedza, A. T., & Sankoh, A. J. (2009). An overview of practical approaches for handling missing data in clinical trials. *Journal of Biopharmaceutical Statistics*, *19*, 1055–1073. https://doi.org/10.1080/10543400903242795

Diggle, P. J., Heagerty, P., Kung-Yee, L., & Zeger, S. L. (2002). *Analysis of Longitudinal Data* (2nd ed.). Oxford University Press.

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford.

Du, H., Alacam, E., Mena, S., & Keller, B. T. (2022). Compatibility in imputation specification. *Behavior Research Methods*, *54*, 2962–2980. https://doi.org/10.3758/s13428-021-01749-5

Edwards, L. J. (2000). Modern statistical techniques for the analysis of longitudinal data in biomedical research. *Pediatric Pulmonology*, *30*, 330–344. https://doi.org/10.1002/1099-0496(200010)30:4<330::AID-PPUL10>3.0.CO;2-D

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.

Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). Guilford Publications.

Enders, C. K. (2023a). Missing data: an update on the state of the art. *Psychological Methods, 1–18*. https://dx.https://doi.org/10.1037/met0000563

Enders, C. K. (2023b). September). *Multiple imputation in BLIMP*. https://www.youtube.com/watch?v=vvrMc9cUrS0

Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and other nonlinear terms. *Psychological Methods*, *25*, 88–112. http://dx.https://doi.org/10.1037/met0000228

Enders, C. K., Hayes, T., & Du, H. (2018). A comparison of multilevel imputation schemes for random coefficient models: fully conditional specification and joint model imputation with random covariance matrices. *Multivariate Behavioral Research*, *53*, 695–713. https://doi.org/10.1080/00273171.2018.1477040

Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*, *23*, 298–317. ().., -. http://dx.https://doi.org/10.1037/met0000148

Enders, C. K., Mistler, S. A., & Keller, B. T. (2016). Multilevel multiple imputation: a review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, *21*, 222–240. https://doi.org/10.1037/met0000063

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. https://doi.org/10.1037/1082-989X.12.2.121

Epstein, J. N., Garner, A. A., Kiefer, A. W., Peugh, J., Tamm, L., MacPherson, R. P., Simon, J. O., & Fisher, D. L. (2022). Trial of training to reduce driver inattention in teens with ADHD. *The New England Journal of Medicine*, *387*, 2056–2066. https://doi.org/10.1056/NEJMoa2204783

Everitt, B. S. (1998). Analysis of longitudinal data. Beyond MANOVA. *The British Journal of Psychiatry: The Journal of Mental Science*, *172*, 7–10. https://doi.org/10.1192/bjp.172.1.7

Fiero, M. H., Hsu, C. H., & Bell, M. L. (2017). A pattern-mixture model approach for handling missing continuous outcome data in longitudinal cluster randomized trials. *Statistics in Medicine*, *36*, 4094–4105. https://doi.org/10.1002/sim.7418

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Wiley.

Garcia, T. P., & Marder, K. (2017). Statistical approaches to longitudinal data analysis in neurodegenerative diseases: Huntington's disease as a model. *Current Neurology and Neuroscience Reports*, *17*, 14. https://doi.org/10.1007/s11910-017-0723-4

Ghisletta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics*, *29*, 421–437. https://doi.org/10.3102/10769986029004421

Girden, E. R. (1992). ANOVA: repeated measures. *In Quantitative Applications in the Social Sciences* (vol. *84*, pp. 41–59). Newbury Park, CA: Sage.

Gomer, B., & Yuan, K. H. (2021). Subtypes of the missing not at random missing data mechanism. *Psychological Methods*, *26*, 559–598. https://doi.org/10.1037/met0000377

Goldstein, H., Carpenter, J. R., & Browne, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *177*, 553–564. https://doi.org/10.1111/rssa.12022

Grady, J. J., & Helms, R. W. (1995). Model selection techniques for the covariance matrix for incomplete longitudinal data. *Statistics in Medicine*, *14*, 1397–1416. https://doi.org/10.1002/sim.4780141302

Graham, J. W. (2012). *Missing data: analysis and design*. Springer Science & Business Media.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science: The Official Journal of the Society for Prevention Research*, *8*, 206–213. https://doi.org/10.1007/s11121-007-0070-9

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: structural equation and multilevel modeling approaches*. Guilford Publications.

Grund, S., Lüdtke, O., & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: simulations and recommendations. *Organizational Research Methods*, *21*, 111–149. https://doi.org/10.1177/1094428117703686

Grund, S., Lüdtke, O., & Robitzsch, A. (2019). Missing data in multilevel research. In Humphrey, S. E., & LeBreton, J. M. (Eds.), *Handbook for multilevel theory, measurement, and analysis* (pp. 365–386). American Psychological Association.

Grund, S., Lüdtke, O., & Robitzsch, A. (2021a). Multiple imputation of missing data in multilevel models with the R package mdmb: a flexible sequential modeling approach. *Behavior Research Methods*, *53*, 2631–2649. https://doi.org/10.3758/s13428-020-01530-0

Grund, S., Lüdtke, O., & Robitzsch, A. (2021b). On the treatment of missing data in background questionnaires in educational large-scale assessments: an evaluation of different procedures. *Journal of Educational and Behavioral Statistics*, *46*, 430–465. https://doi.org/10.3102/1076998620959058

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations*. Chapman & Hall/CRC.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*, 320–338.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Wiley-Interscience.

Helms, R. W. (1992). Intentionally incomplete longitudinal designs: I. Methodology and comparison of some full span designs. *Statistics in Medicine*, *11*, 1889–1913. https://doi.org/10.1002/sim.4780111411

Howell, D. C. (2007). Repeated measures designs. In Howell, D. C. (Ed.), *Statistical methods for psychology* (pp. 439–492). Thomson Wadsworth.

