CHAPTER 12

Fitting Structural Equation Models with Missing Data

Craig K. Enders

Although most of the theoretical and computational underpinnings of contemporary missing data handling procedures were developed in the 1970s and 1980s (Dempster, Laird, & Rubin, 1977; Rubin, 1976, 1987), these methods didn't become a practical reality for researchers in the behavioral and social sciences until the 1990s, when SEM software packages began implementing maximum likelihood missing data estimators (Arbuckle, 1996). Multiple imputation procedures came online around the same time (Schafer, 1997) and provided a second avenue for addressing this common problem. Both approaches have matured considerably since then, and several important developments have appeared in the literature since the first edition of this *Handbook*.

The missing data literature spans many decades and describes numerous and diverse strategies (e.g., listwise and pairwise deletion, mean substitution, regression imputation). There is now broad awareness that older methods possess substantial limitations that should deter their use, and early computer simulation studies from the SEM literature unequivocally support this conclusion (Brown, 1994; Enders & Bandalos, 2001; Olinsky, Chen, & Harlow, 2003). As such, the goal for this chapter is to provide readers with an overview of the two major missing data handling frameworks that have broad support in the literature—full-information maximum likelihood estimation and multiple imputation.

I use data from a longitudinal health psychology study of 500 chronic pain sufferers to illustrate missing data handling procedures. The data include a pain severity composite, gender dummy code, and exercise frequency rating collected at the baseline assessment, a six-item measure of pain interference with daily life obtained at the first follow-up assessment, and a six-item depression measure from the final follow-up. The pain interference items used 7-point scores, and the depression questions used 4-point ratings. Most of the discrete distributions are relatively symmetrical. Figure 12.1 shows a path diagram of the primary analysis model. The pain severity composite and several questionnaire items are incomplete, with missingness rates ranging from 2.8% to 12.6%. The data and all software scripts are available on the companion website.

MISSING DATA PROCESSES

Rubin and colleagues (Little & Rubin, 1987; Rubin, 1976) introduced a classification system for missing data problems that describes three ways in which missingness can relate to the data. A missing completely at random (MCAR) mechanism states that the probability

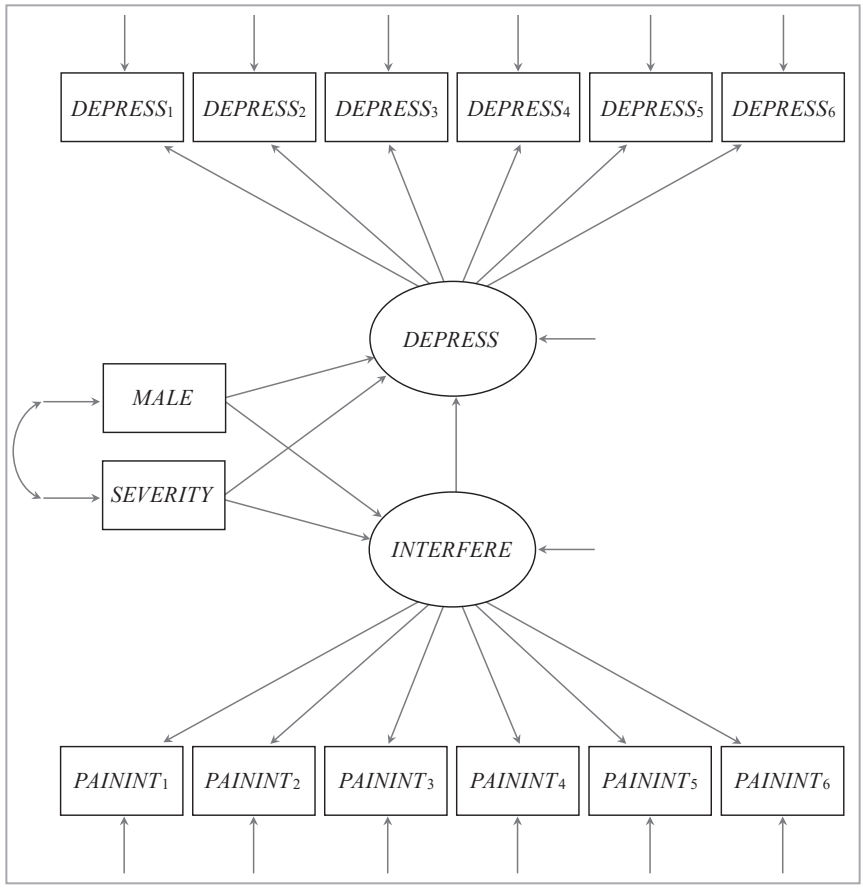

FIGURE 12.1. Structural equation model for the working example.

of missing values is unrelated to both the observed and unseen parts of the realized data set, a missing at random (MAR) process (also called conditionally missing at random; Graham, 2009) posits that missingness is related to the observed parts of the data only, and a missing not at random (MNAR) mechanism states that the probability of missing values is related to the missing parts of the data (and potentially to the observed parts as well). A variety of resources provide detailed descriptions of these mechanisms, which effectively function as assumptions for a missing data analysis (Enders, 2022; Little & Rubin, 2020; Raykov, 2011; Schafer & Graham, 2002).

An MAR process—the default assumption for most maximum likelihood and multiple imputation applications—requires that missingness is purely random after conditioning on the observed data. Adopting a so-called "inclusive" analysis strategy that introduces auxiliary variables into the analysis or imputation model can make this assumption more plausible (Collins, Schafer, & Kam, 2001). One strategy is to identify a small number of prioritized potent auxiliary variables based on the magnitude of their *semipartial* (residual) correlations with the analysis variables; external variables that lack this unique covariation cannot introduce bias, nor can they enhance precision. Raykov and West (2015) described a latent variable modeling approach to estimating these crucial associations. Another strategy is to simply cast a wide net and use principal components analysis to reduce a large set of auxiliary

variables into a smaller number of linear composites (Howard, Rhemtulla, & Little, 2015). In practice, it is often difficult to find more than a few potent auxiliary variables because the focal analysis model taps into the largest reservoir of correlation in the data. Nevertheless, good auxiliary variables can be very useful, if you can find them.

FULL-INFORMATION MAXIMUM LIKELIHOOD

For researchers in the social and behavioral sciences, maximum likelihood estimation became a practical reality in the 1990s when SEM software packages began implementing full-information maximum likelihood (FIML) estimators based on raw rather than summary data (Arbuckle, 1996). These estimators have evolved in recent years, and flexible routines that accommodate mixtures of categorical and continuous variables are widely available (Lüdtke, Robitzsch, & West, 2020a; Muthén, Muthén, & Asparouhov, 2016; Pritikin, Brick, & Neale, 2018).

The classic estimator for multivariate normal data is a good starting point because the normal model is common in published applications, and it performs well in a wide range of situations. Moreover, treating ordered categorical indicators like those from the chronic pain data as normal has support from the literature if items have five or more response options and symmetrical distributions (Rhemtulla, Brosseau-Liard, & Savalei, 2012); the 7-point pain interference ratings are a better approximation to the normal ideal. I sketch the details of normal-theory maximum likelihood estimation and direct interested readers to Savalei and Rosseel (2022) for an in-depth treatment.

To provide a working example, consider a simple factor model with a single latent variable and three indicators. Using generic notation, the measurement and structural models are

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \end{pmatrix} = \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \upsilon_3 \end{pmatrix} + \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \eta_i + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{pmatrix} = \mathbf{Y}_i = \boldsymbol{\upsilon} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon} \tag{12.1}$$
$$\eta_i \sim N_1(\alpha, \psi) \qquad \boldsymbol{\varepsilon} \sim N_3(0, \boldsymbol{\Theta})$$

where the tilde means "distributed as," N_1 and N_3 denote univariate and trivariate normal distribution functions, respectively, α and ψ are the latent variable's mean and variance, respectively, and $\boldsymbol{\Theta}$ is a diagonal matrix containing the residual variances. The factor model parameters combine to produce the following predictions about the population means and variance–covariance matrix, $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, respectively.

