






A tutorial for handling suspected missing not at random data in longitudinal clinical trials

James L. Peugh^{a,b}  , Michael D. Toland^c  & Heather Harrison^{a,b} 

^aDivision of Behavioral Medicine Clinical Psychology, Cincinnati Children's Hospital Medical Center Cincinnati, Ohio

^bDepartment of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio

^cThe Herb Innovation Center, Judith Herb College of Education, University of Toledo, Toledo, Ohio

Abstract ■ Missing data in longitudinal randomized clinical trials, even if assumed to be missing at random (MAR), can result in biased parameter estimates and incorrect treatment conclusions. If missing data are suspected to be missing not at random (MNAR, i.e., missing data due to the unobserved values themselves), accepted missing data handling techniques are inadequate and the problem is compounded due to the introduction of additional bias. The goal of this paper is to provide trialists the analytic tools and methodological steps needed to assess treatment effect veracity if MNAR is suspected, enabling researchers to reach reasonable and defensible treatment effect conclusions. The explanations, steps, and conclusions are demonstrated using trial data involving binge drinking behavior among college students. All analysis model diagrams are presented, and both linear model equations and Mplus syntax scripts for all analyses are included in a supplemental Appendix to further provide trialists the methodological tools needed to further test treatment effect estimates when MNAR is suspected.

Keywords ■ MNAR, missing data, pattern mixture, selection. **Tools** ■ Mplus.

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers

■ Mathieu Caron-Diotte (Université de Montréal)

 James.Peugh@cchmc.org

 [10.20982/tqmp.19.4.p347](https://doi.org/10.20982/tqmp.19.4.p347)

“Missing values represent a potential source of bias in a clinical trial... In reality, however, there will almost always be some missing data... A trial may be regarded as valid, none the less, provided the methods of dealing with missing values are sensible...” (ICH E9 Working Group, 1999, section 5.3, cited in Carpenter & Kenward, 2007, italics added for emphasis).

Introduction

No matter how tightly controlled the data collection portion of a longitudinal clinical trial may be, different amounts and patterns of missing data are commonly viewed as inevitable (Gomer & Yuan, 2021). In addition to valuable treatment efficacy and effectiveness information being lost, missing longitudinal trial data is additionally problematic due to the possibilities of missing data: 1) biasing parameter and standard error estimates, 2) biasing treatment effect size estimates, and 3) leading trialists to incorrect conclusions regarding treatment efficacy and effectiveness (Cro et al., 2020; Gerber & Green, 2012; Gomer & Yuan, 2021; Zhou & Fishbach, 2016). Missing

data generated by a random process that is independent of any (a) past, present, or future (b) observed or unobserved (c) predictor, covariate, or response variable value defines a missing completely at random (MCAR) missing data mechanism (e.g., Demirtas & Schafer, 2003; Enders, 2010, 2011b; Post et al., 2010; Rubin, 1976). MCAR mechanisms are often viewed as unlikely and unrealistic explanations for longitudinal trial missing data (Mallinckrodt et al., 2014; Raykov, 2011). A missing at random (MAR) explanatory mechanism for missing data assumes that missing response variable data are independent of unobserved response variable values, but are related to response variable data collected at a previous or later time point, or to one or more observed predictor or covariate variables included in the data analysis model (Demirtas & Schafer, 2003; Enders, 2011b; Hedeker & Gibbons, 1997; Post et al., 2010; Raykov, 2011). A MAR mechanism further assumes that the response variable distributions for participants who do and do not drop out of a randomized trial are identical (Carpenter & Kenward, 2007). A MAR mechanism in a longitudinal trial design may be a logical assumption, but unfortu-



nately is one that cannot be confirmed (Mallinckrodt et al., 2014). Missing longitudinal trial data assumed to be either MCAR or MAR is often termed ignorable missing data because it can be handled with minimal parameter estimate bias by maximum likelihood (ML) estimation, multiple imputation (MI), inverse probability weighting (IPW), or double sampling and bounds techniques (e.g., Coppock et al., 2017; Enders, 2010, 2022; Gomila & Clark, 2022; Graham, 2012; Heitjan & Rubin, 1991). However, each of these techniques also rely on untestable assumptions regarding the missing data distribution (Cro et al., 2020; Enders, 2011b, 2011a, 2022).