Kashikar-Zuck, S., Ting, T. V., Arnold, L. M., Bean, J., Powers, S. W., Graham, T. B., Passo, M. H., Schikler, K. N., Hashkes, P. J., Spalding, S., Lynch-Jordan, A. M., Banez, G., Richards, M. M., & Lovell, D. J. (2012). Cognitive behavioral therapy for the treatment of juvenile fibromyalgia: a multisite, single-blind, randomized, controlled clinical trial. *Arthritis and Rheumatism*, *64*, 297–305. https://doi.org/10.1002/art.30644

Keller, B. T. (2021). An introduction to factored regression models with Blimp. *Psych*, *4*, 10–37. https://doi.org/10.3390/psych4010002

Keller, B. T., & Enders, C. K. (2022). *Blimp user's guide (version 3)*. www.appliedmissingdata.com/blimp

Keller, B. T., & Enders, C. K. (2023). An investigation of factored regression missing data methods for multilevel models with cross-

level interactions. *Multivariate Behavioral Research*, *58*, 938–963. https://doi.org/10.1080/00273171.2022.2147049

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: a review. *The British Journal of Mathematical and Statistical Psychology*, *54*, 1–20. https://doi.org/10.1348/000711001159357

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied Regression Analysis and Multivariable Methods* (3rd ed.). Doxbury Press.

Kleinke, K., Reinecke, J., Salfrán, D., & Spiess, M. (2020). *Applied multiple imputation: advantages, pitfalls, new developments and applications in R*. Springer.

Kovacs, M. (1992). *Children's depression inventory*. Multi-Health Systems.

Krueger, C., & Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing*, *6*, 151–157. https://doi.org/10.1177/1099800404267682

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974. http://dx.https://doi.org/10.230

Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 649–662.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22. http://dx.https://doi.org/10.1093/biomet/73.1.13

Lininger, M., Spybrook, J., & Cheatham, C. C. (2015). Hierarchical linear model: thinking outside the traditional repeated-measures analysis-of-variance box. *Journal of Athletic Training*, *50*, 438–441. https://doi.org/10.4085/1062-6050-49.5.09

Little, R. J. A. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association*, *87*, 1227–1237.

Little, R. J. A. (2013). In praise of simplicity not mathematistry! Ten simple powerful ideas for the statistical scientist. *Journal of the American Statistical Association*, *108*, 359–369.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. (2014). On the joys of missing data. *Journal of Pediatric Psychology*, *39*, 151–162. https://doi.org/10.1093/jpepsy/jst048.

Little, R. J. A., & Rubin, D. B. (2002). Bayes and multiple imputation. In Little, R. J. A., & Rubin, D. B. (Eds.), *Statistical analysis with missing data* (2nd ed., pp. 200–223). Wiley.

Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse Wishart and separation-strategy priors for Bayesian estimation of covariance parameter matrix in growth curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 354–367. https://doi.org/10.1080/10705511.2015.1057285

Locascio, J. J., & Atri, A. (2011). An overview of longitudinal data analysis methods for neurological research. *Dementia and Geriatric Cognitive Disorders Extra*, *1*, 330–357. https://doi.org/10.1159/000330228

Longford, N. T. (1993). *Random coefficient models*. Oxford University Press.

Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression models involving nonlinear effects with missing data: a sequential modeling approach using Bayesian estimation. *Psychological Methods*, *25*, 157–181. https://doi.org/10.1037/met0000233

Ma, Y., Mazumdar, M., & Memtsoudis, S. G. (2012). Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. *Regional Anesthesia and Pain Medicine*, *37*, 99–105. https://doi.org/10.1097/AAP.0b013e31823ebc74

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, *25*, 610–635. https://doi.org/10.1037/met0000250

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*, 114–140. http://dx.https://doi.org/10.1037/met0000078

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166. https://doi.org/10.1037/0033-2909.105.1.156

Muth, C., Bales, K. L., Hinde, K., Maninger, N., Mendoza, S. P., & Ferrer, E. (2016). Alternative models for small samples in psychological research: applying linear mixed effects models and generalized estimating equations to repeated measures data. *Educational and Psychological Measurement*, *76*, 64–87. https://doi.org/10.1177/0013164415580432

Muthén, B. O. (2010). *Bayesian analysis in MPlus: a brief introduction*. Incomplete draft, version 1. http://www.statmodel.com/download/IntroBayesVersion%201.pdf

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

Myers, J. L. (1979). *Fundamentals of Experimental Design* (3rd ed.). Allyn & Bacon.

Omar, R. Z., Wright, E. M., Turner, R. M., & Thompson, S. G. (1999). Analyzing repeated measurements data: a practical comparison of methods. *Statistics in Medicine*, *18*, 1587–1603. https://doi.org/10.1002/(SICI)1097-0258(19990715)18:13<1587::AID-SIM141>3.0.CO;2-Z

Park, T. (1993). A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements. *Statistics in Medicine*, *12*, 1723–1732. https://doi.org/10.1002/sim.4780121807

Peugh, J. L., & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, *65*, 717–741. https://doi.org/10.1177/0013164405278558

Peugh, J. L., & Heck, R. H. (2017). Conducting three-level longitudinal analyses. *The Journal of Early Adolescence*, *37*, 7–58. https://doi.org/10.1177/0272431616642329

Peugh, J. L., Toland, M. D., & Harrison, H. (2023). A tutorial for handling suspected missing not at random data in longitudinal clinical trials. *The Quantitative Methods for Psychology*, *19*(4), 347–367. https://doi.org/10.20982/tqmp.19.4.p347.