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{pmatrix} \mu_1(\boldsymbol{\theta}) \\ \mu_2(\boldsymbol{\theta}) \\ \mu_3(\boldsymbol{\theta}) \end{pmatrix} = \boldsymbol{\Lambda}\alpha + \boldsymbol{\upsilon}$$
$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma_1^2(\boldsymbol{\theta}) & \sigma_{12}(\boldsymbol{\theta}) & \sigma_{13}(\boldsymbol{\theta}) \\ \sigma_{21}(\boldsymbol{\theta}) & \sigma_2^2(\boldsymbol{\theta}) & \sigma_{23}(\boldsymbol{\theta}) \\ \sigma_{31}(\boldsymbol{\theta}) & \sigma_{32}(\boldsymbol{\theta}) & \sigma_3^2(\boldsymbol{\theta}) \end{pmatrix} = \boldsymbol{\Lambda}\psi\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{12.2}$$

With complete data, taking the natural logarithm of the multivariate normal distribution equation and summing over the N data records gives a log-likelihood function that that summarizes the data's evidence about a specific combination of parameter values in $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. When a participant has missing values, the observed data for that individual no longer contain information about every model parameter. The log-likelihood equation accommodates this feature by eliminating the elements in the data and parameter arrays that correspond to the missing variables. The resulting observed-data log-likelihood function is

$$\ell_{(\text{obs})} = -\sum_{i=1}^{N} \frac{V_i}{2} \ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N} \ln\left|\boldsymbol{\Sigma}_i(\boldsymbol{\theta})\right| - \frac{1}{2}\sum_{i=1}^{N} \left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta})\right)' \boldsymbol{\Sigma}_i(\boldsymbol{\theta})^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta})\right) \tag{12.3}$$

where $\mathbf{Y}_i$ contains participant i's observed data, $\mathbf{V}_i$ is the number of scores in the data vector, and $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ contain the subset of parameters in $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ that correspond to the observed variables in $\mathbf{Y}_i$. The goal of estimation is to identify the parameter values that maximize fit to the observed data (or equivalently, minimize the sum of the squared standardized deviations in the rightmost term). Iterative optimizers such as Newton's algorithm or the expectation maximization (EM) algorithm are used for this purpose.

The log-likelihood equation says that all participants share the same model parameters, but the evidence contained in a person's observed data—and thus contribution to estimation—is restricted to parameters for which there are data. To illustrate this idea, consider participants missing Y_2. The sum of squared standard-

ized deviations between the observed data and model-implied means (the rightmost term in the log-likelihood expression) is as follows:

$$\left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta})\right)' \boldsymbol{\Sigma}_i(\boldsymbol{\theta})^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta})\right) = \left(\begin{pmatrix} Y_{1i} \\ Y_{3i} \end{pmatrix} - \begin{pmatrix} \mu_1(\boldsymbol{\theta}) \\ \mu_3(\boldsymbol{\theta}) \end{pmatrix}\right)' \begin{pmatrix} \sigma_1^2(\boldsymbol{\theta}) & \sigma_{13}(\boldsymbol{\theta}) \\ \sigma_{31}(\boldsymbol{\theta}) & \sigma_3^2(\boldsymbol{\theta}) \end{pmatrix}^{-1} \left(\begin{pmatrix} Y_{1i} \\ Y_{3i} \end{pmatrix} - \begin{pmatrix} \mu_1(\boldsymbol{\theta}) \\ \mu_3(\boldsymbol{\theta}) \end{pmatrix}\right) \quad (12.4)$$

Similarly, the corresponding expression for individuals missing Y_1 and Y_3 simplifies to

$$\left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta})\right)' \boldsymbol{\Sigma}_i(\boldsymbol{\theta})^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta})\right) = \frac{\left(Y_{2i} - \mu_2(\boldsymbol{\theta})\right)^2}{\sigma_2^2(\boldsymbol{\theta})} \quad (12.5)$$

Mathematically, these equations result from marginalizing over (i.e., integrating out) the distribution of missing values; marginalizing over the missing Y_2 scores in the trivariate normal distribution gives a bivariate normal distribution function that includes Equation 12.4, and averaging over both Y_1 and Y_3 returns a univariate normal distribution function that features Equation 12.5.

The previous expressions highlight that the maximum likelihood estimation does not discard incomplete data records, nor does it impute them. Rather, the estimator identifies the parameter values with maximum support from whatever data are available. However, the normal curve does function like an imputation machine in the sense that the estimator can infer the location of an unseen data point from a person's observed scores; conceptually, marginalizing over the missing data is akin to replacing each unseen score with a weighted sum over all possible values of the missing variable, such that higher weights are assigned to score values that are more plausible given a person's observed data, and vice versa. Thus, while the estimator doesn't literally create a filled-in data set, it does use the normal distribution to deduce plausible values for the missing data. Widaman (2006) aptly describes this process as *implicit imputation*.

Manifest Exogenous Variables

An important requirement for parametric missing data handling is that all incomplete variables must have a distribution. This is always true for outcomes but not necessarily true for manifest predictor variables such as pain severity and gender. The easiest strategy is to assume that manifest regressors follow the same multivariate normal distribution as the manifest outcomes (i.e., assign all variables to the $\mathbf{Y}_i$ vector of Equation 12.3 regardless of their role in the analysis). Figure 12.2(a) depicts this specification; the residual arrows

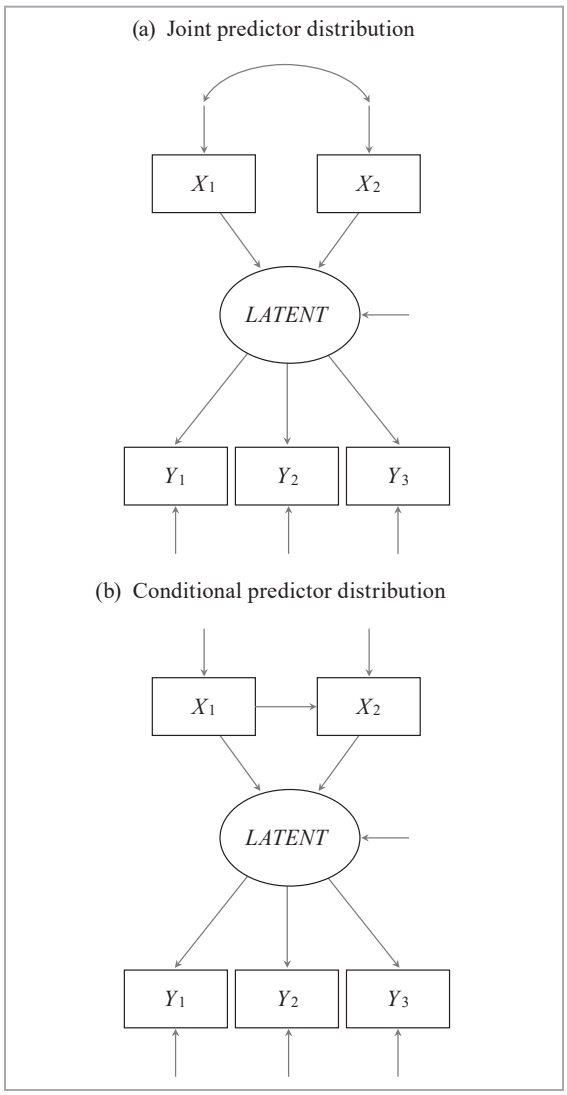

FIGURE 12.2. Two specifications for linking manifest exogenous variables.

pointing to the predictors convey that these variables have a distribution, and correlating the residuals establishes the structural path coefficients as partial slopes. Folding manifest regressors into the multivariate normal distribution is seemingly limited to continuous covariates, but computer simulations suggest that applying this strategy to incomplete binary predictors usually doesn't introduce bias (Muthén et al., 2016). Having logged many hours analyzing incomplete data sets, I usually feel comfortable treating binary covariates as normal because it is rare to find situations where this misspecification meaningfully affects the results (Enders, 2022). Of course, this specification is nonsensical for multicategorical nominal predictors.

