

Missing Data: An Update on the State of the Art

Craig K. Enders

Department of Psychology, University of California Los Angeles

The year 2022 is the 20th anniversary of Joseph Schafer and John Graham's paper titled "Missing data: Our view of the state of the art," currently the most highly cited paper in the history of *Psychological Methods*. Much has changed since 2002, as missing data methodologies have continually evolved and improved; the range of applications that are possible with modern missing data techniques has increased dramatically, and software options are light years ahead of where they were. This article provides an update on the state of the art that catalogs important innovations from the past two decades of missing data research. The paper addresses topics described in the original paper, including developments related to missing data theory, full information maximum likelihood, Bayesian estimation, multiple imputation, and models for missing not at random processes. The paper also describes newer factored regression specifications and missing data handling for multilevel models, both of which have been a focus of recent research. The paper concludes with a summary of the current software landscape and a discussion of several practical issues.

Translational Abstract

Missing data is a common problem in behavioral science research. In 2002, a highly influential paper was published in *Psychological Methods* that detailed the "state of the art" methods for dealing with this issue at the time. Much has changed since then, as missing data methodologies have continually evolved and improved; the range of applications that are now possible has increased dramatically, and software options are far ahead of where they were. This article provides an update on the state of the art that catalogs important innovations from the last two decades of missing data research.

Keywords: missing data, full information maximum likelihood, Bayesian estimation, multiple imputation

This year is the 20th anniversary of Joseph Schafer and John Graham's highly influential paper in *Psychological Methods* titled "Missing data: Our view of the state of the art" (Schafer & Graham, 2002). In her 2017 retrospective, "The making of *Psychological Methods*," former Editor Lisa Harlow reported that the paper was the second most highly cited publication in the journal between 1995 through 2015; the paper had 3,428 citations, roughly 200 fewer than MacKinnon et al. (2002), which had 3,618. Cut to the present, and the article is the most highly cited publication in *Psychological Methods* by a wide margin, with 5,307 references.¹ The paper's continual upward trajectory speaks to its high quality and continued relevance to behavioral science research.

Schafer and Graham's paper (henceforth referred to as SG2002) was published at an inflection point in methodological history when "modern" missing data methods such as full information maximum likelihood and multiple imputation were becoming a practical reality because of an uptick in software options. In both frameworks,

models for multivariate normal missing data were predominant at the time, and flexible approaches for a broader range of applications were still a ways off. Perhaps not surprisingly, much has changed since 2002, as missing data methodologies have continually evolved and improved; the range of applications that are possible with modern missing data techniques has increased dramatically, and software options are light years ahead of where they were. Some of these developments have permeated the collective research consciousness, while recent work may be unfamiliar to many readers. The purpose of this article is to provide an update on the state of the art that catalogs important innovations from the last two decades of research, with the goal of providing methodologists and practicing researchers with a springboard for accessing the most up-to-date missing data handling methodologies. My hope for this paper is to celebrate and supplement a landmark publication that is still a must-read 20 years later.

The structure of the paper is as follows. I begin with missing data theory, detailing extensions and clarifications of Rubin's (Little & Rubin, 2020; Rubin, 1976) classic taxonomy. Next, I describe factored regression specifications that express a multivariate distribution as a sequence of simpler univariate functions. This emerging modeling framework subsumes several recent innovations, and it

Craig K. Enders  <https://orcid.org/0000-0002-7048-8369>

This work was supported by the Institute of Educational Sciences Award R305D22000.

There are no conflicts of interest to report. The ideas in this paper have not been presented elsewhere.

Correspondence concerning this article should be addressed to Craig K. Enders, Department of Psychology, UCLA, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095, United States. Email: cenders@ucla.edu

¹ According to APA's PsycNET website, Schafer and Graham (2002) have 5,307 citations, followed by MacKinnon et al. (2002) with 4,752. The search was conducted on February 16, 2022. Citation counts from other sources may differ.

addresses longstanding limitations of multivariate normal missing data methods. The next three sections of the paper outline full information maximum likelihood estimation, Bayesian estimation, and multiple imputation. Collectively, these topics comprise the majority of the content from SG2002, and I describe how these frameworks have evolved in the past two decades. The next two sections revisit missing not at random analyses and composite scores with item-level missing data, and the final section describes methods for multilevel missing data. The paper concludes with a summary of the current software landscape.

Missing Data Mechanisms

Rubin and colleagues (Little & Rubin, 1987; Rubin, 1976) introduced a theoretical system for missing data problems that describes three ways in which missingness can relate to the realized data. A core part of the theory is the presence of a model that describes the occurrence of missing values, encoded by a missing data indicator M ($0 = \text{score is observed}$, $1 = \text{score is missing}$). The missing completely at random mechanism states that missingness is unrelated to both the observed and unseen parts of the data, a missing at random process posits that missingness is related to the observed parts only, and a missing not at random mechanism states that the probability of missing values relates to the missing parts of the data. Schafer and Graham (2002) provide an excellent summary of Rubin's theory, so I focus on new contributions to the framework. For a critique of Rubin's theory and its underlying assumptions, readers can refer to various works by Manski (2005, 2016, 2022).

To begin, methodologists have proposed useful modifications to the terminology itself. For example, Graham (2009) suggested the phrase *conditionally missing at random* instead of missing at random, as it better conveys that missingness is haphazard after conditioning on the observed data. The current edition of Little and Rubin's (2020) classic missing data text similarly replaces the *not missing at random* phrase from earlier editions with *missing not at random*, arguing that this change improves clarity. Gomer and Yuan's (2021) recent article in *Psychological Methods* further elaborates on this mechanism, describing two distinct missing not at random subtypes; a *focused* missing not at random process occurs when an incomplete variable fully determines its own missingness, and a *diffuse* missing not at random process occurs when additional variables uniquely influence missing data.

A subtle feature of Rubin's definitions is that they describe the conditional distribution of the *realized* missing data patterns given the *realized* sample data. This point, which Schafer and Graham (2002, p. 151) highlight in a footnote, has implications for inference that are not broadly appreciated. S. Seaman et al. (2013) clarified this issue, explaining that frequentist inference based on repeated sampling (e.g., most maximum likelihood applications) requires broader conditions that apply to other missingness patterns and realizations of the data. The essence of this work is that valid inference requires the missing data process to replicate across different random samples. They refer to Rubin's original definition as *realized missing at random* (or *realized missing completely at random*), and they call the requirements for frequentist inference *everywhere missing at random* (or *everywhere missing completely at random*). In related work, Mealli and Rubin (2016) define similar conditions called *missing always completely at random*, *missing always at random*, and *missing always not at random*.² Although SG2002 predates

these expanded definitions, the paper does emphasize a related issue that frequentist standard errors should condition on the observed missing data patterns—a point originally highlighted by Kenward and Molenberghs (1998) and later extended to robust inference by Savalei (2010a).

Another interesting development related to Rubin's framework is the use of graphical models—directed acyclic graphs and missingness graphs (m-graphs)—to represent and understand missing data mechanisms. To illustrate, the m-graphs in Figure 1 depict the focused and diffuse missing not at random processes described by (Gomer & Yuan, 2021). Schafer and Graham (2002) used a similar diagram to represent Rubin's mechanisms, and others have since developed the framework as tool for understanding and clarifying the conditions under which valid estimates are possible (R. M. Daniel et al., 2012; Mohan et al., 2013; Mohan & Pearl, 2021; Moreno-Betancur et al., 2018; Pearl & Mohan, 2013; Thoemmes & Mohan, 2015; Thoemmes & Rose, 2014). For example, Thoemmes and Rose (2014) used graphical models to illustrate the phenomenon of bias-inducing auxiliary variables, and R. M. Daniel et al. (2012) used directed acyclic graphs to determine the conditions under which unbiased estimates of causal effects can be obtained from a complete-case analysis.

A recent paper by Mohan and Pearl (2021) is perhaps the most comprehensive treatment of graphical methods to date. The authors argue that graphical models can answer questions that statistical models cannot, and they discuss the use of graphs for classifying the missing data mechanism for a set of variables and identifying conditional independence propositions that are testable from the data. The paper also builds on the authors' earlier work (Mohan et al., 2013; Pearl & Mohan, 2013), demonstrating the use of graphs to determine whether a parameter of interest is *recoverable* from the observed part of an incomplete data set (i.e., whether a consistent estimator exists for a particular mechanism and set of variables). Translating this work into prescriptions for practicing behavioral science researchers is a fruitful avenue for future work.

Factored Regression Specifications

Missing data techniques based on factored regression specifications arguably top the list of innovations from the past 20 years. Factored regression is an overarching modeling strategy that can be married with maximum likelihood, Bayesian estimation, or multiple imputation. Ibrahim and colleagues provide the theory behind this approach (Huang et al., 2005; Ibrahim, 1990, 1999, 2002; Lipsitz & Ibrahim, 1996), and SG2002 briefly referenced some of this early work in the context of weighting. This modeling strategy has received considerable attention in the recent literature, as it solves several long-standing problems with multivariate normal missing data methods.

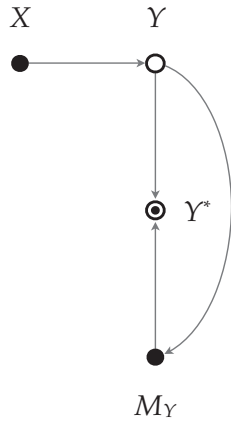
The classic missing data methods that predominated SG2002's era invoke a joint distribution for the incomplete variables, typically the multivariate normal. The parameters of this distribution—a mean vector and covariance matrix—characterize properties of the unseen score values. In contrast, a factored regression specification acknowledges that a multivariate distribution exists, but it does not

² Mealli and Rubin (2016) note that Rubin (1976) eludes to the missing always at random definition in an example on page 584.

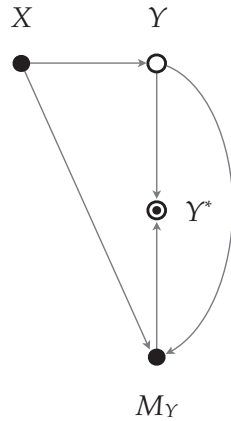
Figure 1

M-Graphs Depicting Focused and Diffuse Missing Not at Random (MNAR) Processes

(a) Focused MNAR



(b) Diffuse MNAR



Note. M-graphs that depict missing not at random processes involving a complete variable, X , an incomplete variable, Y , and a binary missing data indicator, M_Y , coded 0 if Y is observed and 1 if it is missing. The white circle labeled Y represents the hypothetically complete variable (i.e., the combination of the observed and missing data), the circle labeled Y^* represents realized values of Y (i.e., Y^* equals Y when the missing data indicator M_Y equals 0 and is missing whenever M_Y equals 1).

require a researcher to specify or make assumptions about its form. Rather, the procedure uses the probability chain rule to factorize the joint distribution into a sequence of univariate conditional distributions, each of which corresponds to a regression model. In this framework, intercepts, slopes, and residual variances define the distributions of missing values.

To illustrate a factored specification, consider a multiple regression analysis with a pair of incomplete predictors. Using generic function notation, the multivariate distribution of the three variables is $f(Y, X_1, X_2)$. The idea behind a factored specification is to recast the joint distribution as the product of two or more univariate distributions, each of which aligns with a regression model. For example, applying the probability chain rule to the trivariate distribution gives the following expression

$$f(Y, X_1, X_2) = f(Y|X_1, X_2) \times f(X_1|X_2) \times f(X_2) \quad (1)$$

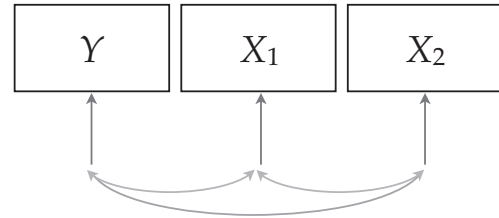
where variables to the left of a vertical pipe function as outcomes, and variables to the right of a pipe are their predictors. Importantly, the $f(Y|X_1, X_2)$ term corresponds to the conditional distribution of Y given the X s—the distribution induced by the focal regression model. The remaining terms are “nuisance models” that describe the conditional distributions of the predictors.³

There are usually several ways to factorize the joint distribution, and these factorizations are not necessarily equivalent. For example, the nuisance models could instead be specified as the product of $f(X_2|X_1)$ and $f(X_1)$. Figure 2 illustrates this distinction with path model diagrams. The key is that one of the terms in the sequence represents the model the researcher would have fit, were the data complete. Practical recommendations for ordering a sequence of factored

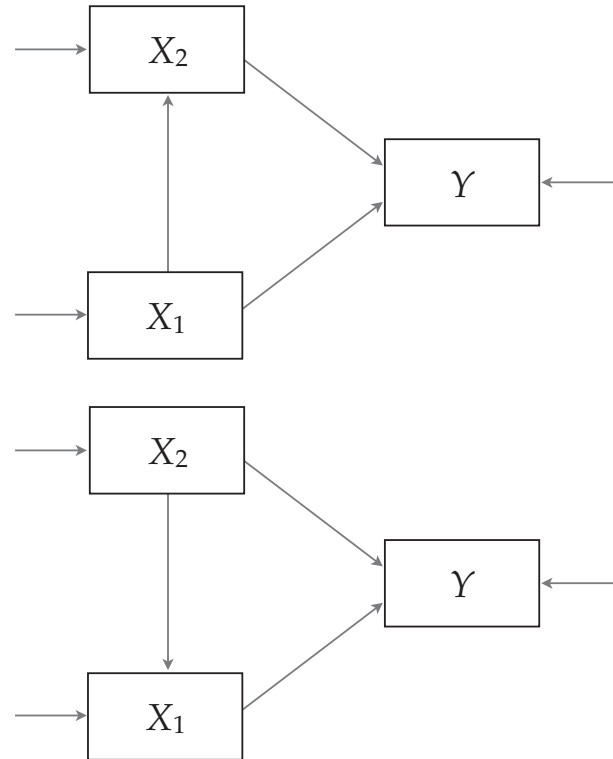
Figure 2

Path Diagrams Depicting a Joint (Multivariate) Distribution and Two Factored Regression Specifications

(a) Joint (multivariate) distribution



(b) Factored regression specifications



Note. Path diagrams depicting a trivariate joint distribution and two factored regression specifications. Double-headed curved arrows represent generic associations, and directed arrows denote regression slopes.