Price, D. D., Bush, F. M., Long, S., & Harkins, S. W. (1994). A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain*, *56*, 217–226. https://doi.org/10.1016/0304-3959(94)90097-3

Quartagno, M., & Carpenter, J. R. (2019). Multiple imputation for discrete data: evaluation of the joint latent normal model. *Biometrical Journal. Biometrische Zeitschrift*, *61*, 1003–1019. https://doi.org/10.1002/bimj.201800222

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (vol. *1*, 2nd ed.). Sage.

Rights, J. D., & Sterba, S. K. (2021). Effect size measures for longitudinal growth analyses: extending a framework of multilevel model R-squareds to accommodate heteroscedasticity, autocorrelation, nonlinearity, and alternative centering strategies. *New Directions for Child and Adolescent Development*, *2021*, 65–110. https://doi.org/10.1002/cad.20387

Rubin, L. H., Witkiewitz, K., Andre, J. S., & Reilly, S. (2007). Methods for handling missing data in the behavioral neurosciences: don't throw the baby rat out with the bath water. *Journal of Undergraduate Neuroscience Education*, *5*, A71–A77.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*, 3–15. https://doi.org/10.1177/096228029900800102

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, *57*, 19–35. https://doi.org/10.1111/1467-9574.00218

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: a data analyst's perspective.

Multivariate Behavioral Research, *33*, 545–571. https://doi.org/10.1207/s15327906mbr3304_5

Schluchter, M. D. (1988). Analysis of incomplete multivariate data using linear models with structured covariance matrices. *Statistics in Medicine*, *7*, 317–324. https://doi.org/10.1002/sim.4780070132

Schober, P., & Vetter, T. R. (2018). Repeated measures designs and analysis of longitudinal data: if at first you do not succeed—try, try again. *Anesthesia and Analgesia*, *127*, 569–575. https://doi.org/10.1213/ANE.0000000000003511

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford University Press.

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

van Buuren, S. (2021). *Flexible imputation of missing data* (2nd ed.). Chapman & Hall/CRC.

van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*, 1049–1064. https://doi.org/10.1080/10629360600810434

van Ginkel, J. R., & Kroonenberg, P. M. (2014). Analysis of variance of multiply imputed data. *Multivariate Behavioral Research*, *49*, 78–91. https://doi.org/10.1080/00273171.2013.855890

Verbeke, G., & Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*, *23*, 541–556. http://dx.https://doi.org/10.1016/S0167-9473(96)00047-3

von Hippel, P. T. (2007). Regression with missing Ys: an improved strategy for analyzing multiply imputed data. *Sociological Methodology*, *37*, 83–117. http://dx.doi.org/10.1111/j.1467-9531.2007.00180.x

von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology*, *39*, 265–291. http://dx.doi.org/10.1111/j.1467-9531.2009.01215.x

Walker, L. S., & Greene, J. W. (1991). The functional disability inventory: measuring a neglected dimension of child health status. *Journal of Pediatric Psychology*, *16*, 39–58. https://doi.org/10.1093/jpepsy/16.1.39

Wang, M., Kong, L., Li, Z., & Zhang, L. (2016). Covariance estimators for generalized estimating equations (GEE) in longitudinal analysis with small samples. *Statistics in Medicine*, *35*, 1706–1721. https://doi.org/10.1002/sim.6817

White, I. R., & Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine*, *29*, 2920–2931. https://doi.org/10.1002/sim.3944

Wulff, J. N., Sajons, G. B., Pogrebna, G., Lonati, S., Bastardoz, N., Banks, G. C., & Antonakis, J. (2023). Common methodological mistakes. *The Leadership Quarterly*, *34*, 101677.

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*, 954–959. https://doi.org/10.1093/biomet/87.4.954

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121–130. https://doi.org/10.2307/2531248

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, *44*, 1049–1060. http://dx.https://doi.org/10.2307/2531734

Zeidan, F., Emerson, N. M., Farris, S. R., Ray, J. N., Jung, Y., McHaffie, J. G., & Coghill, R. C. (2015). Mindfulness meditation-based pain relief employs different neural mechanisms than placebo and sham mindfulness meditation-induced analgesia. *The Journal of Neuroscience*, *35*, 15307–15325. https://doi.org/10.1523/JNEUROSCI.2542-15.2015

Ziegler, A. (2011). *Generalized Estimating Equations*. Springer.

Appendix A

ML for Missing Data: *SPSS, Stata,* and *R* scripts for analyzing the Zeidan et al. (2015) data

The Zeidan et al. (2015) data only contains VAS pretest/posttest (or, "y-side") missing data. In this case, handling missing data with ML estimation is an acceptable approach and simply involves fitting the analysis model to the sample data (Little, 1992; Little & Rubin, 2002; van Buuren, 2021; von Hippel, 2007). Shown below are the data analytic syntax scripts needed to analyze the Zeidan et al. (2015) data using ML estimation in *SPSS*, *Stata*, and *R*. However, multiple imputation scripts in *BLIMP*, and syntax scripts needed to analyze imputed data in *SPSS*, *Stata*, and *R* are also offered below for researchers wanting to use an RM-ANCOVA data analysis model but facing the more likely scenario of having both response variables ("y-side") and control covariate ("x-side") missing data.

SPSS—ML

```
MIXED PRE_POST BY W_FACTOR B_FACTOR WITH Age Gender BASELINE
  /FIXED=W_FACTOR B_FACTOR W_FACTOR*B_FACTOR | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /EMMEANS = TABLES (B_FACTOR*W_FACTOR) compare(B_FACTOR)
  /REPEATED=W_FACTOR | SUBJECT(ID) COVTYPE(CS).
```

(*Note*: The 'WITHIN' and 'BETWEEN' variable names had to be changed to 'W_FACTOR' and 'B_FACTOR' because "within" is an *SPSS* syntax keyword.)