A second strategy for assigning distributions to incomplete predictors is to regress the covariates on each other. Figure 12.2(b) depicts this specification; the residual arrows pointing to the predictors again convey that these variables have a distribution, but the model uses a directed pathway to preserve their association. The primary advantage of this specification is that covariates need not be normal. For example, if both predictors in the figure were binary, X_1's distribution could be modeled with an empty probit or logistic model, and the regression of X_2 on X_1 would similarly be a categorical variable model. Instead of working from a single multivariate distribution, estimators for mixed response types generally disassemble a model into multiple parts that rely on different probability distributions and log-likelihood functions. In many situations, there is no obvious way to define a saturated or restricted model that can be used to construct model fit statistics (beyond information criteria such as the Akaike information criterion [AIC] and Bayesian information criterion [BIC]), and standardized coefficients or modification indices may be unavailable as well.

Test Statistics, Standard Errors, and Non-Normal Data

Maximum likelihood estimation offers three significance testing options: the Wald test, likelihood ratio statistic, and the score test (i.e., modification index). All three procedures are applicable to missing data, and their details are largely the same with or without missing values. Similarly, familiar corrective procedures such as rescaled test statistics, robust (sandwich estimator) standard errors, and the bootstrap have long been available for missing data (Savalei & Yuan, 2009; Yuan & Bentler, 2000), and a good deal of literature supports their use. The one issue to be aware of is that missing data standard errors should be computed using the observed information matrix (i.e., a second derivative matrix that incorporates the observed data), as standard errors based on expected information require an MCAR process. Kenward and Molenberghs (1998) and Savalei (2010) provide additional details about this issue.

Auxiliary Variables

Researchers can use one of two strategies to incorporate extra auxiliary variables into a maximum likelihood analysis. Graham (2003) outlined so-called "saturated" correlates and extra dependent variable specifications that use a particular configuration of residual correlations and regression slopes to connect the auxiliary variables to the focal analysis model. Two-stage estimation is an alternative approach (Savalei & Bentler, 2009) that tackles the missing data in two steps. The first stage estimates the mean vector and variance–covariance matrix of a superset that includes the analysis variables and auxiliary variables, and the second stage uses a subset of these summary statistics as input data for the focal structural model. I focus on Graham's approach because it is easy to implement across software packages.

The saturated correlates model uses a series of correlations and residual correlations to connect auxiliary variables to the focal variables and each other. Graham's (2003) rules for implementing the strategy are (1) correlate each auxiliary variable with all explanatory variables, (2) correlate each auxiliary variable with all other auxiliary variables, (3) correlate each auxiliary variable with the residual terms of all manifest outcome variables, and (4) *do not* correlate auxiliary variables with latent variables or their residual terms. Figure 12.3(a) shows a path diagram of the saturated correlates specification for a factor model with three indicators, an exogenous covariate, and two auxiliary variables. Importantly, the presence of extra variables does not alter the interpretation of the primary model parameters or affect the model's degrees of freedom.

The extra dependent variable specification instead treats auxiliary variables as additional outcomes. The rules for constructing this model are somewhat different: (1) Each auxiliary variable is regressed on all explanatory variables (manifest or latent), (2) each auxiliary variable connects to each outcome (manifest or latent) via a correlated residual, and (3) each pair of

auxiliary variables connects via a correlated residual. Figure 12.3(b) shows a path diagram of the extra outcome specification. Once again, the presence of the extra variables does not alter the interpretation of the primary model parameters, although this option does contribute additional degrees of freedom because it indirectly connects the auxiliary variables to the indicators via the latent variable. The extra dependent variable specification is useful because the saturated correlates model is prone to convergence failures (Howard et al., 2015).

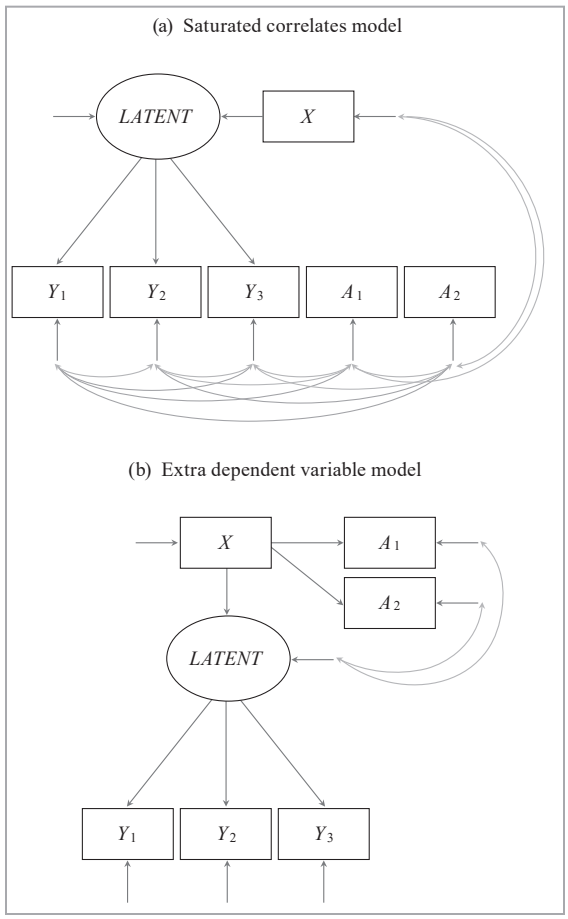

FIGURE 12.3. Two model specifications for auxiliary variables.

ESTIMATORS FOR CATEGORICAL DATA

Maximum likelihood estimators have evolved in recent years, and flexible routines that accommodate mixtures of categorical and continuous variables are widely available (Lüdtke et al., 2020a; Muthén et al., 2016; Pritikin et al., 2018). Adopting a latent response variable formulation with a probit link function is a natural choice for the ordered categorical indicators in Figure 12.1. As detailed in Chapter 15 (Koziol, this volume), the probit model views each questionnaire item as arising from a normally distributed latent response variable, the distribution of which is separated into discrete segments by a set of threshold parameters. The resulting measurement model describes the correlation structure of these latent response variables (i.e., polychoric correlations) rather than the variances and covariances of the discrete indicators.

All things being equal, you might expect that full-information estimators for categorical data are preferable because they are theoretically more correct, but this is not necessarily true. For one, full-information estimation for item-level factor analysis is often restricted to models with relatively few latent factors and indicators (e.g., 20 or fewer latent response variables; Pritikin et al., 2018). Additionally, estimators for mixed response types generally disassemble a model into multiple parts that rely on different probability distributions and log-likelihood expressions. Often there is no obvious way to define a saturated or restricted model that can be used to construct model fit statistics (although AIC and BIC are always options). When an unrestricted model does exist, it is usually too complex to estimate. For example, consider a simple single-factor latent variable model with the six, 7-point pain interference items as indicators. The corresponding saturated or unrestricted model is a massive multivariate contingency table with $7^6 = 117{,}649$ cells rather than the usual sample mean vector and variance–covariance matrix. The multivariate contingency table for a two-factor model with 12 ordinal items is intractably large and inestimable. Unfortunately, the absence of a saturated model rules out global fit assessments.

Weighted or diagonally weighted least squares are alternatives to full-information estimations that also target latent variable associations. Although conceptually similar to maximum likelihood, weighted least squares is a limited information estimator that works from bivariate contingency tables (and associations) rather than high-dimensional multivariate data (see

Chen, Moustaki, & Zhang, Chapter 8, this volume). Estimation happens in two steps. The first step estimates threshold parameters and the polychoric correlation for each pair of latent response variables, and the second step engages an iterative optimization routine that minimizes the sum of the squared standardized differences between the unrestricted model's estimates from the first step and the thresholds and correlations predicted by the model. Importantly, weighted least squares assumes that scores are MCAR, as the first stage uses pairwise deletion to estimate the thresholds and polychoric correlations from the sample data. Most researchers are uncomfortable assuming purely haphazard missingness, rendering the estimator a nonstarter. Multiply imputing the discrete data before deploying weighted least squares is a simple fix.