Missing not at random (MNAR) has been defined as non-ignorable dropout (Demirtas & Schafer, 2003) that is conditional on the unobserved response variable values (Carpenter & Kenward, 2007; Mallinckrodt et al., 2014; Rubin, 1976). In other words, MNAR missing response variable data is assumed to be a direct result of the response variable values that should have been observed but were not. Examples of suspected MNAR in the published literature include: 1) missing data in an investigation of lead exposure in children due to the child being hospitalized for lead toxicity (Roberts & English, 2016), 2) missing data in a study of routine child clinical care due to participants leaving the study to seek care elsewhere due to care quality dissatisfaction (Gachau et al., 2020), 3) missing depression data due to depression symptom severity resulting in study dropout (Shrive et al., 2006), 4) missing data in a children's alcohol abuse study due to leaving the study for fear of alcohol abuse detection (Sharmin et al., 2019), 5) missing weight data in a study of pediatric bariatric weight loss surgical patients due to dropping out before their weight regain is detected (Dewberry et al., 2020), and 6) missing data in a pediatric cancer treatment trial due to mortality (e.g., Bernhard et al., 1998). In contrast to missing data assumed to be MCAR or MAR, suspected MNAR data cannot be handled appropriately by any of the previously mentioned techniques.

The purpose of this paper is to provide a step-by-step tutorial for implementing a series of sensitivity analyses to address missing longitudinal trial data that is suspected to be MNAR (e.g., Beunckens et al., 2008; Demirtas & Schafer, 2003; Enders, 2010, 2011b, 2011a; Gottfredson et al., 2014; Graham, 2012; Hedeker & Gibbons, 1997; Jung et al., 2011; Little, 1995; Molenberghs et al., 2008; Muthén et al., 2011; Muthén & Brown, 2009; Roy, 2003, 2007; Sterba & Gottfredson, 2015; Yuan & Little, 2009). As an aside, suspected MNAR data is not limited to either longitudinal designs or to randomized trials. For example, MNAR could be suspected in an observational newlywed marital satisfaction study due to divorce (DiLillo et al., 2009). This paper proceeds from the assumption that although all missing

data handling techniques rely on untestable assumptions to varying degrees, reasonable answers to longitudinal trial research questions can be obtained even under suspected MNAR (Carpenter & Kenward, 2007). Specifically, a series of sensitivity analysis models, each making different missing data assumptions, will be specified and analyzed using a longitudinal trial dataset that tested treatment interventions designed to reduce college student binge drinking. All linear model equations and Mplus data analysis syntax scripts for all example analyses are available in a supplemental Appendix.

Description of the Murphy et al. (2019) Longitudinal Trial and Dataset

Murphy et al. (2019) examined the effects of a brief motivational interview (brief MI) intervention on problematic binge drinking behavior in a sample of 393 first- and second-year undergraduate college students. A brief MI was used to both illuminate a participant's drinking behaviors and clarify the risks posed to college students' career aspirations. The brief MI intervention was supplemented with a behavioral economic aspect that was geared toward counteracting "delay discounting", defined as the tendency to see the immediate personal and social benefits of drinking behavior as more valuable than the delayed rewards of salary and status associated with career attainment years later. Participants were randomly assigned to the following independent variable conditions: 1) an assessment control condition, 2) brief MI plus relaxation training (RT; i.e., diaphragm breathing, muscle relaxation) and stress management, or 3) brief MI plus substance-free activity sessions (SFAS) designed to replace drinking behaviors with those that are more conducive to long-term goal attainment. Participants were assessed at baseline, and at 1-, 6-, 12-, and 16-month follow ups. Data from baseline, 1-, 6-, and 12-month follow ups are considered here. Although the original Murphy et al. study investigated the effects of RT and SFAS on Daily Drinking Questionnaire (DDQ) and Young Adult Alcohol Consequences Questionnaire (YAACQ) scores, the response variable of interest here is a secondary continuous outcome measure of binge drinking (BINGE) behavior. The specific treatment effect secondary research question of interest is whether RT and SFAS significantly reduce binge drinking behavior versus an assessment control. MNAR could reasonably be suspected in the BINGE drinking trial because dropout missing data may be due to the consequences of binge drinking (Gomer & Yuan, 2021), such as physical symptoms (e.g., hangovers) or hospitalization, academic suspension or expulsion, or arrest and incarceration.