Stata-ML

```
/*Overall Model*/
mixed PRE_POST c. WITHIN##c.BETWEEN BASELINE Age Gender || ID

/*Post-hoc follow-up, reference group = group 3*/
mixed PRE_POST WITHIN##ib3.BETWEEN BASELINE Age Gender || ID
```

R—ML

```
install.packages("lavaan")
library(lavaan)

Zeidan <- read.csv("C:/Users/peuu3c/OneDrive—cchmc/Desktop/Zeidan_ML.csv")

Zeidan$int <- Zeidan$WITHIN*Zeidan$BETWEEN

Zeidan_model <- "
PRE_POST ~ Age + Gender + BASELINE + WITHIN + BETWEEN + int

#These 7 lines must be included, otherwise, means/intercepts are #set to 0 and parameter estimates
are severely biased
PRE_POST ~ 1
Age ~ 1
Gender ~ 1
BASELINE ~ 1
WITHIN ~ 1
BETWEEN ~ 1
int ~ 1"

Zeidan_ML <- lavaan(model = Zeidan_model,
missing = "ML",
data = Zeidan,
cluster = "ID",
auto.var = TRUE,
estimator = "ML")
```

```
summary(Zeidan_ML, fit.measures = TRUE)

#POST-HOC FOLLOW-UP

library(lavaan)
library(haven)
library(fastDummies)

Zeidan <- read_sav("D:/Zeidan.sav")

data <- subset(Zeidan, WITHIN == 2)

require(fastDummies)
data2 <- dummy_cols(data, select_columns="BETWEEN")

Zeidan_model2 <- "
PRE_POST ~ 1 + BETWEEN_1 + BETWEEN_2 + BETWEEN_4"

Zeidan_ML2 <- lavaan(model = Zeidan_model2,
missing = "ML",
data = data2,
auto.var = TRUE,
estimator = "ML")

summary(Zeidan_ML2, fit.measures = TRUE)
```

MI for Missing Data: Zeidan et al. (2015) data specifics to *BLIMP* diagnostic and imputation syntax

The Zeidan et al. (2015) dataset variables are listed in the order in which they appear in the data file after the (`VARIABLES:`) command. Binary variables specifying participant biological sex (`Gender`: 0 = male, 1 = female), the repeated measures effect (`WITHIN`: 0 = pretest, 1 = posttest), and the multinomial randomization variable (`BETWEEN`, see values listed in text) are all listed on the `ORDINAL:` command line. The participant ID variable (`level2id`) is listed on the (`CLUSTERID:`) line to indicate that both pretest and posttest VAS scores are correlated within each participant. `Age, Gender, Baseline`, `WITHIN`, and `BETWEEN` variables are listed on the (`FIXED:`) command line because those variables have no missing data. The (`MISSING: -99;`) command specifies the required single value missing data indicator. The imputation model shown after the (`MODEL:`) command indicates pretest and posttest VAS scores (`PRE_POST`) are predicted by (~) the repeated measures (`WITHIN`) and random assignment (`BETWEEN`) main effects and their interaction (`WITHIN*BETWEEN`) after controlling for covariate effects (`Age, Gender, Baseline`). The (`| 1@0`) specification at the end of the (`MODEL:`) command line specifies no random slopes (e.g., the repeated measures [`WITHIN`] effect is not specified to vary randomly across participants. *BLIMP* estimates a random intercept by default if the `CLUSTERID:` command is used). Both the total number of iterations (`ITERATIONS:`) as well as the number of burn-in (`BURN:`) iterations are set to high values to achieve potential scale reduction (PSR) stability (Keller & Enders, 2022, p. 86). The (`OUTPUT:`) command line requests potential scale reduction (`PSR`) values to be listed in the output window for diagnostic purposes. As shown in Figure 1, multiple imputation diagnostic results showed admissible PSR < 1.05 convergence was obtained after a maximum of 200 burn-in iterations. Despite needing only 200 iterations to achieve PSR convergence, 5000 burn-in (`BURN:`) iterations and 10 000 total iterations (`ITERATIONS:`) are recommended as a minimum for the imputation phase (which should be increased if needed to obtain PSR < 1.05; Keller & Enders, 2023, pp. 85–90), both of which will be used from this point forward.

BLIMP imputation phase syntax is also shown below. The imputation syntax differs from the diagnostic syntax in two ways. First, the (`OUTPUT: PSR;`) command line is replaced with (`NIMPS:, CHAINS:, & SAVE:`) command lines. Second, the number of imputed datasets is specified on the (`NIMPS: 100;`) command line, and the number of chains listed on the (`CHAINS: 100;`) line both were set to a value of 100. This specifies that each imputed dataset be computed by a separate chain to avoid potential autocorrelation across imputed datasets (e.g., the `THIN:` command in *BLIMP* is not needed; see Enders, 2022, p. 267).

BLIMP Diagnostic and Imputation Syntax Scripts for Missing Data

SPSS, Stata, and R syntax for RM-ANCOVA Estimation

BLIMP—Diagnostic

```
DATA: RM_ANCOVA.dat;
VARIABLES: level2id Age Gender Baseline WITHIN BETWEEN PRE_POST;
ORDINAL: Gender WITHIN BETWEEN;
CLUSTERID: level2id;
```

```
MISSING: -99;
FIXED: Age Gender Baseline WITHIN BETWEEN;
MODEL:
PRE_POST ~ Age Gender Baseline WITHIN BETWEEN WITHIN*BETWEEN | 1@0;
SEED: 90291;
BURN: 10000;
ITERATIONS: 10000;
OUTPUT: PSR;
```