MULTIPLE IMPUTATION

A typical application of multiple imputation consists of three major steps. The first is to specify an imputation model and deploy a Bayesian Markov Chain Monte Carlo (MCMC) algorithm that creates several copies of the data (e.g., 20 or more), each containing different estimates of the missing values. This step typically deploys a saturated model that places no restrictions on the sample means and variance–covariance matrix, although it is also possible to tailor imputations around a specific structural model (e.g., so-called H0 imputation; Asparouhov & Muthén, 2010). The second step is to fit a structural equation model to each filled-in data set, and the final step uses "Rubin's rules" (Little & Rubin, 2020; Rubin, 1987) to combine the imputation-specific estimates and standard errors into a single package of results. I restrict the discussion to joint model imputation (Schafer, 1997) and fully conditional specification (van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006) because these are arguably the predominant approaches in published applications.

Why Impute?

In the not-so-distant past, multiple imputation was more limiting because procedures for assessing model fit were not available. Researchers now have the full complement of tools necessary to carry out SEM analyses with multiply imputed data sets, and there are compelling reasons to do so. Perhaps the greatest strength of contemporary imputation procedures is their flexibility with mixtures of categorical and continuous variables. As noted earlier, FIML estimators for mixed response types tend to come with important compromises (e.g., no saturated model for global fit assessments, computationally intensive numerical or Monte Carlo integration techniques, no modification indices). Multiply imputing the data prior to fitting a structural equation model solves virtually all these problems, and it also provides way to marry the conditionally MAR assumption with weighted least squares estimation. Finally, multiple imputation usually tolerates larger numbers of auxiliary variables than maximum likelihood, where Graham's (2003) specification often induces convergence failures.

Imputation Step

Joint model imputation derives from applying a multivariate distribution to a set of incomplete variables. Joe Schafer's (1997) seminal text is arguably responsible for popularizing multiple imputation based on the multivariate normal distribution, and contemporary variants of this framework use a latent response formulation to accommodate a range of discrete metrics (Asparouhov & Muthén, 2021b; Carpenter & Kenward, 2013). Fully conditional specification (also known as chained equations imputation) instead uses a round robin sequence of univariate models where each incomplete variable is regressed on all other variables (complete or previously imputed). The framework was popularized by Stef van Buuren and his R package MICE (van Buuren, 2012; van Buuren & Groothuis–Oudshoorn, 2011), and recent extensions also work directly with latent response variables (Keller & Enders, 2021).

To provide a working example, consider a single-factor analysis model with three ordered categorical indicators, a continuous covariate predicting the latent variable, and two continuous auxiliary variables (see Figure 12.3 for the corresponding FIML model). Joint imputation invokes a saturated model that posits a multivariate normal distribution for the continuous and latent response variables. The model for the working example corresponds to a set of empty regressions with correlated residuals.

$$\begin{pmatrix} Y_{1i}^{*} \\ Y_{2i}^{*} \\ Y_{3i}^{*} \\ X_i \\ A_{1i} \\ A_{2i} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix} + \begin{pmatrix} r_{1i} \\ r_{2i} \\ r_{3i} \\ r_{4i} \\ r_{5i} \\ r_{3i} \end{pmatrix}$$

$$\begin{pmatrix} Y_{1i}^{*} \\ Y_{2i}^{*} \\ Y_{3i}^{*} \\ X_i \\ A_{1i} \\ A_{2i} \end{pmatrix} \sim N_6 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix}, \begin{pmatrix} 1 & & & & & \\ \sigma_{21} & 1 & & & & \\ \sigma_{31} & \sigma_{32} & 1 & & & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & & \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 & \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_6^2 \end{pmatrix} \right) \tag{12.6}$$

Following common notation in the literature, the asterisk superscripts denote normally distributed latent response variables. For identification, the latent variable averages and variances are fixed to 0 and 1, respectively, and the model also includes threshold parameters that divide the latent distributions into discrete regions (see Koziol, Chapter 15, this volume). The latent response formulation also accommodates multicategorical nominal variables (Carpenter & Kenward, 2013), although the multinomial model does not use thresholds.

The equivalent latent fully conditional specification (Keller & Enders, 2021) invokes a univariate regression model for each incomplete variable. For example, the imputation regression model for the first categorical indicator is a probit regression with the latent response variable regressed on all other variables, latent or manifest.

$$\begin{aligned} Y_{1i}^{*} &= \gamma_{01} + \gamma_{11}\left(Y_{2i}^{*}\right) + \gamma_{21}\left(Y_{3i}^{*}\right) + \gamma_{31}\left(X_i\right) + \gamma_{41}\left(A_{1i}\right) \\ &\quad + \gamma_{51}\left(A_{2i}\right) + r_{1i} \\ r_{1i} &\sim N_1\left(0,1\right) \end{aligned} \tag{12.7}$$

The probit model fixes the residual variance to 1 for identification, and it also incorporates threshold parameters that divide the latent response distribution into discrete segments. As a second example, the imputation regression model for X is the following linear regression.

$$\begin{aligned} X_i &= \gamma_{04} + \gamma_{14}\left(Y_{1i}^{*}\right) + \gamma_{24}\left(Y_{2i}^{*}\right) + \gamma_{34}\left(Y_{3i}^{*}\right) \\ &\quad + \gamma_{44}\left(A_{1i}\right) + \gamma_{54}\left(A_{2i}\right) + r_{4i} \\ r_{4i} &\sim N_1\left(0,\sigma_4^2\right) \end{aligned} \tag{12.8}$$

The previous models are somewhat different than the classic multivariate imputation by chained equations (MICE) specification, which applies logistic models to categorical variables and always features discrete variables on the right side of the equations (van Buuren, 2012).

To further illustrate, Figure 12.4(a) shows a path diagram of the joint imputation model from Equation 12.6, and Figure 12.4(b) shows the fully conditional specification model from Equation 12.7 (the path diagram for Equation 12.8 has the same structure but the variables rotate to different positions). I use an oval and rectangle to differentiate the latent response variable and its discrete indicator, respectively, and the broken arrow connecting the two is the link function that maps the unobserved continuum to the discrete responses (e.g., a broken arrow reflects the idea that threshold parameters categorize the latent distribution into discrete segments). The residual terms pointing to the rectangles indicate that all variables have a distribution, and the curved arrows illustrate that the variables link via covariances. To reduce visual clutter, I omit triangle symbols that sometimes denote grand means or intercepts.

The first-stage imputation model is deployed with Bayesian estimation and an MCMC algorithm (sometimes called "data augmentation") that repeatedly performs two broad operations over many iterations: (1) estimate the imputation model parameters conditional on the filled-in data and (2) update the missing values conditional on the current parameter estimates. For joint model imputation, MCMC samples missing values (including the latent response scores) at random from a multivariate normal distribution, whereas the fully conditional specification algorithm draws imputations from a sequence of univariate distributions. Conceptually, the imputation process is akin to computing a predicted score from the estimated model parameters and adding a random normal noise term, the variation of which is determined by the variance parameters. This process yields latent normal imputations for the ordered categorical variables, and the location of the continuous scores relative to the estimated threshold parameters determines a corresponding set of discrete imputes (a latent imputation below the lowest threshold is assigned to the first category, a latent impute between

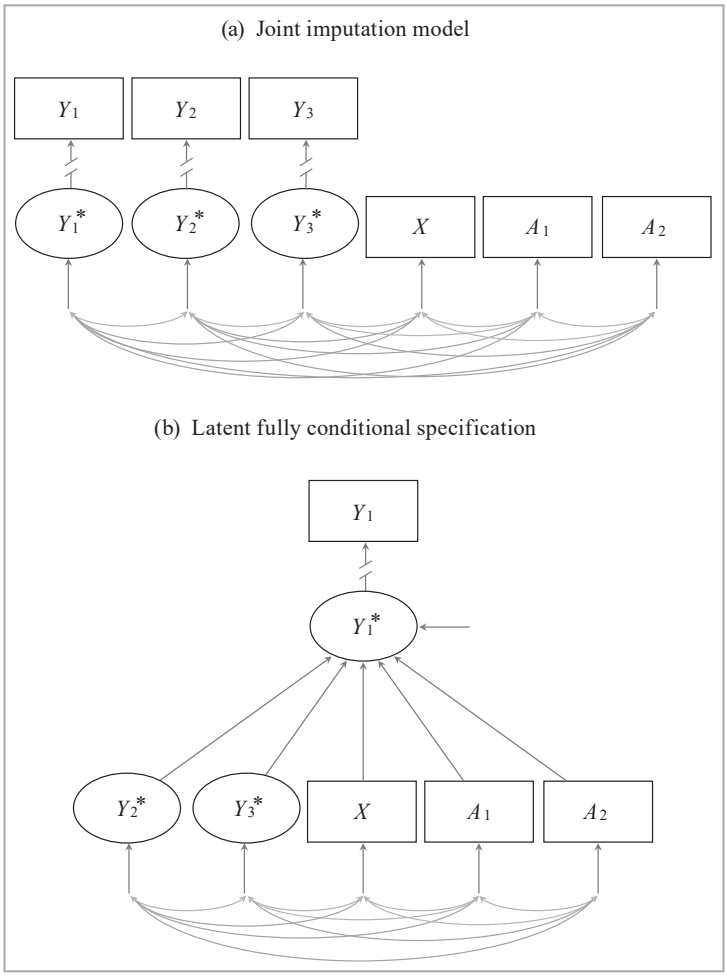

FIGURE 12.4. Joint imputation model and latent fully conditional specification imputation model.

the first two thresholds is assigned to the second category, etc.).