³ Another possible factorization assigns a multivariate normal distribution to the predictors: $f(Y, X_1, X_2) = f(Y|X_1, X_2) \times f(X_1, X_2)$. This specification accommodates binary, ordinal, and multicategorical predictors via the latent response framework, and it is strictly limited to linear associations among the regressors Keller and Enders (2021).

regression models are available in the literature (Lüdtke et al., 2020b; D. Xu et al., 2016).

Factored regression specifications offer compelling advantages over multivariate missing data models. First, variables need not have the same metrics. For example, $f(Y|X_1, X_2)$ could feature a count variable as the outcome, $f(X_1|X_2)$ could be a linear regression with normal errors, and $f(X_2)$ could be a binomial or multinomial distribution induced by an empty logistic or probit model. Incomplete variables can even be continuous and nonnormal (Lüdtke et al., 2020b). A second advantage of factored specifications is that any of the univariate models may include nonlinear terms like interactions, polynomials, and random slopes, among others. For example, the factorization below depicts a scenario where the focal regression features the product of X_1 and X_2 and the first nuisance model expresses X_1 as a quadratic function of X_2 .

$$f(Y, X_1, X_2) = f(Y|X_1, X_2, X_1 \times X_2) \times f(X_1|X_2, X_2^2) \times f(X_2) \quad (2)$$

Such effects are statistically incompatible with a multivariate normal missing data model and are known to introduce bias (J. W. Bartlett et al., 2015; J. C. Liu et al., 2014).

In the factored regression framework, the distributions of missing values depend on every model in which an incomplete variable appears. Returning to Equation 1, the focal regression solely determines the distribution of missing dependent variable scores because $f(Y|X_1, X_2)$ is the only function that includes Y . In contrast, X_1 is a predictor in the focal regression and an outcome in the model linking it to X_2 . Applying Bayes' theorem to Equation 1 tells us that the conditional distribution of missing X_1 values is proportional to the product of two univariate distributions.

$$f(X_1|Y, X_2) \propto f(Y|X_1, X_2) \times f(X_1|X_2) \quad (3)$$

Conceptually, the equation says that the distribution of X_1 given the other analysis variables is a composite function that depends on two sets of regression model parameters. Similarly, the distribution of missing X_2 scores is proportional to the product of three univariate distributions because this variable appears in every term on the right side of Equation 1. Consistent with their classic counterparts, maximum likelihood estimators for factored regression specifications produce estimates that average over these distributions, whereas Bayesian estimation and multiple imputation sample replacement scores from them. I revisit this modeling framework throughout the remainder of the paper.

Maximum Likelihood Estimation

The origins of maximum likelihood missing data handling are quite old and date back to the 1950s (Anderson, 1957; Edgett, 1956; Hartley, 1958; Lord, 1955). Many important breakthroughs came in the 1970s when methodologists developed the theoretical underpinnings of modern missing data handling techniques as well as computational methods to implement them (Beale & Little, 1975; Dempster et al., 1977; Finkbeiner, 1979; Hartley & Hocking, 1971; Orchard & Woodbury, 1972). For researchers in the social and behavioral sciences, maximum likelihood estimation became a practical reality in the 1990s when structural equation modeling software packages began implementing full information

missing data estimators based on raw rather than summary data (Arbuckle, 1996). This classic estimator for multivariate normal data was the primary tool for implementing maximum likelihood when Schafer and Graham (2002) appeared in print.

The classic maximum likelihood estimator uses an iterative optimization algorithm to identify model parameters that minimize the sum of squared standardized distances between a model's predictions and the observed data. The observed-data log-likelihood function that encodes data-model fit is

$$\ell_{(\text{obs})} = -\sum_{i=1}^N \frac{V_i}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln|\Sigma_i| - \frac{1}{2} \sum_{i=1}^N (Y_i - \mu_i)' \Sigma_i^{-1} (Y_i - \mu_i) \quad (4)$$

where Y_i is a vector that contains an individual's observed data, V_i is the number of scores in that vector, and μ_i and Σ_i contain the subset of mean and variance-covariance matrix elements that correspond to the observed variables in Y_i (more generally, the elements in μ and Σ can be functions of another model's parameters). The log-likelihood equation assumes that all participants share the same model parameters, but each observation's contribution to estimation is restricted to the subset of parameters for which there is data. Although the estimator does not fill in missing values, the multivariate normal log-likelihood function functions like an imputation machine in the sense that the estimator can infer the location of unseen data points from the observed scores and adjust the model parameters accordingly.

Auxiliary Variable Methods

A great deal of ensuing methodological research focused on extending normal-theory estimation in structural equation models and understanding its limitations. One such extension involves methods for introducing auxiliary variables and correlates of missingness. The long-standing advice from the multiple imputation literature is to include as many extra variables as possible when treating missing values because doing so minimizes the risk of nonresponse bias (Collins et al., 2001; Rubin, 1996).⁴ However, this prescription is not straightforward to implement with maximum likelihood estimation, which tailors missing data handling to a specific analysis model. Methodologists have developed structural equation modeling specifications that address this issue.

Graham (2003) described two specifications—the saturated correlates and extra dependent variable models—that use residual correlations and directed pathways to introduce auxiliary variables in a way that does not affect the meaning of the focal model's parameters. The saturated correlates approach, which is automated in some structural equation modeling software programs,⁵ is known to produce convergence failures because it introduces an awkward and potentially illogical structure on the residual covariance matrix (Savalei & Bentler, 2009). Possible solutions include adopting a surgical approach that identifies a small number of extra variables that

⁴ Thoemmes and Rose (2014) describe an exception where an auxiliary variable can enhance rather than mitigate nonresponse bias.

⁵ The saturated correlates specification is automated in *Mplus* (L. K. Muthén and B. O. Muthén 1998–2017), *EQS* (Bentler 2000–2008), and the *semTools* package in R (Jorgensen, 2021). Two-stage estimation is available in *EQS* and *semTools*.

correlate with the focal model's residuals (Raykov & West, 2016), or conducting a preliminary data reduction stage that reduces a large set of auxiliary variables into one or two principal components (Howard et al., 2015).

Two-stage estimation is an altogether different strategy that first estimates the mean vector and variance–covariance matrix of a superset of variables beyond those from the focal analysis. The second step uses the “filled-in” summary statistics as input for a complete-data structural equation modeling analysis (Savalei & Bentler, 2009; Yuan & Bentler, 2000), and it uses a sandwich correction to account for differential precision across elements of the first-stage summary statistics. Robust variants of two-stage estimation accommodate nonnormal data (Savalei & Falk, 2014; Yuan et al., 2014; Yuan & Zhang, 2012), and an extension of the procedure is available for composites with missing item responses (Savalei & Rhemtulla, 2017). Graham's specifications and two-stage estimation are both limited to multivariate normal data.

Robustness and Nonnormality

A second important thread of research involves the robustness of maximum likelihood estimation to nonnormal data. Although robust (sandwich estimator) standard errors and test statistics had been developed and evaluated by 2002 (Arminger & Sobel, 1990; Enders, 2001; Yuan & Bentler, 2000), this topic continued to be an active area of research, especially in the context of structural equation modeling. Analytic and simulation work clarified that maximum likelihood estimates are consistent in a broad range of applications, although sandwich corrections and rescaled test statistics are usually preferred for inference (Savalei, 2010a, 2010b; Takai & Kano, 2013; Yuan, 2009; Yuan & Bentler, 2010; Yuan et al., 2012; Yuan & Savalei, 2014). However, missing data can distort approximate fit indices such as the CFI and RMSEA, even when the data are normal (Lai, 2021; X. Zhang & Savalei, 2020). Interested readers can find a comprehensive summary of these technical innovations in Savalei and Rosseel (2022). Bootstrap standard errors and test statistics are an alternative corrective procedure for nonnormal missing data (Enders, 2002; Savalei & Yuan, 2009).

One situation where correctives for nonnormal data may or may not be useful occurs in models that feature a mix of categorical and continuous metrics. The complete-data literature suggests that treating ordinal outcomes as normal is not problematic if the discrete distribution is symmetric and has five or more scale points (Rhemtulla et al., 2012), and this conclusion likely applies to missing data as well. Computer simulations suggest that treating incomplete binary predictors as normal usually does not introduce bias in single-level regression models (B. O. Muthén & L. Muthén, 2016), although it may in multilevel models with discrete Level-2 predictors (Grund et al., 2018). Treating multicategorical nominal predictors as normal is nonsensical.

Factored Regression Specifications

The development of maximum likelihood estimators for factored regression specifications is an important recent innovation. As explained previously, this framework disassembles a model into multiple parts that potentially rely on different probability

distributions and log-likelihood functions (see Equation 1). Ibrahim and colleagues developed the theory and optimization algorithms behind estimators for missing at random and missing not at random processes (Ibrahim, 1990; Ibrahim et al., 1999), and SG2002 mentions this work in the context of weighting. Lüdtke et al. (2020a) provide an accessible and modern account of this work along with an R package that applies Ibrahim's estimator (Grund, Lüdtke, & Robitzsch, 2021b).

As mentioned previously, factored specifications are compelling because they naturally accommodate variables with different metrics. Maximum likelihood estimators have evolved considerably since 2002, and flexible routines that factorizations to accommodate mixtures of categorical and continuous variables are now widely available (Lüdtke et al., 2020a; B. O. Muthén & L. Muthén, 2016; Pritikin et al., 2018; Rabe-Hesketh et al., 2004; Rockwood & Jeon, 2019). However, this functionality varies across software packages, and not all combinations of metrics are currently available; support for binary and ordinal variables is common, but missing data handling for other variable types is more limited.

Disassembling a model into multiple parts that leverage potentially different probability distributions complicates missing data handling considerably. Ibrahim's (1990) seminal work describes an iterative optimization procedure known as numerical integration that fills in the missing parts of the data in an imputation-esque manner. This scheme replaces each missing data point with a fixed grid of values (i.e., nodes) that span the incomplete variable's entire range, and it weights each of these pseudo-imputations according to its likelihood given a person's observed data. After computing weights that condition on the current parameter estimates, the algorithm uses standard weighted least squares solutions to update parameter estimates that average over the grid of missing values (see Lüdtke et al., 2020a). Modern variations of this approach adaptively select the nodes (Rabe-Hesketh et al., 2005), and a related Monte Carlo integration algorithm invokes a similar idea with simulated rather than fixed support nodes. Relative to their normal-theory predecessors, estimators for mixed variable types are computationally intensive and limited in the number of discrete variables they can accommodate (Pritikin et al., 2018).

Another important advantage of factored specifications applies to regression models with interactive and other types of nonlinear effects. Subsequent to SG2002, researchers primarily relied on the so-called just-another-variable approach that treats product and polynomial terms like any other normally distributed variable (von Hippel, 2009). Analytic work shows this strategy requires a restrictive missing completely at random process where missingness does not depend on the data (S. R. Seaman et al., 2012), and computer simulation studies demonstrate its potential for bias (Cham et al., 2017; Enders et al., 2014; Humberg & Grund, 2022; Q. Zhang & Wang, 2017). In contrast, maximum likelihood estimators for factored specifications treat product and polynomial terms as deterministic functions of the pseudo-imputations or nodes; these functions define the center and spread of the univariate distributions to which they contribute, but they do not invoke distributional assumptions, nor do they require their own nuisance models. Recent simulation studies are very promising, showing that factored approaches are superior to their just-another-variable predecessors when applied to models with interactive and polynomial effects (Humberg & Grund, 2022; Lüdtke et al., 2020a).