BLIMP—Imputation

```
DATA: RM_ANCOVA.dat;
VARIABLES: level2id Age Gender Baseline WITHIN BETWEEN PRE_POST;
ORDINAL: Gender WITHIN BETWEEN;
CLUSTERID: level2id;
MISSING: -99;
FIXED: Age Gender Baseline WITHIN BETWEEN;
MODEL:
PRE_POST ~ Age Gender Baseline WITHIN BETWEEN WITHIN*BETWEEN | 1@0;
SEED: 90291;
BURN: 5000;
ITERATIONS: 10000;
NIMPS: 100;
CHAINS: 100;
SAVE: stacked = RM_ANCOVA_imps.dat;
```

SPSS Syntax: RM-ANCOVA Estimation

```
data list free file = 'C : \RM_ANCOVA_imps.dat'
  /imputation_ level2id Age Gender Baseline WFACTOR BFACTOR PRE_POST.
EXECUTE.

*activate pooling algorithm.
sort cases by imputation_.
split file layered by imputation_.

*This code makes 1= post and 2 = pre.
compute WFACTOR = 3 - WFACTOR.
EXECUTE.

*This code makes
*BFACTOR: 1 = LISTEN, 2 = MEDITATION, 3 = SHAM, 4 = CONTROL
DATASET ACTIVATE DataSet1.
RECODE BFACTOR (2 = 1) (3 = 2) (4 = 3) (1 = 4).
EXECUTE.

MIXED PRE_POST BY WFACTOR BFACTOR WITH Baseline Age Gender
  /FIXED=WFACTOR BFACTOR WFACTOR*BFACTOR
  /METHOD=REML
  /PRINT=DESCRIPTIVES SOLUTION TESTCOV
  /REPEATED=WFACTOR | SUBJECT(level2id) COVTYPE(CS)
  /EMMEANS=TABLES(WFACTOR*BFACTOR) compare(BFACTOR).

*Syntax for post-hoc follow-up.
data list free file = 'C : \RM_ANCOVA_imps.dat'
  /imputation_ level2id Age Gender Baseline WFACTOR BFACTOR PRE_POST.
EXECUTE.

*creates IV group dummy codes; '(Mindful) Meditate' is the reference class.
compute Placebo = 0.
if (BFACTOR = 1) Placebo = 1.
compute Listen = 0.
if (BFACTOR = 2) Listen = 1.
```

```
compute Sham = 0.
if (BFACTOR = 4) Sham = 1.
EXECUTE.

*select posttest only.
USE ALL.
COMPUTE filter_$=(WFACTOR = 2).
VARIABLE LABELS filter_$'WFACTOR = 2 (FILTER)'.
VALUE LABELS filter_$0 'Not Selected' 1 'Selected'.
FORMATS filter_$(f1.0).
FILTER BY filter_$.
EXECUTE.

*activate pooling algorithm.
sort cases by imputation_.
split file layered by imputation_.
*Post-hoc.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT PRE_POST
  /METHOD=ENTER Placebo Listen Sham.
```

Stata Syntax: RM-ANCOVA Estimation

```
/*import data and recode missing data*/
use "C : \RM_ANCOVA_Stata_imps.dta"
recode _all (-99 = .)

/*define imported data as multiply imputed*/
mi import flong, m(imputation) id(level2id wfactor) imputed(age gender baseline pre_post) clear

/*estimate the model*/
mi estimate: mixed pre_post c.wfactor##c.bfactor c.baseline c.age gender || level2id:, reml var

/*post-hoc follow-up, reference group = group 3 */
mi estimate: mixed pre_post wfactor##ib3.bfactor c.baseline c.age gender || level2id:, reml var
```

R Syntax: RM-ANCOVA Estimation

```
install.packages("lme4")
install.packages("mitml")
library("lme4")
library("mitml")

imps<-read.table(file = "C:/RM_ANCOVA_imps.dat")

names(imps) <-
c("imputation","level2id","Age","Gender","Baseline","WFACTOR","BFACTOR","PRE_POST")

implist <- mitml::as.mitml.list(split(imps, imps$imputation))

fit<-with(implist, lm(PRE_POST ~ WFACTOR + BFACTOR + WFACTOR*BFACTOR + Baseline + Age + Gender))
```

(*Note:* The above model can be analyzed in *R* using either "lmer" with "LevelID" as a clustering variable "+ [1 | Level2ID]" or as specified above with" lm" and no clustering variable. The two sets of results are an exact match.)

```
mitml::testEstimates(fit, extra.pars = TRUE, df.com = 73)

*Syntax for post-hoc follow-up.
install.packages("fastDummies")
```

```
library("fastDummies")

data<-imps
data<-dummy_cols(data, select_columns="BFACTOR")

data <- subset(data, WFACTOR == 2)

datalist <- mitml::as.mitml.list(split(data, data$imputation))
fit <- with(datalist,
    lm(PRE_POST ~ 1 + BFACTOR_1 + BFACTOR_2 + BFACTOR_4)
    )

testEstimates(fit)
```

Alternate R Syntax: RM-ANCOVA Estimation

(This syntax uses the newly developed 'rblimp' package. The most current versions of *R* and *BLIMP* must be installed prior to using the syntax below. More information can be found in Keller & Enders, 2023, p. 12.)

```
install.packages('remotes')
remotes::install_github('blimp-stats/rblimp')
library(rblimp)

data1 <- as.data.frame(read.table('E:/RM_ANCOVA.dat', na.strings = '-99.00'))
colnames(data1) <-
    c('level2id','Age','Gender','Baseline','WITHIN','BETWEEN','PRE_POST')
RMANCOVA <- rblimp(
    data = data1,
    clusterid = 'level2id',
    ordinal = 'Gender WITHIN BETWEEN',
    fixed = 'Age Gender Baseline WITHIN BETWEEN',
    model = 'PRE_POST ~ Age Gender Baseline WITHIN BETWEEN WITHIN*BETWEEN | 1@0',
    seed = 90291,
    burn = 5000,
    iter = 10000)

output(RMANCOVA)
```