Analysis and Pooling Steps

The analysis step fits a structural equation model to each filled-in data set, and the final pooling step uses "Rubin's rules" (Little & Rubin, 2020; Rubin, 1987) to combine the imputation-specific estimates and standard errors into a single package of results. Fitting a structural equation model like the one in Figure 12.1 to multiply imputed data sets gives *M* estimates of each model parameter, the arithmetic average of which is the multiple imputation point estimate (Rubin, 1987, p. 76, Equation 3.1.2). Multiple imputation standard errors have a two-part composition. Averaging squared standard errors (i.e., sampling variances) from the complete data sets gives a within-imputation variance component that estimates the sampling error that would have resulted had there been no missing values, and the between-imputation variance of the *M* estimates around the average captures the additional influence of missing data on precision. Finally, complete-data sampling variation and missing data uncertainty combine

to form the total variance of an estimate, the square root of which is the standard error (Rubin, 1987, p. 76, Equations 3.1.3 through 3.1.5).

Test Statistics, Model Fit, and Non-Normal Data

Like maximum likelihood estimation, multiple imputation offers the familiar trilogy of significance testing options: the Wald test (Li, Raghunathan, & Rubin, 1991), likelihood ratio statistic (Meng & Rubin, 1992), and the score test or modification index (Mansolf, Jorgensen, & Enders, 2020). All three test statistics have a two-part composition that includes a within- and between-imputation component analogous to standard errors. Additionally, the so-called D_2 statistic is a statistical Swiss Army knife that averages any generic set of chi-square test statistics (Li, Meng, Raghunathan, & Rubin, 1991).

A natural way to compute imputation-based versions of popular absolute and relative fit indices such as the comparative fit index (CFI), Tucker–Lewis index (TLI), and root mean square error of approximation (RMSEA) is to substitute pooled chi-square statistics into familiar complete-data expressions (Enders & Mansolf, 2018). Limited computer simulation results suggest that Meng and Rubin's (1992) likelihood ratio statistic works well for this purpose when data are multivariate normal (Enders & Mansolf, 2018), and the D_2 statistic is ideally suited for pooling weighted least squares test statistics (Liu & Sriutaisuk, 2019).

Familiar corrective procedures like rescaled test statistics, robust (sandwich estimator) standard errors, and the bootstrap are also available for multiply imputed data. For robust standard errors, the within-imputation variance component is computed by averaging squared robust standard errors, and all other parts of Rubin's classic expression are the same (Yuan, Yang-Wallentin, & Bentler, 2012). Analogous substitutions apply to multivariate Wald tests that work with the variance–covariance matrix of the estimates. Likelihood ratio tests can be "robustified" by computing the Satorra–Bentler (1994) scaling factor from each data set and using the arithmetic average to rescale the pooled likelihood ratio test (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2021) or by using the D_2 procedure to pool a set of M rescaled test statistics.

Bootstrap resampling is also an option for multiply imputed data, although there is more than one way to apply the procedure. The "multiple imputation nested within bootstrapping" approach first creates B incomplete data sets by drawing bootstrap samples with replacement from the original data, and it then applies multiple imputation to create M complete data sets from each bootstrap sample (Zhang & Wang, 2013). Reversing the process gives a "bootstrapping nested within multiple imputation" procedure that first applies multiple imputation to the data, then draws B bootstrap samples with replacement from each of the M complete data sets (Wu & Jia, 2013). In either scenario, the analysis phase fits a model to each of the $B \times M$ data sets, and the resulting estimates mix to form empirical sampling distributions.

Non-normal data are also important to consider during the imputation phase. As explained previously, both the joint model and fully conditional specification estimate continuous missing values by sampling imputes from a normal distribution. Applying this procedure to a skewed variable produces a filled-in data set that is a weighted mixture of a normal curve and a non-normal distribution. Though unlikely to cause a problem with small departures from normality and low missing data rates, the imputation process itself can substantially modify non-normal distributions with high rates of missingness.

Two easy approaches to imputing non-normal data are readily available in software. The first is variant of fully conditional specification known as predictive mean matching (Kleinke, 2017; van Buuren, 2012). This approach preserves distribution shapes by sampling imputations from a donor pool of observed scores taken from participants whose predicted values match that of the person with missing data. A second strategy is to transform the variable prior to or during imputation. The Yeo–Johnson power transformation (Yeo & Johnson, 2000) is a flexible option that estimates the shape of the data as the MCMC algorithm iterates and produces imputations that closely approximate the observed-data distribution. The transformation has shown great promise (Lüdtke, Robitzsch, & West, 2020b) and it is available in latent variable modeling software (Keller & Enders, 2021).

EXAMPLE 1: NORMAL-THEORY ANALYSIS MODEL

Having established the key ideas behind maximum likelihood and multiple imputation, I use real data analysis examples to illustrate their application, beginning with the model in Figure 12.1. For the maximum

likelihood analysis, I applied the classic normal-theory estimator to the ordinal indicators, and I used robust corrections to counteract the impact of this misspecification on standard errors and model fit. Following the path diagram in Figure 12.2(a), I also assigned a normal distribution to the manifest predictors, one of which is binary.[1] Finally, I used Graham's (2003) saturated correlates specification from Figure 12.3(a) to incorporate exercise frequency as an auxiliary variable. Although exercise is a correlate of the analysis variables, it does not possess enough unique covariation to meaningfully impact the results. Nevertheless, I include it for illustration.

For the multiple imputation analysis, I used imputation models for categorical data in the first stage and subsequently applied normal-theory estimation with robust corrections in the second stage. Prior to creating imputations, I performed an exploratory MCMC analysis and used potential scale reduction factor diagnostics (Gelman & Rubin, 1992) to evaluate the algorithm's convergence. This step is critically important when imputing ordinal variables because the threshold parameters that slice the latent response distributions into discrete segments tend to converge slowly and require very long burn-in periods (Cowles, 1996). The diagnostic run indicated that a burn-in period of 25,000 iterations was sufficient. Next, I created 100 imputed data sets by saving the filled-in data from the final iteration of 100 unique MCMC chains, each with random starting values and 25,000 iterations; the literature suggests at least 20 data sets (Graham, Olchowski, & Gilreath, 2007), but I used 100 to reduce the impact of Monte Carlo simulation error on inferential quantities such as confidence intervals and probability values (Bodner, 2008). Finally, I fit the model in Figure 12.1 to each imputed data set and used Rubin's (1987) pooling rules to combine the results.

Table 12.1 gives the standardized estimates and robust standard errors from the two analyses (the table omits mean structure parameters and residual variances in the interest of space). It is widely known that given the same variables and assumptions, maximum likelihood and multiple imputation usually give identical results. The tabled results bear this out, as the two sets of results are indistinguishable. I further used the percentile bootstrap to construct 95% confidence intervals for the indirect effect (MacKinnon, 2008). The maximum likelihood bootstrap procedure drew 10,000 samples with replacement from the observed data, whereas the multiple imputation procedure sampled $B = 100$ data sets from each of the $M = 100$ imputations (i.e., bootstrapping nested within multiple imputation; Wu & Jia, 2013). The two sets of confidence interval limits were effectively equivalent, and both revealed that the mediated pathway from pain severity to depression via pain interference was significant.