Missing Data Theory

A fundamental tenet of missing data theory states that if missing response variable data are related to one or more observed analysis variables, there is no relationship between observed response variable data and missing response variable data. This is the definition of conditional independence, which is implicitly invoked when missing data is assumed to be MAR. If the MAR assumption of conditional independence is true, only a data analytic model for the observed variables is needed to handle missing data.

However, if MNAR is suspected, two methodological difficulties immediately arise: 1) the joint distribution of both the observed response variable and missing response variable data quantifying the relationship between the two cannot be known with certainty (Enders, 2011b, 2011a), and 2) MNAR models that propose different possible joint distributional relationships rely heavily on assumptions that cannot be tested. Conditional independence is often invoked as an assumption in MNAR scenarios to suggest that the observed data and the missing data are unrelated assuming certain specific, but untestable, MNAR model specifications and assumptions have been met. A theme that runs throughout this paper, where applicable, is how conditional independence, if assumed, is specified across a variety of different MNAR analysis models. Overall, if MNAR is suspected, then addressing MNAR while simultaneously answering the research question at hand requires specifying the relationship between the observed data analysis model and missing data model (Rubin, 1976). The missing data indicator variables needed to add a missing data model to the longitudinal trial data analysis model are described next.

Coding Missing Data Indicator Variables

Handling suspected MNAR data begins with identifying the number of separate missing data patterns contained in the sample data and creating different types of missing data indicator variables based on the observed patterns (Mplus will output missing data patterns if the command 'OUTPUT: patterns;' is added to the input syntax; see supplemental Appendix A, p. 1). The eight missing data patterns observed in the BINGE dataset are shown in the first column (Pattern) of Table 1 followed by BINGE data examples for each pattern in the next four columns (BINGE_B-BINGE_12). Missing data in longitudinal trials are classified as one of two types: dropout or intermittent. Dropout can occur only after an observed baseline assessment and is defined as any participant who fails to complete the study. Patterns 2 ($n = 26$, 6.6%), 4 ($n = 13$, 3.3%), 6 ($n = 2$, 0.5%), and 8 ($n = 9$, 2.3%) are examples of dropout missing data. Intermittent missing data is defined as any participant who

completes the study but has missing data for one or more intervening assessment time points. Patterns 3 ($n = 21$, 5.3%), 5 ($n = 10$, 2.5%), and 7 ($n = 6$, 1.5%) are examples of intermittent missing data. Pattern 1 shows complete data representing participants with no missing data for any assessment time point ($n = 306$, 77.9%).