*Syntax for post-hoc follow-up.

```
library("fastDummies")
data1 <- dummy_cols(data1, select_columns="BETWEEN")
data1 <- subset(data1, WITHIN == 2)

POSTHOC <- rblimp(
    data = data1,
    ordinal = 'BETWEEN_1 BETWEEN_2 BETWEEN_4',
    fixed = 'BETWEEN_1 BETWEEN_2 BETWEEN_4',
    model = 'PRE_POST ~ 1 BETWEEN_1 BETWEEN_2 BETWEEN_4',
    seed = 75061,
    burn = 5000,
    iter = 10000)

output(POSTHOC)
```

Appendix B

Linking Epstein et al. (2022) data specifics to *BLIMP* diagnostic syntax

Many of the command specifications remain the same as in the previous example, but additional attention is needed for the GEE (`NOMINAL: & MODEL:`) specifications. In the previous RM-ANCOVA example, the `WITHIN` (pretest/posttest indicator) and `BETWEEN` (IV group indicator) variables were included on the `ORDINAL:` line because the main effects for `BETWEEN` and `WITHIN`, as well as the `BETWEEN * WITHIN` interaction effect, needed to be specified consistent with RM-ANCOVA logic. In this GEE example, the IV group indicator (`TX_group`) is also listed on the `ORDINAL:` line, but the `drive` variable (coded 0 for

the baseline assessment, 1 for the 1-month assessment, and 2 for the 6-month assessment) is treated as a multinomial variable and listed on the NOMINAL: line for two reasons. First, adding a (0) to the end of the nominal command specifies that the baseline assessment is the reference, and *BLIMP* will automatically create two dummy codes for the 1-month (drive.1) and 6-month (drive.2) assessments. Second, these dummy codes can then be included on the MODEL: line to specify both main effects (drive.1 & drive.2) and interactions (drive.1*TX_group & drive.2*TX_group) included in the imputation model. This both satisfies Graham's (2012) golden rule of imputation and ensures that research questions regarding FOCAL+ treatment efficacy (drive.1*TX_group) and maintenance of treatment gains (drive.2*TX_group) after controlling for covariates (MeanSped, age_lic) can both be answered in the GEE analysis.

BLIMP Diagnostic and Imputation Syntax Scripts for Missing Data

SPSS, Stata, and R syntax for GEE Estimation

BLIMP—Diagnostic

```
DATA: SDLP.dat;
VARIABLES: level2id drive TX_group SDLaneP MeanSped age_lic;
CLUSTERID: level2id;
MISSING: -99;
FIXED: drive TX_group age_lic;
ORDINAL: TX_group;
NOMINAL: drive(0);
MODEL:
SDLaneP ~ drive.1 drive.2 TX_group drive.1*TX_group drive.2*TX_group MeanSped age_lic | 1@0;
SEED: 75061;
BURN: 10000;
ITERATIONS: 10000;
OUTPUT: PSR;
```

BLIMP—Imputation

```
DATA: SDLP.dat;
VARIABLES: level2id drive TX_group SDLaneP MeanSped age_lic;
CLUSTERID: level2id;
MISSING: -99;
FIXED: drive TX_group age_lic;
ORDINAL: TX_group;
NOMINAL: drive(0);
MODEL:
SDLaneP ~ drive.1 drive.2 TX_group drive.1*TX_group drive.2*TX_group MeanSped age_lic | 1@0;
SEED: 75061;
BURN: 5000;
ITERATIONS: 10000;
CHAINS: 100;
NIMPS: 100;
SAVE: stacked = SDLP_imps.dat;
```

SPSS Syntax: GEE Estimation

```
data list free file = 'C : \SDLP_imps.dat'
  /imputation_ level2id drive drive_0 drive_1 drive_6 TX_group SDLaneP MeanSped age_lic.
EXECUTE.

*Grand-mean centering.
aggregate
  /outfile = * mode = addvariables
  /break = imputation_
  /m_MeanSped = mean(MeanSped)
  /m_age_lic = mean(age_lic).

compute c_MeanSped = MeanSped—m_MeanSped.
compute c_age_lic = age_lic—m_age_lic.
EXECUTE.

sort cases by imputation_.
```

```
split file layered by imputation_.

GENLIN SDLaneP BY TX_group drive (order = descending) WITH c_MeanSped c_age_lic
/MODEL drive TX_group drive*TX_group INTERCEPT=YES
DISTRIBUTION=NORMAL LINK=IDENTITY
/CRITERIA SCALE=MLE PCONVERGE = 1E-006 (ABSOLUTE) SINGULAR = 1E-012 ANALYSISTYPE = 3 (WALD)
CILEVEL = 95 LIKELIHOOD=FULL
/REPEATED SUBJECT=level2id WITHINSUBJECT=drive SORT=YES CORRTYPE= AR(1) ADJUSTCORR=YES
COVB=ROBUST
MAXITERATIONS = 100 PCONVERGE = 1e-006 (ABSOLUTE) UPDATECORR = 1
/PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION.
```

Stata Syntax: GEE Estimation

```
/*import data and recode missing data*/
use "C : \GEE_Stata_imps.dta"
recode _all (-99 = .)

/*define imported data as multiply imputed*/
mi import flong, m(imputation) id(level2id drive) imputed(sdlanep meansped age_lic) clear

/*define the hierarchical structure of the data*/
mi xtset level2id drive

/*estimate the model*/
mi estimate: xtgee sdlanep c.tx_group##drive meansped age_lic, family(gaussian) corr(ar1)
```