Turning to model fit, the Satorra–Bentler rescaled test statistic from the maximum likelihood analysis was statistically significant, $T_{SB}(73) = 237.68$, $p < .001$; the corresponding fit indices were CFI = .96, TLI = .94, and RMSEA = .07. For the multiple imputation analysis, I used the average Satorra–Bentler scaling term to rescale Meng and Rubin's (1992) pooled likelihood ratio statistic (Jorgensen et al., 2021). This approach gave a test statistic that was well calibrated to that of maximum likelihood, $T_{SB}(73) = 241.43$, $p < .001$. The corresponding fit indices were nearly identical as well: CFI = .96, TLI = .95, and RMSEA = .07.

Modification indices (also called "score tests") have a long history in the SEM literature and require caution because they capitalize on chance (MacCallum, Roznowski, & Necowitz, 1992). I include them here as a further point of comparison. The maximum likelihood analysis produced two large indices. The first pointed to an omitted cross-loading from the pain interference factor to the depression item "I found it difficult to work up the initiative to do things" ($\chi^2(1) = 76.56$, $p < .001$), and the second pointed to a residual covariance between depression items "I felt I wasn't worth much as a person" and "I felt that life was meaningless" ($\chi^2(1) = 78.18$, $p < .001$). The multiple imputation modification indices (Mansolf et al., 2020) identified the same sources of potential misfit, and the numerical values of the test statistics were comparable.

EXAMPLE 2: MODELING LATENT RESPONSE VARIABLES

The second set of examples features three analyses that apply more appropriate distributional assumptions to the categorical indicators: FIML estimation with a probit link, fully conditional specification with multiply imputed latent response scores, and multiple imputation followed by weighted least squares estimation. For the latter, I simply applied a different estimator to the multiply imputed data from the first example; I do not consider the estimator's natural behavior, which invokes pairwise deletion and a strict MCAR mechanism.

Beyond using different link function, the FIML anal-

TABLE 12.1. Normal-Theory Standardized Estimates with Robust Standard Errors

	FIML		MI	
Parameter	Est.	*SE*	Est.	*SE*
Structural Paths				
MALE → *INTERFERE*	0.06	0.04	0.06	0.04
SEVERITY → *INTERFERE*	0.57	0.03	0.57	0.03
MALE → *DEPRESS*	0.01	0.05	0.01	0.05
SEVERITY → *DEPRESS*	–0.002	0.06	0.003	0.06
INTERFERE → *DEPRESS*	0.34	0.05	0.34	0.05
Pain Interference Loadings				
INTERFERE → $PAININT_1$	0.89	0.01	0.89	0.01
INTERFERE → $PAININT_2$	0.86	0.01	0.86	0.02
INTERFERE → $PAININT_3$	0.90	0.01	0.90	0.01
INTERFERE → $PAININT_4$	0.88	0.01	0.88	0.01
INTERFERE → $PAININT_5$	0.63	0.03	0.63	0.02
INTERFERE → $PAININT_6$	0.63	0.03	0.63	0.03
Depression Loadings				
DEPRESS → $DEPRESS_1$	0.85	0.02	0.85	0.02
DEPRESS → $DEPRESS_2$	0.78	0.02	0.78	0.02
DEPRESS → $DEPRESS_3$	0.88	0.01	0.88	0.01
DEPRESS → $DEPRESS_4$	0.85	0.02	0.85	0.02
DEPRESS → $DEPRESS_5$	0.72	0.03	0.72	0.03
DEPRESS → $DEPRESS_6$	0.72	0.03	0.72	0.03

ysis changed in two ways. First, to simplify estimation, I did not assign a distribution to the complete dummy code, and I instead linked the manifest predictors by regressing pain severity ratings on gender (see Figure 12.2(b)). Second, the estimator did not accommodate the saturated correlate's residual correlation structure, so I applied the extra dependent variable specification to the auxiliary variable (see Figure 12.3(b)). The maximum likelihood analysis has three noteworthy limitations: (1) No fit indices were available because the saturated model was inestimable, (2) no modifications were available, and (3) using the bootstrap to construct confidence intervals for the indirect effect was extraordinarily time-consuming (fitting the model to only 500 bootstrap samples required nearly 20 hours of computational time on a capable 10-core iMac Pro desktop).

The fully conditional specification approach described earlier offers the possibility of saving and analyzing the latent response scores in lieu of the categorical indicators. Doing so has the advantage of converting a complex categorical variable model into a simpler one for multivariate normal data. Moreover, the latent response variables have a natural substantive interpretation as continuous estimates of the themes captured by the questionnaire items. The procedure is conceptually equivalent to full-information estimation with a probit link but carries all the advantages of normal-theory estimation. Latent imputations have a rich history in the psychometrics literature and are routinely used in large-scale assessment settings, where they are referred to as "plausible values" (von Davier, Gonzalez, & Mislevy, 2009).

Because the latent item responses have 100% missing data, more data sets are needed to maximize precision and minimize Monte Carlo simulation error. In my experience, increasing the number of imputations

from 100 to 500 can have a meaningful impact on test statistics and probability values, but additional data sets provide diminishing returns. To this end, I used latent fully conditional specification (Keller & Enders, 2021) to create 500 filled-in data sets. Following my earlier procedure, I specified 500 separate MCMC chains with random starting values and 25,000 iterations each, and I saved the filled-in data and latent response scores from the final iteration of each chain. There is no need to apply robust corrections to this analysis because the latent response variables are normal by construction.

Table 12.2 gives the standardized estimates and robust standard errors from the analysis. Analyzing plausible values produced estimates that were comparable to those of the full-information estimator with a probit link function, and both sets of results were noticeably different from weighted least squares; structural paths differed by up to one standard error unit, and several loadings differed by more than two standard errors. Even with this relatively simple model, maximum likelihood estimation was incapable of generating a model fit statistic, and modification indices were unavailable. For the plausible values analysis, Meng and Rubin's (1992) pooled likelihood ratio test was significant, $\chi^2(73) = 198.00$, $p < .001$; the corresponding fit indices were TLI = .93, CFI = .95, and RMSEA = .06. Appling the D_2 pooling procedure to the weighted least squares test statistics (Liu & Sriutaisuk, 2019) also gave a significant chi-square test, $\chi^2(73) = 142.21$, $p < .001$; the corresponding fit statistics were TLI = .98, CFI = .98, and RMSEA = .04. Finally, both procedures produced large modification indices that identified same cross-loading and residual correlation described in the first example.

TABLE 12.2. Standardized Estimates and Standard Errors from the Categorical Variable Models

	FIML		Latent MI		WLS	
Parameter	Est.	*SE*	Est.	*SE*	Est.	*SE*
Structural Paths						
MALE → *INTERFERE*	0.07	0.04	0.07	0.04	0.06	0.04
SEVERITY → *INTERFERE*	0.59	0.03	0.59	0.03	0.60	0.03
MALE → *DEPRESS*	–0.002	0.05	–0.001	0.05	0.01	0.05
SEVERITY → *DEPRESS*	–0.03	0.06	–0.03	0.06	–0.04	0.06
INTERFERE → *DEPRESS*	0.38	0.06	0.36	0.06	0.41	0.06
Pain Interference Loadings						
INTERFERE → $PAININT_1$	0.92	0.01	0.92	0.01	0.92	0.01
INTERFERE → $PAININT_2$	0.89	0.01	0.89	0.01	0.89	0.01
INTERFERE → $PAININT_3$	0.92	0.01	0.92	0.01	0.92	0.01
INTERFERE → $PAININT_4$	0.91	0.01	0.91	0.01	0.91	0.01
INTERFERE → $PAININT_5$	0.75	0.03	0.75	0.03	0.75	0.03
INTERFERE → $PAININT_6$	0.66	0.03	0.66	0.03	0.65	0.03
Depression Loadings						
DEPRESS → $DEPRESS_1$	0.90	0.02	0.90	0.02	0.89	0.02
DEPRESS → $DEPRESS_2$	0.84	0.02	0.84	0.02	0.88	0.02
DEPRESS → $DEPRESS_3$	0.93	0.01	0.93	0.01	0.92	0.01
DEPRESS → $DEPRESS_4$	0.90	0.02	0.91	0.02	0.90	0.02
DEPRESS → $DEPRESS_5$	0.82	0.02	0.82	0.03	0.83	0.02
DEPRESS → $DEPRESS_6$	0.81	0.03	0.81	0.03	0.84	0.02