Bayesian Estimation

The Bayesian paradigm views parameters as random variables, and a probability distribution called the posterior encodes our subjective knowledge about different realizations of a parameter after collecting and analyzing data (Levy & Mislevy, 2016). SG2002 largely presented Bayesian estimation as a method in service of multiple imputation. In that application, a researcher uses Bayesian methods to fit a generic model to the data (e.g., one for multivariate normal data), and the resulting estimates define model-predicted distributions of the missing scores. An iterative Markov chain Monte Carlo (MCMC) algorithm repeatedly cycles between estimating the model parameters conditional on the filled-in data, then sampling replacement scores from distributions that condition on the estimates. Saving a small number of imputed data sets from this iterative sequence and fitting the focal analysis model to the filled-in data gives frequentist estimates that average over the distributions of missing values. Importantly, this application puts imputations at the fore, and the Bayesian parameter estimates themselves play no inferential role.

Bayesian analyses have since gained a strong foothold in social and behavioral science disciplines (Andrews & Baguley, 2013; van de Schoot et al., 2017), and user-friendly software tools abound (e.g., Stan, Gelman et al., 2015; Blimp, Keller & Enders, 2021; Mplus, L. K. Muthén & B. O. Muthén, 1998–2017; JASP, Wagenmakers et al., 2018). Unlike the fairly narrow lens depicted in SG2002, modern Bayesian analyses are an attractive direct estimation competitor to maximum likelihood. In this application, a researcher fits a model to the incomplete data and uses the resulting estimates to inform their substantive research questions; missing data imputation is a means to a more important end, which is to use Bayesian estimates and posterior summaries for inference.

Factored Regression Specifications

Like maximum likelihood estimators of the day, the multivariate normal distribution was the predominant Bayesian model described in SG2002. Multivariate models are still very useful, but Bayesian estimation's natural fit with factored regression specifications makes it a formidable missing data tool. Ibrahim and colleagues published the seminal work on Bayesian factorizations for missing at random and missing not at random processes around the time SG2002 was published (Huang et al., 2005; Ibrahim et al., 2002), but broader interest in their approach was still more than a decade away.

Subsequent work has extended factored regression specifications to moderated and curvilinear regression models (Asparouhov & Muthén, 2021a; S.-Y. Lee et al., 2007; Lüdtke et al., 2020b; Q. Zhang & Wang, 2017), regression models with discrete metrics and nonnormal continuous outcomes (Asparouhov & Muthén, 2021b; M. C. Lee & Mitra, 2016; Lüdtke et al., 2020b), latent variable models (Keller & Enders, 2021; S.-Y. Lee & Shi, 2000; Lüdtke et al., 2020b; Merkle & Rosseel, 2018; Palomo et al., 2007), models for missing not at random processes (Du, Enders, et al., 2022), auxiliary variable models (M. J. Daniels et al., 2014; Lüdtke et al., 2020b), multilevel models (Enders et al., 2020; Erler et al., 2016, 2019; Goldstein et al., 2014; Grund, Lüdtke, & Robitzsch, 2021b), and scale scores with missing item responses (Alacam et al., 2022), among others. The development of Bayesian missing data handling procedures has arguably outpaced

that of maximum likelihood, as the aforementioned specifications encompass a much broader cache of applications than is currently possible with likelihood-based approaches.

As explained previously, the factored regression framework defines the conditional distribution of an incomplete variable as the product of two or more univariate distributions (see Equation 3). Unlike maximum likelihood optimizers, which average over a grid of replacement scores, Bayesian estimation routines randomly sample imputations from these composite distributions. A typical MCMC algorithm cycles between two broad steps: the first step estimates the parameters of two or more regression models, conditional on the filled-in data (e.g., the three regressions depicted on the right side of Equation 1); and the second step uses the resulting parameter values to define distributions of missing values, from which it random samples new imputations.

Composite distributions like the one from Equation 3 can be quite complex, especially with models that feature nonlinear terms or categorical variables. These distributions sometimes have an analytic form (Levy & Enders, 2021), but in many cases, they do not and require specialized algorithmic approximations to generate the imputations (e.g., the Metropolis–Hastings algorithm). Regardless of origin, the imputations are just a means to an end, which is to use the Bayesian parameter estimates and posterior summaries for inference. Importantly, this application mimics the logic of maximum likelihood missing data handling: a researcher fits a model to the incomplete data and uses the resulting estimates to inform their substantive research questions.

Multiple Imputation

A typical application of multiple imputation consists of an imputation phase that fills in the data multiple times, an analysis phase where one or more analysis models are fit to the filled-in data, and a pooling phase that uses “Rubin’s rules” (Little & Rubin, 2020; Rubin, 1987) to combine imputation-specific point estimates and standard errors into a single package of results. Focusing on the imputation phase, the multivariate normal model was by far the predominant approach in the SG2002 era, as Joseph Schafer had earlier popularized this method in his classic textbook (Schafer, 1997). The MCMC algorithm for multivariate normal missing data repeatedly updates the model parameters—a mean vector and covariance matrix—conditional on the current filled-in data, after which it samples new synthetic scores from normal distributions based on the updated parameter values. Saving a small number of imputed data sets from a long iterative sequence and performing analyses on the filled-in data sets gives estimates that average over the distributions of missing values. Descriptions of this classic procedure are widely available in the literature, including in SG2002.

Robustness and Nonnormality

A major thread of methodological research in the early aughts focused on extending the multivariate normal imputation model's utility and understanding its limitations. As an example, several studies considered the impact of applying the model to nonnormal variables (Demirtas et al., 2008; K. J. Lee & Carlin, 2017; von Hippel, 2013; Yuan et al., 2012). A common finding from this work is that important estimands like means and regression coefficients are largely unaffected by distributional misspecifications,

and analytic evidence suggests that certain estimands are consistent (Yuan et al., 2012). In related work, several authors discuss the possibility of transforming nonnormal variables prior to imputation, then back-transforming to the original metric after. Possibilities include Box–Cox, logarithmic, square root, inverse, fourth-root, and logit transformations, among others (Goldstein et al., 2009; K. J. Lee & Carlin, 2017; Schafer & Olsen, 1998; Su et al., 2011; van Buuren, 2018; von Hippel, 2013). Authoritative studies conclude that this transform-then-impute strategy should be avoided because it can exacerbate rather than mitigate biases due to nonnormal data (K. J. Lee & Carlin, 2017; von Hippel, 2013).

A related strand of literature describes various methods for generating nonnormal imputations. Predictive mean matching leverages the same estimate–impute sequence described earlier, but it draws imputations from a donor pool of observed scores taken from participants whose predicted values are similar to that of the person with missing data (Kleinke, 2017; K. J. Lee & Carlin, 2017; van Buuren, 2018; Vink et al., 2014). The imputation step for this procedure is nonparametric because it does not assume anything about the distribution of missing values. Van Buuren’s (2018) multiple imputation text provides details about the procedure, which is available in several software programs, including his popular R package mice (van Buuren & Groothuis-Oudshoorn, 2011). Morris et al. (2014) catalog variations of the procedure in software and provide recommendations about the optimal matching criterion and donor pool size.

Another strategy is to sample imputations directly from nonnormal distributions (de Jong et al., 2016; Demirtas, 2017; Demirtas & Hedeker, 2008a, 2008b; He & Raghunathan, 2009). Among these possibilities, the Yeo–Johnson normal distribution is broadly applicable and readily available in computer software. The Yeo–Johnson approach invokes a normally distributed transformed variable and a shape parameter that maps transformed scores to a skewed raw score distribution (Lüdtke et al., 2020b; Yeo & Johnson, 2000). Embedding the transformation and shape parameter into the MCMC estimation process tailors imputations to match the observed-data distribution. This approach, which subsumes a number of common transformations (e.g., inverse, logarithmic, square root, and Box–Cox), has shown promise when paired with factored regression specifications (Keller & Enders, 2023; Lüdtke et al., 2020b). The procedure is available in the Blimp application (Keller & Enders, 2021) as well as the R package mdmb (Grund, Lüdtke, & Robitzsch, 2021b). Although less is known about this approach, imputation methods based on generalized additive models are similar in the sense that they incorporate shape parameters that accommodate a broad range of distributions for the missing data (de Jong et al., 2016).

Categorical Variables

Schafer’s (1997) classic text described multiple imputation approaches for multivariate nominal data and mixtures of nominal and continuous variables, but these approaches had important practical limitations that limited their broad adoption (Belin et al., 1999). Research succeeding SG2002 explored the utility of applying a normal imputation model to categorical variables and rounding the continuous imputes to discrete values (Ake, 2005; Allison, 2005; Bernaards et al., 2007; Graham, 2009; Horton et al., 2003; van Ginkel et al., 2007; Yucel et al., 2008). These studies largely showed that rounding is unnecessary and even detrimental in some cases. Fortunately, ad hoc rounding approaches are now a historical

footnote, as mature imputation approaches for mixed response types abound.

Contemporary variants of the multivariate normal model use a latent response (i.e., probit regression) framework to accommodate incomplete binary, ordinal, and multicategorical nominal variables (Asparouhov & Muthén, 2010; Carpenter & Kenward, 2013; Demirtas, 2017; Goldstein et al., 2009; Quartagno & Carpenter, 2019). This approach envisions binary and ordinal scores originating from a normally distributed latent variable, the distribution of which is carved into discrete segments by one or more threshold parameters (Johnson & Albert, 1999). The variation for multicategorical nominal variables instead uses category-specific latent variables that are ultimately expressed as a set of latent difference scores, much like a dummy coding scheme (Aitchison & Bennett, 1970; Carpenter & Kenward, 2013; Enders, 2022; Goldstein et al., 2009). Recent work has extended the latent response framework to count and other types of outcomes (Asparouhov & Muthén, 2021b; Demirtas, 2017; Polson et al., 2013).

Fully Conditional Specification

The introduction of fully conditional specification (van Buuren, 2007, 2018; van Buuren et al., 2006; van Buuren & Groothuis-Oudshoorn, 2011) was another major innovation that substantially broadened multiple imputation’s applicability. Rather than working directly from a multivariate distribution, the procedure imputes variables one at a time using a sequence of univariate regression models, each of which typically features an incomplete variable regressed on all other variables. Importantly, each regression in the sequence is tailored to the incomplete variable’s metric, allowing for a potentially diverse collection of generalized linear imputation models.

Variations of van Buuren’s procedure are available for imputing latent response variables (Enders, 2022; Grund, Lüdtke, & Robitzsch, 2021c; Keller & Enders, 2021), incomplete covariates in moderated regression models (J. W. Bartlett et al., 2015; J. W. Bartlett & Morris, 2015), multilevel data structures (Enders, Keller, et al., 2018; van Buuren, 2011), count and zero-inflated variables (Kleinke & Reinecke, 2013), classification and regression trees (Doove et al., 2014; Shah et al., 2014), and regularized regression (Deng et al., 2016; Zhao & Long, 2016), among others. Stef van Buuren’s excellent multiple imputation book (van Buuren, 2018) provides a detailed account of the fully conditional specification framework and its extensions.

A potential Achilles heel of fully conditional specification is that its regressions may be incompatible, meaning that the univariate conditional distributions induced by two or more regression models cannot derive from the same multivariate distribution (J. W. Bartlett et al., 2015; Du, Alacam, et al., 2022; J. C. Liu et al., 2014). Note that this concept is distinct from congeniality, which refers to the match or mismatch between the variables used in the imputation and analysis models (Meng, 1994; Schafer, 2003). Simulation studies suggest that routine incompatibilities often have a negligible impact on the final estimates (Arnold et al., 2001; Raghunathan et al., 2001; van Buuren et al., 2006). One such example of an innocuous incompatibility occurs when deploying a logistic imputation model for an incomplete binary variable and a linear imputation model for a continuous variable, as known multivariate distributions cannot induce these regressions.

An important example of incompatibility occurs with models that feature incomplete interactive or nonlinear effects, where conventional fully conditional specification approaches based on just-another-variable imputation can produce substantial bias, even when the missing at random assumption is satisfied (Kim et al., 2015, 2018; S. R. Seaman et al., 2012; von Hippel, 2009). This incompatibility results from misspecifying the distributions of incomplete predictors, which are potentially asymmetric and feature heteroscedastic variation (see Enders, 2022, equation 5.22). Newer variations of fully conditional specification called substantive model-compatible imputation address this shortcoming (J. W. Bartlett et al., 2015; J. W. Bartlett & Morris, 2015), and generalized additive models with smoothing splines can accommodate a variety of nonlinear relations and nonnormal distribution shapes (de Jong et al., 2016).

Factored Regression Specifications

Parametric imputation schemes that sample replacement scores from a distribution generally rely on Bayesian estimation as their mathematical machinery. Thus, all of the aforementioned innovations in the factored regression space also extend to multiple imputation—creating multiple imputations is simply a matter of saving the filled-in data sets from a small subset of MCMC iterations. Readers who dig deeper into this literature will encounter three typologies of factored regression specification: a sequential specification like the one in Equation 1 (Erler et al., 2016, 2019; Lüdtke et al., 2020b), a factorization that assigns a multivariate distribution to a set of predictors (Carpenter & Kenward, 2013; Enders et al., 2020; Goldstein et al., 2014), and a variation on fully conditional specification called substantive model-compatible imputation (J. W. Bartlett et al., 2015; J. W. Bartlett & Morris, 2015). The sequential specification is perhaps the most flexible because it can accommodate a broader range of predictor metrics (e.g., skewed continuous, count variables) as well as nonlinear associations among the covariates. However, in most applications, these parameterizations are effectively equivalent, although software packages have different capabilities.