R Syntax: GEE Estimation

```
install.packages("geepack")
library("geepack")
install.packages("mitml")
library("mitml")

imps2<-read.table(file = "C:/SDLP_imps.dat")
names(imps2) <- c("imputation","level2id","drive","drive_0","drive_1","drive_6","TX_group","
SDLaneP", "MeanSped", "age_lic")

imps2<-imps2[order(imps2$imputation, imps2$level2id, imps2$drive),]

imps2 <- mitml::as.mitml.list(split(imps2, imps2$imputation))

gee.sdlp<-with(imps2, geeglm(SDLaneP ~ TX_group*factor(drive) + MeanSped + age_lic, id = lev-
el2id, family = gaussian, corstr = "ar1"))

vcov.geeglm<-function(x) summary(x)$cov.scaled

testEstimates(gee.sdlp)
```

Appendix C

Linking Kashikar-Zuck et al. (2012) data specifics to *BLIMP* diagnostic syntax

The syntax needed to impute missing data from the Kashikar-Zuck et al. (2012) trial differs from previous syntax specifications in several ways. First, and because depression is needed to answer the second research question, CDI_T scores are group mean centered (`GROUPMEAN`) using the (`CENTERING:`) command (see Enders & Tofighi, 2007) (tender point examination scores [`TENDERPT`] are not centered because they are included only as a control covariate and are not of theoretical interest). Second, and more importantly, the `MODEL:` command now contains `Focal.model:` and `Predictor.model:` sections. The `Focal.model:` contains the data analysis model needed to answer both research questions. As shown, and after including dummy coded sites (`SITE2` & `SITE3`) and tender point examination scores (`TENDERPT`) as control covariates, main effects, all two-way interaction effects, and a three-way interaction are all included for random assignment (`IV_GROUP`), assessment time point (`TIME`, coded in months as 0, 2, and 6), and depression (`CDI_T`) scores.

A `Predictor.model:` section is also needed because a priori testing results from both Kolmogorov–Smirnov (KS: statistic = 0.08; $p < .001$) and Shapiro-Wilk (SW: statistic = .971; $p < .001$) tests showed CDI_T scores were significantly nonnormally distributed (KS and SW tests for tender point examination scores [`TENDERPT`] were not significant). Shown in the second line below the `Predictor.model:` is the use of the Yeo and Johnson (2000) transform (`yjt`) procedure. Readers interested in additional details can consult several sources (Enders, 2022; Keller & Enders, 2023, pp. 47, 65, 120–125; Lüdtke et al., 2020; Yeo & Johnson, 2000), but the Yeo-Johnson transform procedure can be described briefly here. As implemented in *BLIMP*, the (`yjt`) procedure proceeds from the assumption that plausible imputed values for missing data should not be drawn from a normal posterior distribution if the original variable with missing values is nonnormally distributed.

The (`yjt`) procedure can be summarized in six steps: (1) *BLIMP* estimates parameters consistent with the `MODEL:` command from a posterior distribution conditional on FDI and non-normal CDI_T scores, (2) plausible imputed values for missing FDI scores are obtained conditional on the parameters estimated in step 1 and CDI_T scores, (3) *BLIMP* estimates a "shape" parameter that essentially quantifies the extent to which CDI_T scores deviate from a normal distribution conditional on current CDI_T scores and parameter estimates, (4) CDI_T scores are transformed based on the "shape" parameter estimated in step 3 resulting in normally distributed CDI_T scores, (5) *BLIMP* draws plausible imputed values from a posterior distribution assuming normally distributed CDI_T scores conditional on steps 1–4, and (6) CDI_T observed and imputed scores are then back transformed using the "shape" parameter estimated in step 3 and the inverse of the (`yjt`) procedure to place values back on their original non-normal distribution (B. Keller, personal communication, October 16, 2023).

Recall that having a nonnormally distributed continuous predictor variable with missing values is a condition that calls for factored regression or sequential specification rather than the default "all-in-turn" specification for predictor variables with missing values. As shown in the first line below the `Predictor.model:` line, tender point examination scores are predicted by the four variables with complete data (`IV_GROUP, SITE2, SITE3, & TIME`). On the second line below the `Predictor.model:` line, and consistent with factored regression/sequential specification, Yeo-Johnson CDI_T scores are predicted by the four variables with complete data plus tender point examination scores. To clarify what might at first glance appear to be redundant centering command lines, the centering specified with the (`yjt`) procedure (`CDI_T - 50`) is needed to assist in imputation model convergence (Keller & Enders, 2023, pp. 120–125). The `CENTERING: GROUPMEAN =` command is needed for accurate Bayesian parameter estimates (described further in the Summary section of the article).

BLIMP Diagnostic and Imputation Syntax Scripts for Missing Data

SPSS, Stata, and R syntax for LLMM Estimation

BLIMP—Diagnostic

```
DATA: LLMM.dat;
VARIABLES: ID TIME FDI IV_GROUP SITE2 SITE3
CDI_T TENDERPT;
CLUSTERID: ID;
ORDINAL: TIME IV_GROUP SITE2 SITE3;
MISSING: -999;
FIXED: TIME IV_GROUP SITE2 SITE3;
CENTERING: GROUPMEAN = CDI_T;
MODEL:
Focal.model:
FDI ~ TIME IV_GROUP SITE2 SITE3 TENDERPT CDI_T
TIME*IV_GROUP TIME*CDI_T IV_GROUP*CDI_T
TIME*IV_GROUP*CDI_T | TIME;
Predictor.model:
TENDERPT ~ IV_GROUP SITE2 SITE3 TIME;
yjt(CDI_T - 50) ~ TENDERPT IV_GROUP SITE2 SITE3 TIME;
SEED: 45103;
BURN: 10000;
ITERATIONS: 10000;
OUTPUT: PSR;
```