INTERACTIVE AND NONLINEAR EFFECTS

The emergence of Bayesian and multiple imputation methods for interactive and nonlinear effects is a relatively recent development in the literature (Enders, Du, & Keller, 2020; Lüdtke et al., 2020b), and one that readily extends to latent variable models (Asparouhov & Muthén, 2021a; Keller & Enders, 2021). To illustrate, suppose it is of interest to assess whether gender moderates the structural paths connecting pain severity to pain interference and pain interference to depression. The presence of an interactive term implies that the constituent variables' missing data distributions feature nonlinearities that are mathematically inconsistent or incompatible with the normal distribution (Bartlett, Seaman, White, & Carpenter, 2015). This incompatibility is the source of bias associated with so-called "just-another-variable procedures" that treat product terms as normal variables (Cham, Reshetnyak, Rosenfeld, & Breitbart, 2017).

Incomplete interactive effects require an SEM framework that works with a collection of variable-specific conditional log-likelihood functions rather than a single multivariate function like Equation 12.3 (Keller & Enders, 2021; Merkle & Rosseel, 2018). This conditional specification treats each factor loading or structural path as a unique regression equation with its own univariate log-likelihood, and the log-likelihood for the entire model is the sum of the univariate parts. In this framework, factor scores are missing data, just like latent response variables and incomplete numeric variables. The distributions of missing values from which MCMC draws imputations are complex, multipart functions that depend on every univariate model in which a variable appears. Importantly, this specification does assign a distribution to a product term. Rather, MCMC tailors imputations—manifest or latent—to the fitted model and creates filled-in values that anticipate the multiplication of the lower-order variables. Enders (2022, Ch. 10) describes these models in more detail.

To illustrate missing data handling for an analysis with moderated structural paths, consider the path diagram in Figure 12.5. The dashed arrows depict a manifest-by-manifest interaction involving the gender dummy code and the pain severity index and a latent-by-manifest interaction between gender and the pain interference factor. The structural regression equations are as follows:

$$\begin{aligned} INTERFERE_i^* &= \beta_{01} + \beta_{11}\left(SEVERITY_i\right) + \beta_{21}\left(MALE_i\right) \\ &\quad + \beta_{31}\left(SEVERITY_i\right)\left(MALE_i\right) + \varepsilon_{1i} \\ DEPRESS_i^* &= \beta_{02} + \beta_{12}\left(SEVERITY_i\right) + \beta_{22}\left(MALE_i\right) \\ &\quad + \beta_{32}\left(INTERFERE_i^*\right) \\ &\quad + \beta_{42}\left(INTERFERE_i^*\right)\left(MALE_i\right) + \varepsilon_{2i} \end{aligned} \tag{12.9}$$

Following earlier conventions, I use an oval and rectangle to differentiate a latent response variable and its discrete indicator, respectively, and the broken arrow connecting the two is the link function that maps the unobserved continuum to the discrete responses (e.g., a broken arrow reflects the idea that threshold parameters categorize the latent distribution into discrete segments). Although the gender dummy code is complete and does not require a distribution, the diagram shows that categorical (including multicategorical nominal) predictors can be modeled with latent response variables. I used the Blimp application (Keller & Enders, 2021) and Bayesian estimation to fit the model, and the same run can produce multiple imputations for a frequentist analysis. Neither interaction had support from the data.

MODELS FOR MNAR PROCESSES

An MNAR process is one where the unseen score values carry information about missingness beyond that in the observed data. This definition suggests that nonresponse is entangled with the outcome variable in a way that cannot be ignored when analyzing data. The two major modeling frameworks for MNAR processes—selection models and pattern mixture models—mitigate nonresponse bias by introducing an additional model that describes the occurrence of missing data, albeit in different ways. I briefly describe two variants of the selection model for longitudinal growth models (the context in which a great deal of methodological work is situated). Numerous resources provide more detailed treatments (e.g., Enders, 2011; Muthén, Asparouhov, Hunter, & Leuchter, 2011).

The Diggle–Kenward selection model (Diggle & Kenward, 1994) for outcome-dependent missing data links dropout at occasion *t* to scores from the concurrent and prior measurement occasions. The model augments the focal growth curve model with additional regression equations that predict binary missing data indica-

tors. To illustrate, Figure 12.6 shows a path diagram of a three-wave latent curve model that incorporates an explicit missingness model. The rectangles labeled M_2 and M_{3d} are binary dropout indicators,[2] and the dashed arrows pointing to these variables are logistic or probit regressions that predict missingness.

The shared parameter selection model for trajectory-dependent missingness instead uses individual random intercepts and slopes as predictors of dropout (Albert & Follmann, 2009). Figure 12.6(b) shows the path diagram for the three-wave growth curve model. The dashed lines are again logistic or probit regressions. This model is useful for a situation where one's underlying growth trajectory could be responsible for missing data rather than time-specific realizations of the dependent variable (e.g., participants experiencing

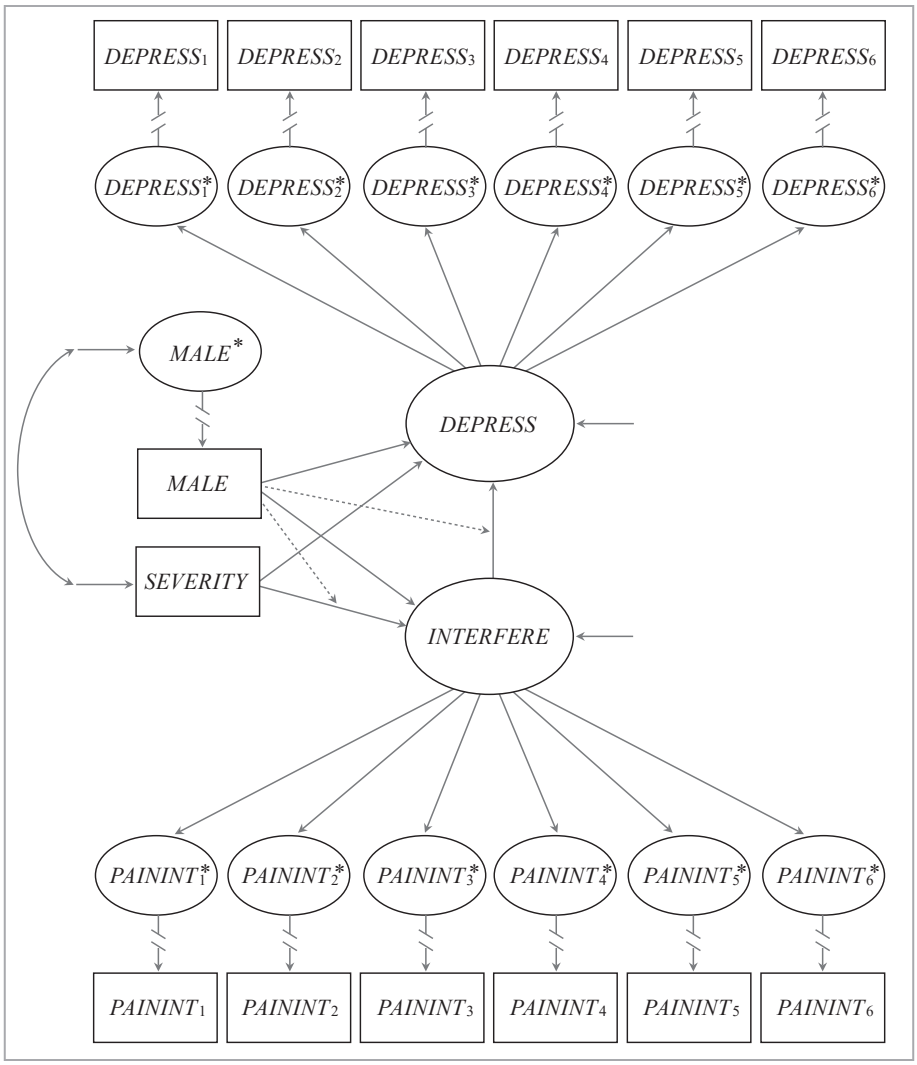

FIGURE 12.5. Structural model with a manifest-by-manifest variable interaction and a manifest-by-latent variable interaction (the dashed arrows).

the most rapid declines in depressive symptoms might quit the study because they judge that treatment is no longer necessary, whereas individuals with elevated and flatter trajectories might drop out to seek treatment elsewhere).