Multiple Imputation Inference

Although most of the major innovations have centered on improving the imputation stage, inferential methods for the pooling stage have also been a focus of recent research. For example, modern simulation studies have broadened our understanding of the classic Wald and likelihood ratio tests for multiply imputed data (Grund et al., 2016c; Grund, Lüdtke, & Robitzsch, 2021a; Y. Liu & Enders, 2017; Y. Liu & Sriutaisuk, 2020), and new degrees of freedom expressions can improve the small-sample properties of some of these tests (Licht, 2010; Reiter, 2007). Chan and Meng (2021) also proposed a new likelihood ratio test (the D_4 statistic) that performs comparably to its classic counterpart from Meng and Rubin (1992) while being easier to compute.

For researchers in the behavioral and social sciences, the development of multiple imputation inference for structural equation modeling analyses is an important innovation. Multiple imputation was not a viable technique for this application when SG2002 was published, but researchers now have the full complement of tools necessary to fit structural equation models to imputed data sets. Recent developments include new methods for assessing global model fit (Chung &

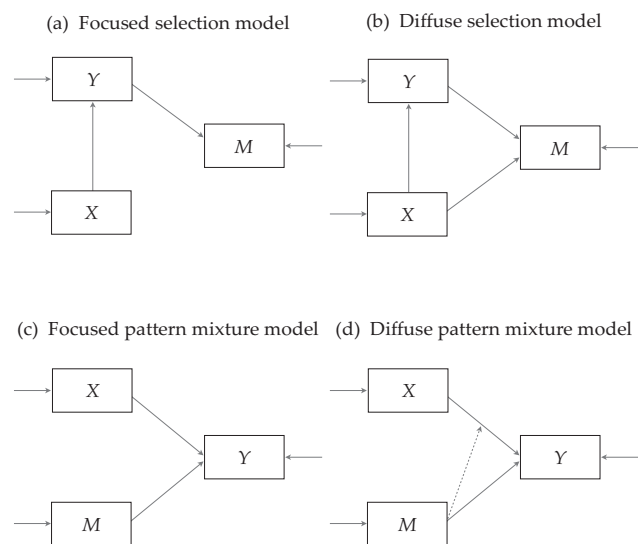
Cai, 2019; T. Lee & Cai, 2012; Y. Liu et al., 2021; Y. Liu & Sriutaisuk, 2020), fit indices based on pooled test statistics (Enders & Mansolf, 2018), pooling methods for rescaled test statistics (Jorgensen, 2021), and model modification indices (i.e., score tests; Mansolf et al., 2020). The R package *semTools* (Jorgensen, 2021) features several procedures that are not widely available in commercial software. T. Lee and Shi (2021) provide a comprehensive simulation study that compares the performance of maximum likelihood and multiple imputation for structural equation modeling analyses.

Missing Not at Random Processes

The two major modeling frameworks for missing not at random processes—selection models and pattern mixture models—mitigate nonresponse bias by introducing a model that describes the occurrence of missing data, codified by a binary missing data indicator ($M = 0$ if a score is observed, and $M = 1$ if it is missing). The two frameworks use the missing data indicator in very different ways; a selection model features a regression equation with a missing data indicator as a dependent variable, whereas a pattern mixture model uses the missing data indicator as a predictor. Figure 3 shows path diagrams for a simple regression model where outcome scores follow a missing not at random process. The figure highlights that a selection model aligns with a mediating process where analysis variables influence missingness directly or indirectly via the variable with missing data, and a pattern mixture model corresponds to a moderation process where model parameters differ between missingness groups.

Selection and pattern mixture models were well established by the time SG2002 was published (Diggle & Kenward, 1994; Hedeker &

Figure 3
Path Diagrams Depicting Selection and Pattern Mixture Models for Focused and Diffuse Missing Not at Random Processes



Note. Path diagrams that depict selection and pattern mixture models for focused and diffuse missing not at random processes. The models involve a complete variable, X , an incomplete variable, Y , and a binary missing data indicator, M_Y , coded 0 if Y is observed and 1 if it is missing.

Gibbons, 1997), as were shared parameter models, which are variants of the selection model that feature one or more latent variables predicting missingness (M. C. Wu & Carroll, 1988). The authors provide an excellent discussion of both approaches, and a number of accessible tutorial papers were published in the years following as these models became available in software (Albert & Follmann, 2009; Enders, 2011; Feldman & Rabe-Hesketh, 2012; B. Muthén et al., 2011; S. Xu & Blozis, 2011). Simultaneously, methodological work published in behavioral science journals clarified the strengths and weaknesses of different modeling approaches and extended their reach to a broader audience (Gomer & Yuan, 2021; Gottfredson, Bauer, & Baldwin, 2014; Gottfredson, Bauer, Baldwin, & Okiishi, 2014; Gottfredson et al., 2017; Sterba & Gottfredson, 2015; Yang et al., 2015; Yang & Maxwell, 2014).

SG2002 was not overly bullish on missing not at random analyses, suggesting they be reserved for sensitivity analyses in contexts where the reasons for dropout are plausibly related to the unseen score values (e.g., clinical studies). The idea behind this approach is to pair the focal model with different missingness models in order to examine whether one's substantive conclusions are consistent across different assumptions about the missing data process. Although journal pages are no less constrained than they were 20 years ago, online supplemental documents now provide an ideal vehicle for reporting multiple sets of results that speak to the stability (or lack thereof) of key findings. Enders (2022) provides an illustration and recipe for sensitivity analyses based on Gomer and Yuan's (2021) missing not at random subtypes, and Carpenter and Kenward (2013) also provide an extensive discussion of this topic.

While it is true that selection models and pattern mixtures require special considerations and careful application, these models are arguably underutilized in 2022. For one, software tools are no longer a barrier to implementation; researchers have multiple options, and it is easier than ever to fit these models. Likewise, researchers now have access to a wealth of accessible tutorial papers that were not available 20 years ago, many of which include real data analysis examples with software scripts. Additionally, the literature on these models is considerably more mature; we know more about the strengths and weaknesses of different approaches, and simple diagnostics can help identify models with dubious support from the data (Enders, 2022; Sterba & Gottfredson, 2015). Finally, new graphical methods for studying parameter recovery (Mohan et al., 2013; Mohan & Pearl, 2021; Pearl & Mohan, 2013) provide a complementary tool for understanding whether a missing not at random analysis can produce meaningful estimates.

Composite Scores and Item-Level Missing Data

Researchers collecting self-report data routinely use questionnaires with multiple items that tap into different features of the target construct. When analyzing such data, the focus is usually a composite score, often a sum (or average) of the item responses or a factor score based on a weighted average; McNeish and Wolf (2020) and Widaman and Revelle (2022) discuss the relative merit of these two approaches. SG2002 addressed the common problem of item-level missing data, cautioning against the widespread practice of computing scale scores by averaging the available item responses (this practice is also known as person-mean imputation, and the resulting composite is referred to as a prorated scale score). Subsequent research confirmed their intuition about this procedure,

demonstrating its propensity for bias in the common (if not prevailing) case where item means or intercorrelations differ (Mazza et al., 2015).

SG2002 instead advocated for a multiple imputation procedure that computes composite scores from filled-in item responses. Subsequent research convincingly demonstrated that item-level imputation substantially enhances power, and it does so without imposing structure on the data (Eekhout et al., 2014, 2015a, 2015b; Gottschall et al., 2012; Mazza et al., 2015; Savalei & Rhemtulla, 2017; van Buuren, 2010). The literature describes a variety of other approaches to imputing item responses (Bernaards & Sijtsma, 2000; Sijtsma & van der Ark, 2003; van Ginkel et al., 2007, 2010), but the multiple imputation procedure recommended by SG2002 is the gold standard.

In practice, multiply imputing discrete item responses is often difficult to implement because categorical imputation models suffer from the same "curse of dimensionality" problem that plagues the estimation of many psychometric models (Cai, 2010). Factored regression specifications are an emerging solution for imputing scale scores with large number of items. Alacam et al. (2022) described factorizations that use equality constraints on item-level regression parameters to reduce imputation model complexity. The procedure is potentially advantageous because it builds item-level missing data handling into a direct estimation routine for a regression model with composite scores as predictors or the outcome. Preliminary computer simulations suggest the procedure can produce approximately unbiased estimates in situations where fully conditional specification or joint model imputation routines fail.

Factor scores computed from psychometric models are another way to accommodate item-level missing data. Of course, factor scores have a long history in the psychometric literature, and recent studies have extended familiar approaches to maximum likelihood estimation with missing data (Estabrook & Neale, 2013; Lawes & Eid, 2022; Loncke et al., 2018). Pairing Bayesian estimation with factored regression specifications provides interesting opportunities for latent variable modeling and factor score computation that are currently difficult or impossible with likelihood-based methods. In this context, the latent variable scores in a psychometric model are missing data to be imputed, just like any other incomplete variable (Aßmann et al., 2015; Keller & Enders, 2021; S.-Y. Lee & Shi, 2000; Merkle & Rosseel, 2018; Palomo et al., 2007; M. Wu, 2005). Viewing latent variables as missing data paves the way for estimating plausible values in models with latent variable interactions and interactions between latent and manifest variables (Enders, 2022; Keller & Enders, 2021).

Multilevel Missing Data

The emergence of missing data handling methods for multilevel models is arguably one of the more important developments for behavioral science researchers since SG2002. Methodologists have extended all three analytic pillars to accommodate a broad range of multilevel applications, and numerous accessible tutorial papers and chapters are available that summarize these developments (Enders, 2022; Enders et al., 2016; Grund et al., 2016b, 2019; Grund, Lüdtke, & Robitzsch, 2021b; van Buuren, 2011).

Rewinding to 2002, maximum likelihood estimators in mixed modeling software could accommodate incomplete outcomes, but they had no capacity for treating incomplete predictors. This is

still mostly true today, with two notable exceptions: the HLM software (Raudenbush, 2019) estimates random intercept models with incomplete normal predictors (Shin & Raudenbush, 2007, 2010, 2013), as do multilevel structural equation modeling programs. Some structural equation modeling frameworks also allow for incomplete random slope predictors (e.g., *Mplus*; L. K. Muthén & B. O. Muthén, 1998–2017). At least for now, the literature suggests that maximum likelihood is better suited for random intercept models (Grund et al., 2019), as limited simulation evidence suggests that Monte Carlo integration procedures for random coefficient models can introduce bias (Enders et al., 2020; Enders, Hayes, et al., 2018). Grund et al. (2018) report simulation results that evaluate the maximum likelihood estimator for random intercept models, and work on incomplete random slope predictors is ongoing (Rockwood, 2020).

Prior to the advent of sophisticated imputation techniques for multilevel data, researchers could use a fixed effect imputation procedure that dummy codes the Level-2 units and introduces the code variables as predictors in a single-level imputation scheme. Fixed effect imputation is computationally simple and produces approximately unbiased parameter estimates in certain random intercept applications (Lüdtke et al., 2017; Reiter et al., 2006), although it may inflate standard errors and distort confidence intervals (Andridge, 2011; van Buuren, 2011). Nevertheless, the procedure may be desirable when the number of clusters is very small (a situation that makes random effect estimation difficult).

Joseph Schafer extended his popular joint model imputation framework to multilevel data structures around the same time as SG2002 (Schafer, 2001; Schafer & Yucel, 2002), and a number of flexible variations of his approach subsequently appeared in the literature (Asparouhov & Muthén, 2010; Carpenter et al., 2011; Carpenter & Kenward, 2013; Goldstein et al., 2009, 2014; Yucel, 2008). These newer approaches generally allow missing data at any level of the data hierarchy, and they use a latent response formulation to accommodate incomplete categorical variables. Importantly, most incarnations of joint model imputation are limited to random intercept analyses and have no capacity for preserving random associations among incomplete variables. The exception is a variant of the joint model that treats within-cluster covariance matrix elements as random parameters (Quartagno & Carpenter, 2016, 2020; Yucel, 2011).

In a similar vein, the literature also describes multilevel extensions of Stef van Buuren's fully conditional specification imputation for two- and three-level models (Audigier et al., 2018; Enders, Keller, et al., 2018; Keller, 2015; Resche-Rigon & White, 2018; van Buuren, 2011). Fully conditional specification should also be reserved for random intercept models because the procedure fails to capture heteroscedastic variation in the conditional distribution of random slope predictors (Enders et al., 2020, equation 20). Simulation studies show that applying fully conditional specification (i.e., "reverse random coefficient" imputation) to such an analysis can introduce substantial bias (Enders et al., 2016, 2020; Grund et al., 2016a, 2018). Several studies have compared fully conditional specification to joint model imputation, and the procedures perform well and are effectively equivalent when applied to random intercept models (Grund et al., 2017, 2018, 2019; Lüdtke et al., 2017; Mistler & Enders, 2017).