BLIMP-Imputation

```
DATA: LLMM.dat;
VARIABLES: ID TIME FDI IV_GROUP SITE2 SITE3
CDI_T TENDERPT;
CLUSTERID: ID;
ORDINAL: TIME IV_GROUP SITE2 SITE3;
MISSING: -999;
FIXED: TIME IV_GROUP SITE2 SITE3;
CENTERING: GROUPMEAN = CDI_T;
```

```
MODEL:
Focal.model:
FDI ~ TIME IV_GROUP SITE2 SITE3 TENDERPT CDI_T
TIME*IV_GROUP TIME*CDI_T IV_GROUP*CDI_T
TIME*IV_GROUP*CDI_T | TIME;
Predictor.model:
TENDERPT ~ IV_GROUP SITE2 SITE3 TIME;
yjt(CDI_T - 50) ~ TENDERPT IV_GROUP SITE2 SITE3 TIME;
SEED: 45103;
BURN: 5000;
ITERATIONS: 10000;
CHAINS: 100;
NIMPS: 100;
SAVE: stacked = LLMM_imputed.dat;
```

SPSS Syntax: LLMM Estimation

```
data list free file = 'C : \LLMM_Imputed.dat'
  /imputation_ ID TIME AGE FDI IV_GROUP SITE2 SITE3 MIGRAINE CDI_T EXTERNAL TENDERPT.
EXECUTE.
*Group-mean centering.
aggregate
  /outfile = * mode = addvariables
  /break = imputation_ ID
  /m_CDI_T = mean(CDI_T).

compute CWC_CDI_T = CDI_T—m_CDI_T.
EXECUTE.

sort cases by imputation_.
split file layered by imputation_.

MIXED FDI WITH IV_GROUP SITE2 SITE3 TIME TENDERPT CWC_CDI_T
/PRINT = SOLUTION TESTCOV
/METHOD = ML
/FIXED = INTERCEPT TIME IV_GROUP SITE2 SITE3 TENDERPT CWC_CDI_T
TIME*IV_GROUP TIME*CWC_CDI_T IV_GROUP*CWC_CDI_T
TIME*IV_GROUP*CWC_CDI_T
/RANDOM INTERCEPT TIME | SUBJECT(ID) COVTYPE(VC).
```

Stata Syntax: LLMM Estimation

```
/*import the data and recode missing data*/
use "C : \LLMM_Imputed_Stata.dta"
recode _all (-999 = .)

/*define imported data as imputed */
mi import flong, m(imputation) id(id time) imputed(age fdi migraine cdi_t external tenderpt) clear

/*define longitudinal data structure*/
mi xtset id time

/*group mean centering*/
mi passive: egen means_cdit = mean(cdi_t), by(id)
mi passive: gen cwc_cdit = cdi_t—means_cdit

/*estimate model with a random linear slope for time */
mi estimate: mixed fdi time iv_group site2 site3 tenderpt cwc_cdit c.time#iv_group c.time#c.
cwc_cdit iv_grou#c.cwc_cdit c.time#iv_group#c.cwc_cdit || id: time
```

R Syntax: LLMM Estimation

```
oo <- options(repos = "https://cran.r-project.org/")
install.packages("Matrix")
install.packages("lme4")
options(oo)

install.packages("lme4", type = "source")
install.packages("mitml")
library("lme4")
library("mitml")

LLMM <- read.table("C:/Users/peuu3c/Desktop/LLMM_Imputed.dat")
names(LLMM) <-
    c("Imputation","ID","TIME","AGE","FDI","IV_GROUP","SITE2","SITE3", "MIGRAINE", "CDI_T",
    "EXTERNAL", "TENDERPT")

implist <- mitml::as.mitml.list(split(LLMM, LLMM$Imputation))

new.implist <- within(implist, Means.CDI_T <- clusterMeans(CDI_T, ID))

new.implist2 <- within(new.implist, {cwc.CDI_T <- CDI_T-Means.CDI_T})

llmmod <- "FDI ~ TIME + IV_GROUP + SITE2 + SITE3 + TENDERPT + cwc. CDI_T + TIME*IV_GROUP + TIME*cwc.
     CDI_T + IV_GROUP*cwc.CDI_T + TIME*IV_GROUP*cwc.CDI_T + (1 + TIME | ID)"

#Clusters (N) = 114 - (Predictors) 10-1
ddf <- 103

result <- with(new.implist2, lme4::lmer(llmmod, REML = T))

mitml::testEstimates(result, extra.pars = T, df.com = ddf)
```

Alternate R Syntax: LLMM Estimation

(This syntax uses the newly developed 'rblimp' package. The most current versions of *R* and *BLIMP* must be installed prior to using the syntax below. More information can be found in Keller & Enders, 2023, p. 12.)

```
library(rblimp)
data3 <- as.data.frame(read.table('E:/LLMM_Test1.dat', na.strings = '-999.00'))

colnames(data3) <-
    c('ID','TIME','AGE','FDI','IV_GROUP','SITE2','SITE3','MIGRAINE','CDI_T','EXTERNAL',
    'TENDERPT')

LLMM <- rblimp(
    data = data3,
    clusterid = 'ID',
    ordinal = 'TIME IV_GROUP SITE2 SITE3 MIGRAINE',
    fixed = 'TIME IV_GROUP SITE2 SITE3',
    center = 'GROUPMEAN = CDI_T',
    model = 'Focal.model:
    FDI ~ TIME IV_GROUP SITE2 SITE3 TENDERPT CDI_T
    TIME*IV_GROUP TIME*CDI_T IV_GROUP*CDI_T
    TIME*IV_GROUP*CDI_T | TIME;
    Predictor.model:
    TENDERPT ~ IV_GROUP SITE2 SITE3 TIME;
    yjt(CDI_T - 50) ~ TENDERPT IV_GROUP SITE2 SITE3 TIME',
    seed = 45103,
    burn = 5000,
    iter = 10000)

output(LLMM)
```