Selection models for MNAR processes require strict, unverifiable assumptions (e.g., multivariate normality, a correctly specified missingness model). Because simple model misspecifications such as omitting a key determinant of missingness can produce biased estimates, a common view is that these models are best suited for sensitivity analyses. To apply this strategy to longitudinal data, one could fit growth curve models that leverage three different assumptions about the missing data: a conditionally MAR process, an outcome-dependent MNAR mechanism, and a trajectory-dependent missingness process. An online supplemental document can present side-by-side comparisons of two or more sets of analysis results, with any discrepant findings noted in main body of text. Enders (2022, Ch. 9) provides a detailed illustration of a sensitivity analysis.

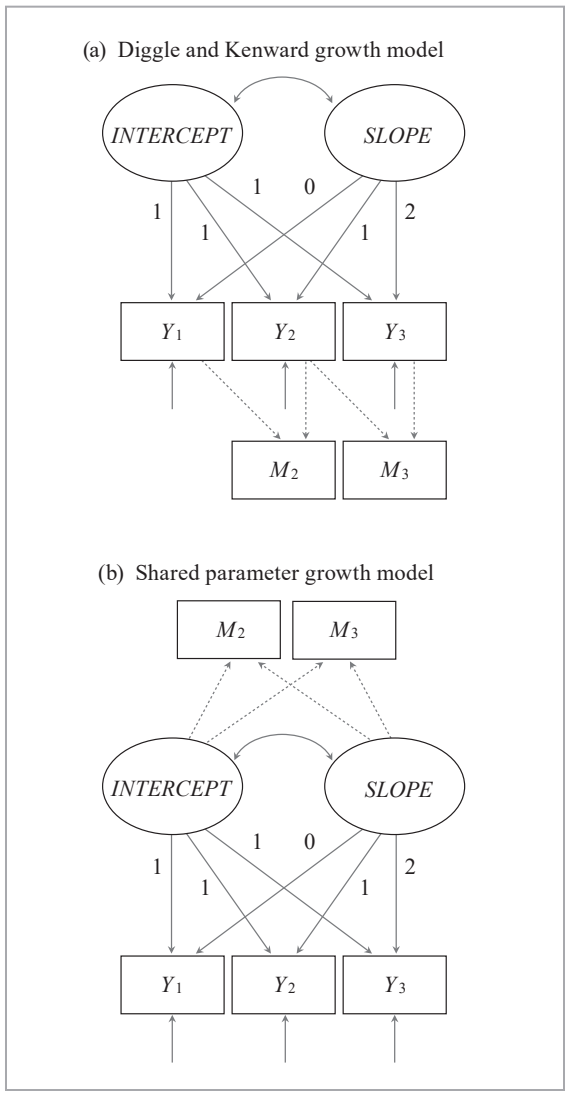

FIGURE 12.6. Diggle and Kenward selection model and shared parameter growth models.

CONCLUSION

The goal for this chapter was to provide readers with an overview of the two predominant strategies for handling missing data in structural equation models—maximum likelihood estimation and multiple imputation. Both approaches have matured since the first edition of this *Handbook*, and the types of problems that researchers can tackle with these methods are broader than ever. The analysis examples highlighted that when given the same data and assumptions, the two methods usually produce indistinguishable results. Thus, in most situations there is no reason beyond personal preference to prefer one method to the other. Perhaps the major exceptions to this conclusion occur when fitting models with mixtures of categorical and continuous variables or nonlinear effects, in which case Bayesian estimation and multiple imputation are often more flexible.

NOTES

1. Alternatively, I could have used a specification like Figure 12.2(b) to apply a logistic model to the binary variable. The choice of specification did not meaningfully influence the results.

2. M_1 is not needed because baseline scores are complete.

REFERENCES

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). Boca Raton, FL: Chapman & Hall.

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. 243–277). Mahwah, NJ: Erlbaum.

Asparouhov, T., & Muthén, B. (2021a). Bayesian estimation of single and multilevel models with latent variable interactions. *Structural Equation Modeling: A Multidisciplinary Journal*, 28, 314–328.

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*, 622–637.

Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus*. www.statmodel.com/download/imputations7.pdf

Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research, 24*, 462–487.

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal, 15*, 651–675.

Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal, 1*, 287–316.

Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. West Sussex, UK: 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*, 12–30.

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*, 330–351.

Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing, 6*, 101–111.

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 B: Statistical Methodology, 39*, 1–38.

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society C: Applied Statistics, 43*, 49–93.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*, 1–16.

Enders, C. K. (2022). *Applied missing data* (2nd ed.). New York: Guilford Press.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*, 430–457.

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

Enders, C. K., & Mansolf, M. (2018). Assessing the fit of structural equation models with multiply imputed data. *Psychological Methods, 23*, 76–93.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 10*, 80–100.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed?: Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206–213.

Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research, 50*, 285–299.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *Package 'semTools.'* https://cran.r-project.org/web/packages/semtools/semtools.pdf

Keller, B. T., & Enders, C. K. (2021). *Blimp user's guide* (Version 3). www.appliedmissingdata.com/multilevel-imputation.html

Kenward, M. G., & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science, 13*, 236–247.

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*, 371–404.

Li, K.-H., Meng, X.-L., Raghunathan, T. E., & Rubin, D. B. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica, 1*, 65–92.

Li, K. H., Raghunathan, T. E., & Rubin, D. B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association, 86*, 1065–1073.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.

Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd ed.). Hoboken, NJ: Wiley.

Liu, Y., & Sriutaisuk, S. (2019). Evaluation of model fit in structural equation models with ordinal missing data: An

examination of the D_2 method. *Structural Equation Modeling: A Multidisciplinary Journal, 27*, 561–583.

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*, 361–381.

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*, 157–181.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490–504.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.

Mansolf, M., Jorgensen, T. D., & Enders, C. K. (2020). A multiple imputation score test for model modification in structural equation models. *Psychological Methods, 25*, 393–411.

Meng, X. L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika, 79*, 103–111.

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software, 85*, 1–30.

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*, 17–33.

Muthén, B., Muthén, L., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Authors.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research, 151*, 53–79.

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*, 490–500.

Raykov, T. (2011). On testability of missing data mechanisms in incomplete data sets. *Structural Equation Modeling: A Multidisciplinary Journal, 18*, 419–429.

Raykov, T., & West, B. T. (2015). 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*, 45–53.

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*, 354–373.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. v. Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods, 15*, 352–367.

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*, 477–497.

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*, 163–181.

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*, 741–763.

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

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

van Buuren, S. (2012). *Flexible imputation of missing data*. New York: Chapman & 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*, 1049–1064.

van Buuren, S., & Groothuis–Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*, 1–67.

von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 2*, 9–36.

Widaman, K. F. (2006). Missing data: What to do with or without them. *Monographs of the Society for Research in Child Development, 71*, 42–64.

Wu, W., & Jia, F. (2013). A new procedure to test mediation with missing data through nonparametric bootstrapping and multiple imputation. *Multivariate Behavioral Research, 48*, 663–691.

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87*, 954–959.

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*, 165–200.

Yuan, K.-H., Yang-Wallentin, F., & Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. *Sociological Methods and Research, 41*, 598–629.

Zhang, Z., & Wang, L. (2013). Methods for mediation analysis with missing data. *Psychometrika, 78*, 154–184.