A good deal of recent research has focused on the development of missing data handling for random coefficient and multilevel

moderated regression models (Enders et al., 2020, 2016, 2019; Goldstein et al., 2014; Grund et al., 2018; Grund, Lüdtke, & Robitzsch, 2021b; Keller & Enders, 2023). These analyses are challenging because they induce predictor distributions with heteroscedastic variation, a feature that not all estimators can readily accommodate. Currently, the literature supports Bayesian estimation and multiple imputation routines based on factored regression specifications. These procedures are available in specialized Bayesian programs like JAGS (Erlor et al., 2016) as well as more user-friendly programs like Blimp (Keller & Enders, 2021) and the R package *mdmb* (Grund, Lüdtke, & Robitzsch, 2021b).

Current Software Landscape

This final section provides brief summary of the current software landscape. Structural equation modeling software programs are often the most general way to implement maximum likelihood estimation. Commercial software packages like SAS (CALIS; SAS Institute Inc., 2011), SPSS (AMOS; Arbuckle, 2019), and Stata (*gllamm*; Rabe-Hesketh et al., 2004) all have structural equation modeling modules, and most readers are no doubt aware of specialized applications like EQS (Bentler, 2000–2008), LISREL (Jöreskog & Sörbom, 2018), and *Mplus* (L. K. Muthén & B. O. Muthén, 1998–2017). On the R platform, *OpenMx* (Boker et al., 2011), *PLmixed* (Rockwood & Jeon, 2019), and *lavaan* (Rosseel, 2012) with *semTools* (Jorgensen, 2021) are options. While these programs offer broad toolkits that include innovations described earlier, they generally invoke multivariate normality. *Mplus*, *OpenMX*, *gllamm*, and *PLmixed* are exceptions that implement estimators for mixtures of categorical and continuous variables. The R package *mdmb* (Lüdtke et al., 2020a) offers factored regression specifications for single-level regression models with interactive or nonlinear effects, and the moderated latent structural equation model facilities in *Mplus* and the R package *nlsem* (Umbach et al., 2017) are an alternative approach to incomplete interactive effects (Cham et al., 2017).

Turning to Bayesian estimation, *Mplus* offers a powerful feature set for multivariate normal data, but that multivariate focus carries the same limitations as it does with maximum likelihood. Several R packages implement factored regression specifications with missing data, including *rstan* (Guo et al., 2020), *rjags* (Plummer, 2019), *R2OpenBUGS* (Sturtz et al., 2019), *brms* (Bürkner, 2021), *blavaan* (Merkle & Rosseel, 2018), *mdmb* (Grund, Lüdtke, & Robitzsch, 2021b), *NIMBLE* (de Valpine et al., 2017), and *JointAI* (Erlor, 2021); among these many options, *blavaan*, *brms*, *mdmb*, and *JointAI* are among the most user-friendly. Finally, *Blimp* (Keller & Enders, 2021) is an all-purpose data analysis and latent variable modeling program that harnesses the flexibility of factored regression specifications in a user-friendly application that requires minimal scripting and no deep-level knowledge about the Bayesian paradigm.

Turning to multiple imputation, most commercial software packages offer variations of joint model imputation, fully conditional specification, or both. The aforementioned *Blimp* package also offers single-level and multilevel fully conditional specification as well as model-based multiple imputation based on factored regression specifications. Not surprisingly, multiple imputation options also abound in R. A partial list includes Stef van Buuren's popular *MICE* package (van Buuren & Groothuis-Oudshoorn, 2011), *pan* (Grund et al., 2016b; Schafer, 2018), *jomo* (Quartagno &

Carpenter, 2016, 2020), Amelia (Honaker et al., 2021), and smcfs (J. Bartlett et al., 2021; J. W. Bartlett et al., 2015), among others. Most of the aforementioned Bayesian analysis programs also allow users to save imputations constructed from the estimated model (i.e., model-based multiple imputation). Regardless of where the imputations originate, the R package *mitml* (Grund, Robitzsch, & Luedtke, 2021) provides a comprehensive toolkit for pooling estimates and conducting significance tests, as does the *semTools* package.

Discussion

This year marks the 20th anniversary of Schafer and Graham's (2002) highly cited paper "Missing data: Our view of the state of the art." Not surprisingly, the past two decades of missing data research have brought numerous, exciting developments; the range of applications that are possible with modern missing data techniques has increased dramatically, and software options are light years ahead of where they were in 2002. The purpose of this article was to provide an update on the state of the art that catalogs important innovations from the last two decades of research, with the goal of providing methodologists and practicing researchers with a springboard for accessing the most up-to-date missing data handling methodologies.

One lesson from the past 20 years is that complex analyses often require specific solutions that preserve important distributional features of the incomplete predictors (e.g., heteroscedasticity and non-normality). Common examples include models with interactive effects, curvilinear terms, or random coefficients, all of which require a missing data routine tailored to that specific analysis. A focused strategy is not foreign to maximum likelihood missing data handling—which is inherently model-based—but it is at odds with widespread perceptions about multiple imputation. In particular, the need for specificity precludes the possibility of creating filled-in data sets that cater to many different researchers with diverse substantive goals (an oft-cited advantage of multiple imputation that dates to its origins). In truth, the idea of creating general-use imputations did not originate in the behavioral sciences (e.g., see Rubin, 1996), and it probably is not suited for most behavioral science data sets, where sample sizes are not large enough to support high-dimensional estimation problems with dozens or hundreds of variables. Although it may be contrary to how many researchers think about imputation, there is absolutely nothing wrong with creating filled-in data sets on an analysis-by-analysis basis (van Buuren, 2018).

The conclusion that models with interactive or nonlinear effects require tailored solutions raises interesting questions about missing data handling in predictive modeling contexts, where the goal is to discover such effects rather than evaluate theoretically-derived propositions about them (e.g., see Yarkoni & Westfall, 2017). There is growing interest in missing data methods for machine learning methods, including applications in the social and behavioral sciences (Golino & Gomes, 2016; Gunn et al., 2022; Hayes et al., 2015). A recent review of published machine learning applications suggests that the majority of studies use deletion methods, despite the fact that many software packages offer better alternatives (Nijman et al., 2022). To the extent that this review's conclusions generalize across disciplines, there are ripe opportunities for methodological innovations in this space going forward. Several accessible discussions of missing data handling methods for machine learning models

are available in the literature (Emmanuel et al., 2021; Provost & Saar-Tsechanski, 2007; Thomas et al., 2020; van Buuren, 2018, section 3.5).

Factored regression specifications have received considerable attention in the recent literature, and this framework is certainly one of the more important advances described in this paper. Although the research is generally promising, we still know relatively little about these methods, especially when compared to classic approaches described in SG2002. Methodologists are just beginning to exploit the framework's flexibility, and the procedure will undoubtedly be an area of ongoing interest during the next 20 years. For the remainder of this section, I reflect on some of the challenges associated with factored regression specifications.

From a practical perspective, factored regression specifications are more conceptually challenging than their normal-theory predecessors. Almost all of us encounter the multivariate normal distribution at some point in our graduate training, and the idea that a multidimensional bell curve could describe the distributions of missing values is somewhat intuitive, even for researchers with no formal exposure to missing data techniques. In contrast, factored regression specifications are opaquer because the probability rules that give rise to this framework are less likely to surface in a typical graduate statistics sequence. Moreover, the idea that a distribution of missing values is obtained by multiplying two or more distributions (see Equation 3) is even more mysterious without churning through tedious algebra to gain deeper insight. Although these conceptual challenges are not barriers to implementing factored specifications, they do imply a steeper learning curve for researchers who want deeper mastery of the methodology—the alchemy under the hood is unquestionably more complex.

A former colleague often jokes that he "clicks the FIML button" for his analyses. This comment is a reminder that some of the classic methods from SG2002 are largely plug and play. While factored regression specifications are easily within the grasp of most researchers, they are not fully self-driving and can be challenging to implement, depending on the application. As mentioned previously, there are usually several ways to factorize a joint distribution (see Figure 2 and the text surrounding Equation 1), and there is often ambiguity about how to deploy regression models that link incomplete predictors. Factored regression specifications are essentially path models, and ordering recommendations from the literature (Lüdtke et al., 2020b; D. Xu et al., 2016) can produce a configuration of nonsensical directed pathways that defy temporal logic. Importantly, this feature is not a statistical or substantive problem because the nuisance regressions are just a mathematical tool for linking incomplete predictors. Although factored specifications are not invariant with respect to ordering (D. Xu et al., 2016), my experience is that different sequences usually produce equivalent results. When in doubt, a researcher can conduct a sensitivity analysis that deploys different orderings.

Finally, it probably comes as no surprise that factored regression specifications are more computationally intensive than their classic counterparts, especially for models with many categorical variables. This is true for maximum likelihood optimizers that rely on complex numeric or Monte Carlo integration schemes, as well as for MCMC algorithms that sample imputations from complex composite functions. Fortunately, computing advances make these models a practical reality for a broad swath of realistic applications. However,

massive data sets with hundreds of thousands or millions of observations are becoming increasingly common in behavioral science applications, and big data pose formidable challenges for computationally intensive missing data methods. Unfortunately, it may not be possible to process huge data matrices and apply newer approaches in a tolerable amount of time.

In closing, I want to emphasize the continued relevance of SG2002 on its 20th anniversary. It is a captivating paper about an interdisciplinary methodological problem, and it is still a must-read for anyone interested in learning about missing data. The methods described in SG2002 have continually evolved and improved, and I hope this paper serves as a useful update and supplement to this landmark publication.

References

- Aitchison, J., & Bennett, J. A. (1970). Polychotomous quantal response by maximum likelihood. *Biometrika*, 57(2), 253–262. <https://doi.org/10.1093/biomet/57.2.253>
- Ake, C. F. (2005). *Rounding after multiple imputation with non-binary categorical covariates*. SAS Users Group International.
- Alacam, E., Enders, C. K., Du, H., & Keller, B. T. (2022). *A factored regression model for composite scores with item-level missing data* [Manuscript submitted for publication].
- Albert, P. S., & Follmann, D. A. (2009). Shared-parameter models. In G. Fitzmaurice, M. Davidian, G. Vebeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 433–452). Chapman & Hall.
- Allison, P. D. (2005). *Imputation of categorical variables with PROC MI*. SAS Users Group International.
- Anderson, T. W. (1957). Maximum-likelihood estimates for a multivariate normal-distribution when some observations are missing. *Journal of the American Statistical Association*, 52(278), 200–203. <https://doi.org/10.1080/01621459.1957.10501379>
- Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, 66(1), 1–7. <https://doi.org/10.1111/bmsp.12004>
- Andridge, R. R. (2011). Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biometrical Journal*, 53(1), 57–74. <https://doi.org/10.1002/bimj.201000140>
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–278). Lawrence Erlbaum Associates.
- Arbuckle, J. L. (2019). *Amos 26.0 user's guide*. IBM SPSS.
- Arminger, G., & Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association*, 85(409), 195–203. <https://doi.org/10.1080/01621459.1990.10475326>
- Arnold, B. C., Castillo, E., & Sarabia, J. M. (2001). Conditionally specified distributions: An introduction (with comments and a rejoinder by the authors). *Statistical Science*, 16(3), 268–269. <https://doi.org/10.1214/ss/1009213728>
- Aßmann, C., Gaasch, C., Pohl, S., & Carstensen, C. H. (2015). Bayesian Estimation in IRT models with missing values in background variables. *Psychological Test and Assessment Modeling*, 57(4), 595–618. https://www.psychologie-aktuell.com/journale/search/ergebnis.html?tx_news_pi1%5Bnews%5D=299&cHash=97d000e1d2de08b1e0b35e1049de84d9
- Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus*. www.statmodel.com/download/Imputations7.pdf
- Asparouhov, T., & Muthén, B. (2021a). Bayesian Estimation of single and multilevel models with latent variable interactions. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(2), 314–328. <http://www.statmodel.com/download/Imputations7.pdf>
- Asparouhov, T., & Muthén, B. (2021b). Expanding the Bayesian structural equation, multilevel and mixture models to logit, negative-binomial and nominal variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(4), 622–637. <https://doi.org/10.1080/10705511.2021.1878896>
- Audigier, V., White, I. R., Jolani, S., Debray, T. P. A., Quartagno, M., Carpenter, J., van Buuren, S., & Resche-Rigon, M. (2018). Multiple imputation for multilevel data with continuous and binary variables. *Statistical Science*, 33(2), 160–183. <https://doi.org/10.1214/18-Sts646>
- Bartlett, J., Keogh, R., Bonneville, E. F., & Ekström, C. T. (2021). *Package "smcfcs"* [Computer software]. <https://cran.r-project.org/web/packages/smcfcs/smcfcs.pdf>
- Bartlett, J. W., & Morris, T. P. (2015). Multiple imputation of covariates by substantive-model compatible fully conditional specification. *The Stata Journal*, 15(2), 437–456. <https://doi.org/10.1177/1536867X1501500206>
- Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., & the Alzheimer's Disease Neuroimaging Initiative*. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research*, 24(4), 462–487. <https://doi.org/10.1177/0962280214521348>
- Beale, E. M. L., & Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(1), 129–145. <https://doi.org/10.1111/j.2517-6161.1975.tb01037.x>
- Belin, T. R., Hu, M. Y., Young, A. S., & Grusky, O. (1999). Performance of a general location model with an ignorable missing-data assumption in a multivariate mental health services study. *Statistics in Medicine*, 18(22), 3123–3135. [https://doi.org/10.1002/\(SICI\)1097-0258\(19991130\)18:22<3123::AID-SIM277>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-0258(19991130)18:22<3123::AID-SIM277>3.0.CO;2-2)
- Bentler, P. M. (2000–2008). *EQS 6 structural equations program manual*. Multivariate Software.
- Bernaards, C. A., Belin, T. R., & Schafer, J. L. (2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Statistics in Medicine*, 26(6), 1368–1382. <https://doi.org/10.1002/sim.2619>
- Bernaards, C. A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research*, 35(3), 321–364. https://doi.org/10.1207/S15327906MBR3503_03
- Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., & Fox, J. (2011). Openmx: An open source extended structural equation modeling framework. *Psychometrika*, 76(2), 306–317. <https://doi.org/10.1007/s11336-010-9200-6>
- Bürkner, P.-C. (2021). *Package 'brms'* [Computer software]. <https://cran.r-project.org/web/packages/brms/brms.pdf>
- Cai, L. (2010). Metropolis–Hastings Robbins–Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35(3), 307–335. <https://doi.org/10.3102/1076998609353115>
- Carpenter, J. R., Goldstein, H., & Kenward, M. G. (2011). REALCOM-IMPUTE software for multilevel multiple imputation with mixed response types. *Journal of Statistical Software*, 45(5), 1–14. <https://doi.org/10.18637/jss.v045.i05>
- Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. Wiley.
- Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behavioral Research*, 52(1), 12–30. <https://doi.org/10.1080/00273171.2016.1245600>
- Chan, K. W., & Meng, X.-L. (2021, December 30). *Multiple improvements of multiple imputation likelihood ratio tests*. <https://arxiv.org/abs/1711.08822>
- Chung, S., & Cai, L. (2019). Alternative multiple imputation inference for categorical structural equation modeling. *Multivariate Behavioral Research*, 54(3), 323–337. <https://doi.org/10.1080/00273171.2018.1523000>

- Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4), 330–351. <https://doi.org/10.1037/1082-989X.6.4.330>
- Daniel, R. M., Kenward, M. G., Cousens, S. N., & De Stavola, B. L. (2012). Using causal diagrams to guide analysis in missing data problems. *Statistical Methods in Medical Research*, 21(3), 243–256. <https://doi.org/10.1177/0962280210394469>
- Daniels, M. J., Wang, C., & Marcus, B. H. (2014). Fully Bayesian inference under ignorable missingness in the presence of auxiliary covariates. *Biometrics*, 70(1), 62–72. <https://doi.org/10.1111/biom.12121>
- de Jong, R., van Buuren, S., & Spiess, M. (2016). Multiple imputation of predictor variables using generalized additive models. *Communications in Statistics—Simulation and Computation*, 45(3), 968–985. <https://doi.org/10.1080/03610918.2014.911894>
- Demirtas, H. (2017). A multiple imputation framework for massive multivariate data of different variable types: A Monte-Carlo technique. In D. G. Chen & J. D. Chen (Eds.), *Monte-Carlo simulation-based statistical modeling* (pp. 143–162). Springer Nature Singapore.
- Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiple imputing non-Gaussian continuous outcomes: A simulation assessment. *Journal of Statistical Computation and Simulation*, 78(1), 69–84. <https://doi.org/10.1080/10629360600903866>
- Demirtas, H., & Hedeker, D. (2008a). Imputing continuous data under some non-Gaussian distributions. *Statistica Neerlandica*, 62(2), 193–205. <https://doi.org/10.1111/j.1467-9574.2007.00377.x>
- Demirtas, H., & Hedeker, D. (2008b). Multiple imputation under power polynomials. *Communications in Statistics—Simulation and Computation*, 37(8), 1682–1695. <https://doi.org/10.1080/03610910802101531>
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Deng, Y., Chang, C., Ido, M. S., & Long, Q. (2016). Multiple imputation for general missing data patterns in the presence of high-dimensional data. *Scientific Reports*, 6(1), 1–10. <https://doi.org/10.1038/srep21689>
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413. <https://doi.org/10.1080/10618600.2016.1172487>
- Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1), 49–93. <https://doi.org/10.2307/2986113>
- Doove, L. L., Van Buuren, S., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis*, 72(April), 92–104. <https://doi.org/10.1016/j.csda.2013.10.025>
- Du, H., Alacam, E., Mena, S., & Keller, B. T. (2022). Compatibility in imputation specification. *Behavior Research Methods*, 54(6), 2962–2980. <https://doi.org/10.3758/s13428-021-01749-5>
- Du, H., Enders, C., Keller, B. T., Bradbury, T. N., & Karney, B. R. (2022). A Bayesian latent variable selection model for nonignorable missingness. *Multivariate Behavioral Research*, 57(2-3), 478–512. <https://doi.org/10.1080/00273171.2021.1874259>
- Edgett, G. L. (1956). Multiple regression with missing observations among the independent variables. *Journal of the American Statistical Association*, 51(273), 122–131. <https://doi.org/10.2307/2280808>
- Eekhout, I., de Vet, H. C., Twisk, J. W., Brand, J. P., de Boer, M. R., & Heymans, M. W. (2014). Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *Journal of Clinical Epidemiology*, 67(3), 335–342. <https://doi.org/10.1016/j.jclinepi.2013.09.009>
- Eekhout, I., Enders, C. K., Twisk, J. W. R., de Boer, M. R., de Vet, H. C. W., & Heymans, M. W. (2015a). Analyzing incomplete item scores in longitudinal data by including item score information as auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 588–602. <https://doi.org/10.1080/10705511.2014.937670>
- Eekhout, I., Enders, C. K., Twisk, J. W. R., de Boer, M. R., de Vet, H. C. W., & Heymans, M. W. (2015b). Including auxiliary item information in longitudinal data analyses improved handling missing questionnaire outcome data. *Journal of Clinical Epidemiology*, 68(6), 637–645. <https://doi.org/10.1016/j.jclinepi.2015.01.012>
- Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, 8(1), 1–37. <https://doi.org/10.1186/s40537-021-00516-9>
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6(4), 352–370. <https://doi.org/10.1037/1082-989X.6.4.352>
- Enders, C. K. (2002). Applying the Bollen-Stine bootstrap for goodness-of-fit measures to structural equation models with missing data. *Multivariate Behavioral Research*, 37(3), 359–377. https://doi.org/10.1207/S15327906MBR3703_3
- Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, 16(1), 1–16. <https://doi.org/10.1037/a0022640>
- Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). Guilford Press.
- Enders, C. K., Baraldi, A. N., & Cham, H. (2014). Estimating interaction effects with incomplete predictor variables. *Psychological Methods*, 19(1), 39–55. <https://doi.org/10.1037/a0035314>
- 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(1), 88–112. <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(5), 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(2), 298–317. <https://doi.org/10.1037/met0000148>
- Enders, C. K., & Mansolf, M. (2018). Assessing the fit of structural equation models with multiply imputed data. *Psychological Methods*, 23(1), 76–93. <https://doi.org/10.1037/met0000102>
- 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(2), 222–240. <https://doi.org/10.1037/met0000063>
- Erler, N. S. (2021). *Package “JointAI”* [Computer software]. <https://cran.r-project.org/web/packages/JointAI/JointAI.pdf>
- Erler, N. S., Rizopoulos, D., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2019). Bayesian Imputation of time-varying covariates in linear mixed models. *Statistical Methods in Medical Research*, 28(2), 555–568. <https://doi.org/10.1177/0962280217730851>
- Erler, N. S., Rizopoulos, D., Rosmalen, J., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2016). Dealing with missing covariates in epidemiologic studies: A comparison between multiple imputation and a full Bayesian approach. *Statistics in Medicine*, 35(17), 2955–2974. <https://doi.org/10.1002/sim.6944>
- Estabrook, R., & Neale, M. (2013). A comparison of factor score estimation methods in the presence of missing data: Reliability and an application to

- nicotine dependence. *Multivariate Behavioral Research*, 48(1), 1–27. <https://doi.org/10.1080/00273171.2012.730072>
- Feldman, B. J., & Rabe-Hesketh, S. (2012). Modeling achievement trajectories when attrition is informative. *Journal of Educational and Behavioral Statistics*, 37(6), 703–736. <https://doi.org/10.3102/1076998612458701>
- Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, 44(4), 409–420. <https://doi.org/10.1007/BF02296204>
- Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5), 530–543. <https://doi.org/10.3102/1076998615606113>
- Goldstein, H., Carpenter, J., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling*, 9(3), 173–197. <https://doi.org/10.1177/1471082X0800900301>
- 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(2), 553–564. <https://doi.org/10.1111/rssa.12022>
- Golino, H. F., & Gomes, C. M. (2016). Random forest as an imputation method for education and psychology research: Its impact on item fit and difficulty of the Rasch model. *International Journal of Research & Method in Education*, 39(4), 401–421. <https://doi.org/10.1080/1743727X.2016.1168798>
- Gomer, B., & Yuan, K.-H. (2021). Subtypes of the missing not at random missing data mechanism. *Psychological Methods*, 26(5), 559–598. <https://doi.org/10.1037/met0000377>
- Gottfredson, N. C., Bauer, D. J., & Baldwin, S. A. (2014). Modeling change in the presence of non-randomly missing data: Evaluating a shared parameter mixture model. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 196–209. <https://doi.org/10.1080/10705511.2014.882666>
- Gottfredson, N. C., Bauer, D. J., Baldwin, S. A., & Okiishi, J. C. (2014). Using a shared parameter mixture model to estimate change during treatment when termination is related to recovery speed. *Journal of Consulting and Clinical Psychology*, 82(5), 813–827. <https://doi.org/10.1037/a0034831>
- Gottfredson, N. C., Sterba, S. K., & Jackson, K. M. (2017). Explicating the conditions under which multilevel multiple imputation mitigates bias resulting from random coefficient-dependent missing longitudinal data. *Prevention Science*, 18(1), 12–19. <https://doi.org/10.1007/s11121-016-0735-3>
- Gottschall, A. C., West, S. G., & Enders, C. K. (2012). A comparison of item-level and scale-level multiple imputation for questionnaire batteries. *Multivariate Behavioral Research*, 47(1), 1–25. <https://doi.org/10.1080/00273171.2012.640589>
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 80–100. https://doi.org/10.1207/S15328007sem1001_4
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2016a). Multiple imputation of missing covariate values in multilevel models with random slopes: A cautionary note. *Behavior Research Methods*, 48(2), 640–649. <https://doi.org/10.3758/s13428-015-0590-3>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2016b). Multiple imputation of multilevel missing data: An introduction to the R package pan. *Sage Open*, 6(4), 1–17. <https://doi.org/10.1177/2158244016668220>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2016c). Pooling ANOVA results from multiply imputed datasets: A simulation study. *Methodology*, 12(3), 75–88. <https://doi.org/10.1027/1614-2241/a000111>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2017). Multiple imputation of missing data at level 2: A comparison of fully conditional and joint modeling in multilevel designs. *Journal of Educational and Behavioral Statistics*, 43(3), 316–353. <https://doi.org/10.3102/1076998617738087>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, 21(1), 111–149. <https://doi.org/10.1177/1094428117703686>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2019). Missing data in multilevel research. In S. E. Humphrey & J. M. LeBreton (Eds.), *Handbook for multilevel theory, measurement, and analysis* (pp. 365–386). American Psychological Association.
- Grund, S., Robitzsch, A., & Lüdtke, O. (2021). Package “mitml” [Computer software]. <https://cran.r-project.org/web/packages/mitml/mitml.pdf>
- 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(6), 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(4), 430–465. <https://doi.org/10.3102/1076998620959058>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2021c). Pooling methods for likelihood ratio tests in multiply imputed data sets. <https://psyarxiv.com/d459g/>
- Gunn, H. J., Hayati Rezvan, P., Fernández, M. I., & Comulada, W. S. (2022). How to apply variable selection machine learning algorithms with multiply imputed data: A missing discussion. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000478>
- Guo, J., Gabry, J., Goodrich, B., Weber, S., Lee, D., Sakrejda, K., Martin, M., Sklyar, O., Oehlschlaegel-Akiyoshi, J., Maddock, J., Bristow, P., Agrawal, N., Kormanyos, C., & Steve, B. (2020). Package “rstan” [Computer software]. <https://cran.r-project.org/web/packages/rstan/rstan.pdf>
- Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14(2), 174–194. <https://doi.org/10.2307/2527783>
- Hartley, H. O., & Hocking, R. R. (1971). The analysis of incomplete data. *Biometrics*, 27(4), 783–823. <https://doi.org/10.2307/2528820>
- Hayes, T., Usami, S., Jacobucci, R., & McArdle, J. J. (2015). Using Classification and Regression Trees (CART) and random forests to analyze attrition: Results from two simulations. *Psychology and Aging*, 30(4), 911–929. <https://doi.org/10.1037/pag0000046>
- He, Y. L., & Raghunathan, T. E. (2009). On the performance of sequential regression multiple imputation methods with non normal error distributions. *Communications in Statistics—Simulation and Computation*, 38(4), 856–883. <https://doi.org/10.1080/03610910802677191>
- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2(1), 64–78. <https://doi.org/10.1037/1082-989x.2.1.64>
- Honaker, J., King, G., & Blackwel, M. (2021). Package “Amelia” [Computer software]. <https://cran.r-project.org/web/packages/Amelia/Amelia.pdf>
- Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A potential for bias when rounding in multiple imputation. *American Statistician*, 57(4), 229–232. <https://doi.org/10.1198/0003130032314>
- Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research*, 50(3), 285–299. <https://doi.org/10.1080/00273171.2014.999267>
- Huang, L., Chen, M. H., & Ibrahim, J. G. (2005). Bayesian Analysis for generalized linear models with nonignorably missing covariates. *Biometrics*, 61(3), 767–780. <https://doi.org/10.1111/j.1541-0420.2005.00338.x>

- Humbert, S., & Grund, S. (2022). Response surface analysis with missing data. *Multivariate Behavioral Research*, 57(4), 581–602. <https://doi.org/10.1080/00273171.2021.1884522>
- Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, 85(411), 765–769. <https://doi.org/10.1080/01621459.1990.10474938>
- Ibrahim, J. G., Chen, M.-H., & Lipsitz, S. R. (2002). Bayesian methods for generalized linear models with covariates missing at random. *Canadian Journal of Statistics*, 30(1), 55–78. <https://doi.org/10.2307/3315865>
- Ibrahim, J. G., Lipsitz, S. R., & Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1), 173–190. <https://doi.org/10.1111/1467-9868.00170>
- Johnson, V. E., & Albert, J. H. (1999). *Ordinal data modeling*. Springer.
- Jöreskog, K. G., & Sörbom, D. (2018). *LISREL 10 for Windows* [Computer software]. Scientific Software International.
- Jorgensen, T. D. (2021). *Package “semTools”* [Computer software]. <https://cran.r-project.org/web/packages/semTools/semTools.pdf>
- Keller, B. T. (2015). *Three-level multiple imputation: A fully conditional specification approach* [Master's thesis]. Arizona State University, Proquest.
- Keller, B. T., & Lenders, C. K. (2021). *Blimp user's guide* (Version 3). <https://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*. Advance online publication. <https://doi.org/10.1080/00273171.2022.2147049>
- Kenward, M. G., & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, 13(3), 236–247. <https://doi.org/10.1214/ss/1028905886>
- Kim, S., Belin, T. R., & Sugar, C. A. (2018). Multiple imputation with non-additively related variables: Joint-modeling and approximations. *Statistical Methods in Medical Research*, 27(6), 1683–1694. <https://doi.org/10.1177/0962280216667763>
- Kim, S., Sugar, C. A., & Belin, T. R. (2015). Evaluating model-based imputation methods for missing covariates in regression models with interactions. *Statistics in Medicine*, 34(11), 1876–1888. <https://doi.org/10.1002/sim.6435>
- Kleinke, K. (2017). Multiple imputation under violated distributional assumptions: A systematic evaluation of the assumed robustness of predictive mean matching. *Journal of Educational and Behavioral Statistics*, 42(4), 371–404. <https://doi.org/10.3102/1076998616687084>
- Kleinke, K., & Reinecke, J. (2013). Multiple imputation of incomplete zero-inflated count data. *Statistica Neerlandica*, 67(3), 311–336. <https://doi.org/10.1111/stan.12009>
- Lai, K. (2021). Correct estimation methods for RMSEA under missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(2), 207–218. <https://doi.org/10.1080/10705511.2020.1755864>
- Lawes, M., & Eid, M. (2022). Factor score estimation in multimethod measurement designs with planned missing data. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000483>
- Lee, K. J., & Carlin, J. B. (2017). Multiple imputation in the presence of non-normal data. *Statistics in Medicine*, 36(4), 606–617. <https://doi.org/10.1002/sim.7173>
- Lee, M. C., & Mitra, R. (2016). Multiply imputing missing values in data sets with mixed measurement scales using a sequence of generalised linear models. *Computational Statistics & Data Analysis*, 95, 24–38. <https://doi.org/10.1016/j.csda.2015.08.004>
- Lee, S.-Y., & Shi, J.-Q. (2000). Joint Bayesian analysis of factor scores and structural parameters in the factor analysis model. *Annals of the Institute of Statistical Mathematics*, 52(4), 722–736. <https://doi.org/10.1023/A:1017529427433>
- Lee, S.-Y., Song, X.-Y., & Tang, N.-S. (2007). Bayesian Methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 404–434. <https://doi.org/10.1080/10705510701301511>
- Lee, T., & Cai, L. (2012). Alternative multiple imputation inference for mean and covariance structure modeling. *Journal of Educational and Behavioral Statistics*, 37(6), 675–702. <https://doi.org/10.3102/1076998612458320>
- Lee, T., & Shi, D. (2021). A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data. *Psychological Methods*, 26(4), 466–485. <https://doi.org/10.1037/met0000381>
- Levy, R., & Enders, C. (2021). Full conditional distributions for Bayesian multilevel models with additive or interactive effects and missing data on covariates. *Communications in Statistics—Simulation and Computation*. Advance online publication. <https://doi.org/10.1080/03610918.2021.1921799>
- Levy, R., & Mislevy, R. J. (2016). *Bayesian psychometric modeling*. CRC Press.
- Licht, C. (2010). *New methods for generating significance levels from multiply-imputed data* [Doctoral Dissertation]. Universität Bamberg.
- Lipsitz, S. R., & Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika*, 83(4), 916–922. <https://doi.org/10.1093/biomet/83.4.916>
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. Wiley.
- Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd ed.). Wiley.
- Liu, J. C., Gelman, A., Hill, J., Su, Y.-S., & Kropko, J. (2014). On the stationary distribution of iterative imputations. *Biometrika*, 101(1), 155–173. <https://doi.org/10.1093/biomet/ast044>
- Liu, Y., & Enders, C. K. (2017). Evaluation of multi-parameter test statistics for multiple imputation. *Multivariate Behavioral Research*, 52(3), 371–390. <https://doi.org/10.1080/00273171.2017.1298432>
- Liu, Y., & Sriutaisuk, S. (2020). Evaluation of model fit in structural equation models with ordinal missing data: An examination of the D₂ method. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(4), 561–583. <https://doi.org/10.1080/10705511.2019.1662307>
- Liu, Y., Sriutaisuk, S., & Chung, S. (2021). Evaluation of model fit in structural equation models with ordinal missing data: A comparison of the D₂ and MI2S methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(5), 740–762. <https://doi.org/10.1080/10705511.2021.1919118>
- Loncke, J., Eichelsheim, V. I., Branje, S. J. T., Buysse, A., Meeus, W. H. J., & Loeys, T. (2018). Factor score regression with social relations model components: A case study exploring antecedents and consequences of perceived support in families. *Frontiers in Psychology*, 9, 1699. <https://doi.org/10.3389/fpsyg.2018.01699>
- Lord, F. M. (1955). Estimation of parameters from incomplete data. *Journal of the American Statistical Association*, 50(271), 870–876. <https://doi.org/10.2307/2281171>
- Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, 22(1), 141–165. <https://doi.org/10.1037/met0000096>
- Lüdtke, O., Robitzsch, A., & West, S. G. (2020a). Analysis of interactions and nonlinear effects with missing data: A factored regression modeling approach using maximum likelihood estimation. *Multivariate Behavioral Research*, 55(3), 361–381. <https://doi.org/10.1080/00273171.2019.1640104>
- Lüdtke, O., Robitzsch, A., & West, S. G. (2020b). Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation. *Psychological Methods*, 25(2), 157–181. <https://doi.org/10.1037/met0000233>
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other

- intervening variable effects. *Psychological Methods*, 7(1), 83–104. <https://doi.org/10.1037/1082-989X.7.1.83>
- Manski, C. F. (2005). Partial identification with missing data: Concepts and findings. *International Journal of Approximate Reasoning*, 39(2–3), 151–165. <https://doi.org/10.1016/j.ijar.2004.10.006>
- Manski, C. F. (2016). Credible interval estimates for official statistics with survey nonresponse. *Journal of Econometrics*, 191(2), 293–301. <https://doi.org/10.1016/j.jeconom.2015.12.002>
- Manski, C. F. (2022). *Inference with imputed data: The allure of making stuff up*. <https://arxiv.org/abs/2205.07388>
- Mansolf, M., Jorgensen, T. D., & Enders, C. K. (2020). A multiple imputation score test for model modification in structural equation models. *Psychological Methods*, 25(4), 393–411. <https://doi.org/10.1037/met0000243>
- Mazza, G. L., Enders, C. K., & Ruehlman, L. S. (2015). Addressing item-level missing data: A comparison of prorating and full information maximum likelihood estimation. *Multivariate Behavioral Research*, 50(5), 504–519. <https://doi.org/10.1080/00273171.2015.1068157>
- McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52(6), 2287–2305. <https://doi.org/10.3758/s13428-020-01398-0>
- Mealli, F., & Rubin, D. B. (2016). Clarifying missing at random and related definitions, and implications when coupled with exchangeability. *Biometrika*, 103(2), 491–491. <https://doi.org/10.1093/biomet/asw017>
- Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, 9(4), 538–558. <https://www.jstor.org/stable/2246252>
- Meng, X.-L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79(1), 103–111. <https://doi.org/10.2307/2337151>
- Merkle, E. C., & Rosseel, Y. (2018). Blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, 85(4), 1–30. <https://doi.org/10.18637/jss.v085.i04>
- Mistler, S. A., & Enders, C. K. (2017). A comparison of joint model and fully conditional specification imputation for multilevel missing data. *Journal of Educational and Behavioral Statistics*, 42(4), 432–466. <https://doi.org/10.3102/1076998617690869>
- Mohan, K., & Pearl, J. (2021). Graphical models for processing missing data. *Journal of the American Statistical Association*, 116(534), 1023–1037. <https://doi.org/10.1080/01621459.2021.1874961>
- Mohan, K., Pearl, J., & Tian, J. (2013). Graphical models for inference with missing data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 1277–1285). Curran Associates
- Moreno-Betancur, M., Lee, K. J., Leacy, F. P., White, I. R., Simpson, J. A., & Carlin, J. B. (2018). Canonical causal diagrams to guide the treatment of missing data in epidemiologic studies. *American Journal of Epidemiology*, 187(12), 2705–2715. <https://doi.org/10.1093/aje/kwy173>
- Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, 14(1), 1–13. <https://doi.org/10.1186/1471-2288-14-75>
- Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, 16(1), 17–33. <https://doi.org/10.1037/a0022634>
- Muthén, B. O., & Muthén, L. (2016). *Regression and mediation analysis using Mplus*.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.).
- Nijman, S., Leeuwenberg, A. M., Beekers, I., Verkouter, I., Jacobs, J., Bots, M. L., Asselbergs, F. W., Moons, K., & Debray, T. (2022). Missing data is poorly handled and reported in prediction model studies using machine learning: A literature review. *Journal of Clinical Epidemiology*, 142(February), 218–229. <https://doi.org/10.1016/j.jclinepi.2021.11.023>
- Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and applications. In *Proceedings from the sixth Berkeley symposium on mathematical statistics and probability: Theory of statistics* (Vol. 1, pp. 697–715). University of California Press.
- Palomo, J., Dunson, D. B., & Bollen, K. (2007). Bayesian structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 163–188). Elsevier.
- Pearl, J., & Mohan, M. (2013, October 26). *Recoverability and testability of missing data: Introduction and summary of results*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2343873
- Plummer, M. (2019). *Package "rjags"* [Computer software]. <https://cran.r-project.org/web/packages/rjags/rjags.pdf>
- Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian Inference for logistic models using pólya-gamma latent variables. *Journal of the American Statistical Association*, 108(504), 1339–1349. <https://doi.org/10.1080/01621459.2013.829001>
- Pritikin, J. N., Brick, T. R., & Neale, M. C. (2018). Multivariate normal maximum likelihood with both ordinal and continuous variables, and data missing at random. *Behavior Research Methods*, 50(2), 490–500. <https://doi.org/10.3758/s13428-017-1011-6>
- Provost, F., & Saar-Tsechanski, M. (2007). Handling missing values when applying classification models. *Journal of Machine Learning Research*, 8, 1625–1657. <https://www.jmlr.org/papers/v8/>
- Quartagno, M., & Carpenter, J. (2020). *Package "jomo"* [Computer software]. <https://cran.r-project.org/web/packages/jomo/jomo.pdf>
- Quartagno, M., & Carpenter, J. R. (2016). Multiple imputation for IPD meta-analysis: Allowing for heterogeneity and studies with missing covariates. *Statistics in Medicine*, 35(17), 2938–2954. <https://doi.org/10.1002/sim.6837>
- Quartagno, M., & Carpenter, J. R. (2019). Multiple imputation for discrete data: Evaluation of the joint latent normal model. *Biometrical Journal*, 61(4), 1003–1019. <https://doi.org/10.1002/bimj.201800222>
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 167–190. <https://doi.org/10.1007/Bf02295939>
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128(2), 301–323. <https://doi.org/10.1016/j.jeconom.2004.08.017>
- Raghuathan, T. E., Lepkowski, J., Van Hoewyk, J. H., & Solenberger, P. W. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27(1), 85–95. <https://www150.statcan.gc.ca/n1/pub/12-001-x/2001001/article/5857-eng.pdf>
- Raudenbush, S. W. (2019). *HLM for Windows* [Computer software]. Scientific Software International.
- Raykov, T., & West, B. T. (2016). On enhancing plausibility of the missing at random assumption in incomplete data analyses via evaluation of response-auxiliary variable correlations. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 45–53. <https://doi.org/10.1080/10705511.2014.937848>
- Reiter, J. P. (2007). Small-sample degrees of freedom for multi-component significance tests with multiple imputation for missing data. *Biometrika*, 94(2), 502–508. <https://doi.org/10.1093/biomet/asm028>
- Reiter, J. P., Raghuathan, T. E., & Kinney, S. K. (2006). The importance of modeling the survey design in multiple imputation for missing data. *Survey Methodology*, 32(2), 143–150. <https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X20060029548>
- Resche-Rigon, M., & White, I. R. (2018). Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Statistical Methods in Medical Research*, 27(6), 1634–1649. <https://doi.org/10.1177/0962280216666564>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous

- and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
- Rockwood, N. J. (2020). Maximum likelihood estimation of multilevel structural equation models with random slopes for latent covariates. *Psychometrika*, 85(2), 275–300. <https://doi.org/10.1007/s11336-020-09702-9>
- Rockwood, N. J., & Jeon, M. (2019). Estimating complex measurement and growth models using the R package PLmixed. *Multivariate Behavioral Research*, 54(2), 288–306. <https://doi.org/10.1080/00273171.2018.1516541>
- Rosseel, Y. (2012). Lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473–489. <https://doi.org/10.1080/01621459.1996.10476908>
- SAS Institute Inc. (2011). *SAS/STAT® 9.3 user's guide*. SAS Institute
- Savalei, V. (2010a). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, 15(4), 352–367. <https://doi.org/10.1037/a0020143>
- Savalei, V. (2010b). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 241–264. <https://doi.org/10.1080/10705511003659375>
- Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 477–497. <https://doi.org/10.1080/10705510903008238>
- Savalei, V., & Falk, C. F. (2014). Robust two-stage approach outperforms robust full information maximum likelihood with incomplete nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 280–302. <https://doi.org/10.1080/10705511.2014.882692>
- Savalei, V., & Rhemtulla, M. (2017). Normal theory two-stage estimator for models with composites when data are missing at the item level. *Journal of Educational and Behavioral Statistics*, 42(4), 405–431. <https://doi.org/10.3102/1076998617694880>
- Savalei, V., & Rosseel, Y. (2022). Computational options for standard errors and test statistics with incomplete normal and nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(2), 163–181. <https://doi.org/10.1080/10705511.2021.1877548>
- Savalei, V., & Yuan, K. H. (2009). On the model-based bootstrap with missing data: Obtaining a p-value for a test of exact fit. *Multivariate Behavioral Research*, 44(6), 741–763. <https://doi.org/10.1080/00273170903333590>
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall.
- Schafer, J. L. (2001). Multiple imputation with PAN. In A. G. Sayer & L. M. Collins (Eds.), *New methods for the analysis of change* (pp. 355–377). American Psychological Association.
- Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, 57(1), 19–35. <https://doi.org/10.1111/1467-9574.00218>
- Schafer, J. L. (2018). *Package "pan"* [Computer software]. <https://cran.r-project.org/web/packages/pan/pan.pdf>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33(4), 545–571. https://doi.org/10.1207/s15327906mbr3304_5
- Schafer, J. L., & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, 11(2), 437–457. <https://doi.org/10.1198/106186002760180608>
- Seaman, S., Galati, J., Jackson, D., & Carlin, J. (2013). What is meant by “missing at random”? *Statistical Science*, 28(2), 257–268. <https://doi.org/10.1214/13-Sts415>
- Seaman, S. R., Bartlett, J. W., & White, I. R. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: An evaluation of statistical methods. *BMC Medical Research Methodology*, 12(1), 1–13. <https://doi.org/10.1186/1471-2288-12-46>
- Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*, 179(6), 764–774. <https://doi.org/10.1093/aje/kwt312>
- Shin, Y., & Raudenbush, S. W. (2007). Just-identified versus overidentified two-level hierarchical linear models with missing data. *Biometrics*, 63(4), 1262–1268. <https://doi.org/10.1111/j.1541-0420.2007.00818.x>
- Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics*, 35(1), 26–53. <https://doi.org/10.3102/1076998609345252>
- Shin, Y., & Raudenbush, S. W. (2013). Efficient analysis of Q-level nested hierarchical general linear models given ignorable missing data. *International Journal of Biostatistics*, 9(1), 109–133. <https://doi.org/10.1515/ijb-2012-0048>
- Sijtsma, K., & van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, 38(4), 505–528. https://doi.org/10.1207/s15327906mbr3804_4
- Sterba, S. K., & Gottfredson, N. C. (2015). Diagnosing global case influence on MAR versus MNAR model comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2), 294–307. <https://doi.org/10.1080/10705511.2014.936082>
- Sturtz, S., Ligges, U., & Gelman, A. (2019). *R2OpenBUGS: A package for running OpenBUGS from R* [Computer software]. <https://cran.r-project.org/web/packages/R2OpenBUGS/vignettes/R2OpenBUGS.pdf>
- Su, Y.-S., Gelman, A., Hill, J., & Yajima, M. (2011). Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. *Journal of Statistical Software*, 45(2), 1–31. <https://doi.org/10.18637/jss.v045.i02>
- Takai, K., & Kano, Y. (2013). Asymptotic inference with incomplete data. *Communications in Statistics—Theory and Methods*, 42(17), 3174–3190. <https://doi.org/10.1080/03610926.2011.621577>
- Thoemmes, F., & Mohan, K. (2015). Graphical representation of missing data problems. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 631–642. <https://doi.org/10.1080/10705511.2014.937378>
- Thoemmes, F., & Rose, N. (2014). A cautious note on auxiliary variables that can increase bias in missing data problems. *Multivariate Behavioral Research*, 49(5), 443–459. <https://doi.org/10.1080/00273171.2014.931799>
- Thomas, R. M., Bruin, W., Zhutovsky, P., & Wingen, G. (2020). Dealing with missing data, small sample sizes, and heterogeneity in machine learning studies of brain disorders. In A. Mechelli & S. Viera (Eds.), *Methods and applications to brain disorders* (pp. 249–266). Academic Press.
- Umbach, N., Naumann, K., Hoppe, D., Brandt, H., Kelava, A., & Schmitz, B. (2017). *Package "nlsem"* [Computer software]. <https://cran.r-project.org/web/packages/nlsem/nlsem.pdf>
- van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 16(3), 219–242. <https://doi.org/10.1177/0962280206074463>
- van Buuren, S. (2010). Item imputation without specifying scale structure. *Methodology*, 6(1), 31–36. <https://doi.org/10.1027/1614-2241/a000004>

- van Buuren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 173–196). Routledge.
- van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman and Hall.
- van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049–1064. <https://doi.org/10.1080/10629360600810434>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2), 217–239. <https://doi.org/10.1037/met0000100>
- van Ginkel, J. R., Sijtsma, K., van der Ark, L. A., & Vermunt, J. K. (2010). Incidence of missing item scores in personality measurement, and simple item-score imputation. *Methodology*, 6(1), 17–30. <https://doi.org/10.1027/1614-2241/a000003>
- van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2007). Multiple imputation for item scores when test data are factorially complex. *British Journal of Mathematical and Statistical Psychology*, 60(2), 315–337. <https://doi.org/10.1348/000711006X117574>
- Vink, G., Frank, L. E., Pannekoek, J., & van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, 68(1), 61–90. <https://doi.org/10.1111/stan.12023>
- von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology*, 39(1), 265–291. <https://doi.org/10.1111/j.1467-9531.2009.01215.x>
- von Hippel, P. T. (2013). Should a normal imputation model be modified to impute skewed variables? *Sociological Methods and Research*, 42(1), 105–138. <https://doi.org/10.1177/0049124112464866>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E. J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., ... Morey, R. D. (2018). Bayesian Inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review*, 25(1), 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Widaman, K. F., & Revelle, W. (2022). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-01849-w>
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114–128. <https://doi.org/10.1016/j.stueduc.2005.05.005>
- Wu, M. C., & Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1), 175–188. <https://doi.org/10.2307/2531905>
- Xu, D., Daniels, M. J., & Winterstein, A. G. (2016). Sequential BART for imputation of missing covariates. *Biostatistics*, 17(3), 589–602. <https://doi.org/10.1093/biostatistics/kxw009>
- Xu, S., & Blozis, S. A. (2011). Sensitivity analysis of mixed models for incomplete longitudinal data. *Journal of Educational and Behavioral Statistics*, 36(2), 237–256. <https://doi.org/10.3102/1076998610375836>
- Yang, M., & Maxwell, S. E. (2014). Treatment effects in randomized longitudinal trials with different types of nonignorable dropout. *Psychological Methods*, 19(2), 188–210. <https://doi.org/10.1037/a0033804>
- Yang, M., Wang, L., & Maxwell, S. E. (2015). Bias in longitudinal data analysis with missing data using typical linear mixed-effects modelling and pattern-mixture approach: An analytical illustration. *British Journal of Mathematical and Statistical Psychology*, 68(2), 246–267. <https://doi.org/10.1111/bmsp.12043>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>
- Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954–959. <https://doi.org/10.1093/biomet/87.4.954>
- Yuan, K.-H. (2009). Normal distribution based pseudo ML for missing data: With applications to mean and covariance structure analysis. *Journal of Multivariate Analysis*, 100(9), 1900–1918. <https://doi.org/10.1016/j.jmva.2009.05.001>
- Yuan, K.-H., & Bentler, P. M. (2000). 5. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30(1), 165–200. <https://doi.org/10.1111/0081-1750.00078>
- Yuan, K.-H., & Bentler, P. M. (2010). Consistency of normal distribution based pseudo maximum likelihood estimates when data are missing at random. *American Statistician*, 64(3), 263–267. <https://doi.org/10.1198/tast.2010.09203>
- Yuan, K.-H., & Savalei, V. (2014). Consistency, bias and efficiency of the normal-distribution-based MLE: The role of auxiliary variables. *Journal of Multivariate Analysis*, 124(February), 353–370. <https://doi.org/10.1016/j.jmva.2013.11.006>
- Yuan, K.-H., Tong, X., & Zhang, Z. (2014). Bias and efficiency for SEM with missing data and auxiliary variables: Two-stage robust method versus two-stage ML. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2), 178–192. <https://doi.org/10.1080/10705511.2014.935750>
- Yuan, K.-H., Yang-Wallentin, F., & Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. *Sociological Methods & Research*, 41(4), 598–629. <https://doi.org/10.1177/0049124112460373>
- Yuan, K.-H., & Zhang, Z. Y. (2012). Robust structural equation modeling with missing data and auxiliary variables. *Psychometrika*, 77(4), 803–826. <https://doi.org/10.1007/s11336-012-9282-4>
- Yucel, R. M. (2008). Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philosophical Transactions of the Royal Society A: Mathematical and Physical Sciences*, 366(1874), 2389–2403. <https://doi.org/10.1098/rsta.2008.0038>
- Yucel, R. M. (2011). Random-covariances and mixed-effects models for imputing multivariate multilevel continuous data. *Statistical Modelling*, 11(4), 351–370. <https://doi.org/10.1177/1471082X1001100404>
- Yucel, R. M., He, Y., & Zaslavsky, A. M. (2008). Using calibration to improve rounding in imputation. *The American Statistician*, 62(2), 125–129. <https://doi.org/10.1198/000313008X300912>
- Zhang, Q., & Wang, L. (2017). Moderation analysis with missing data in the predictors. *Psychological Methods*, 22(4), 649–666. <https://doi.org/10.1037/met0000104>
- Zhang, X., & Savalei, V. (2020). Examining the effect of missing data on RMSEA and CFI under normal theory full-information maximum likelihood. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(2), 219–239. <https://doi.org/10.1080/10705511.2019.1642111>
- Zhao, Y., & Long, Q. (2016). Multiple imputation in the presence of high-dimensional data. *Statistical Methods in Medical Research*, 25(5), 2021–2035. <https://doi.org/10.1177/0962280213511027>

Received March 25, 2022

Revision received November 10, 2022

Accepted November 28, 2022 ■