Table 1 also shows the first two types of binary missing data indicators, dummy dropout indicators (d1 – d3; columns 7-9) and survival dropout indicators (s1 – s3; columns 10-12). Dummy indicators assign a value of 1 for post-baseline dropout, 0 otherwise; survival indicators assign 0 for observed data, 1 for post-baseline dropout, and -99 (missing data indicator value) for all post-dropout assessments. Notice both dummy and survival indicators treat intermittent missing data (e.g., not dropout missing data) essentially the same as observed data by invoking a MAR assumption (Enders, 2011b, 2011a; Muthén et al., 2011) and assigning values of 0 for patterns 3, 5, and 7. Notice also that both dummy and survival indicators assign a value of 1 for the time point at which dropout occurred. The two sets of indicators differ only in their respective post-dropout codes as shown in patterns 4 and 8: dummy indicators assign post-dropout codes a value of 0 (as will be shown in the pattern mixture model example, dummy dropout indicators will serve as predictors of missing data, so complete data for these indicators is needed to avoid listwise deletion); survival indicators assign post-dropout codes a missing value indicator (-99). Consistent with discrete-time survival analysis, survival indicators treat dropout as an “at risk” event and assign a post-dropout missing value indicator (-99). This effectively removes participants who have dropped out from the “risk set” that contains participants who have not dropped out, but still could later.

In contrast to dummy and survival dropout indicators, multinomial missing data indicators (m1 – m3 in Table 1) assign observed data a value of 0 (reference class), intermittent missing data a value of 1 (assuming possible MAR), and dropout missing data a value of 2 (assuming possible MNAR; Albert et al., 2002; see also Lin et al., 2004). A comparison of predictions of intermittent (suggesting possible MAR) and dropout missing data (suggesting possible MNAR) can be performed using both survival (s1-s3) and multinomial (m1-m3) missing data codes in the missing data model. Finally, two additional single value missing data indicators are shown in the last two columns of Table 1: ‘Droptime’ (Roy, 2003, 2007; Yuan & Little, 2009) and ‘Summary’ (Gottfredson et al., 2014; Rose et al., 2010). Both have the same values for complete data (4, see pattern 1) and dropout patterns 2, 4, and 8. ‘Droptime’ codes the last assessment timepoint at which response data was observed; ‘Summary’ is a count of the number of observed



Table 1 ■ Missing Data Codings for MNAR Models

| Pattern | BINGE_B | BINGE_1 | BINGE_6 | BINGE_12 | n | d1 | d2 | d3 | s1 | s2 | s3 | m1 | m2 | m3 | Droptime | Summary |
|---------|---------|---------|---------|----------|-----|----|----|----|----|-----|-----|----|----|----|----------|---------|
| 1 | 5 | 20 | 2 | 1 | 306 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 |
| 2 | 4 | 0 | 1 | -99 | 26 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 2 | 3 | 3 |
| 3 | 4 | 3 | -99 | 2 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 3 |
| 4 | 8 | 8 | -99 | -99 | 13 | 0 | 1 | 0 | 0 | 1 | -99 | 0 | 2 | 2 | 2 | 2 |
| 5 | 1 | -99 | 0 | 2 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 4 | 3 |
| 6 | 10 | -99 | 13 | -99 | 2 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 3 | 2 |
| 7 | 6 | -99 | -99 | 3 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 4 | 2 |
| 8 | 16 | -99 | -99 | -99 | 9 | 1 | 0 | 0 | 1 | -99 | -99 | 2 | 2 | 2 | 1 | 1 |

Note. BINGE refers to response variable values collected at baseline (_B), 1-month (_1), 6-month (_6) and 12-month (_12) follow-ups, d1 - d3 are dummy missing data codes; s1 - s3 are survival missing data codes, m1 - m3 are multinomial missing data codes, -99 is a missing data indicator value. For d1-d3 and s1-s3, 0 = observed data, 1 = missing data. For m1 - m3, 0 = observed data, 1 = intermittent missing data, 2 = dropout missing data. Droptime indicates the assessment time point at which dropout occurred. Summary is a count of the number of observed response variable scores.

data points. First is a description of MNAR model types and diagnostic statistics, then each of these missing data indicator variables will be used in the MNAR sensitivity analyses that follow.

MNAR Models and Diagnostics

The two general types of MNAR models can be differentiated based on how the added missing data model is specified (e.g., Little, 1995). Pattern mixture models add dummy missing data codes (Table 1, columns 7-9) to the analysis model as additional predictors assuming that observed response variable data is conditional on missing data. Selection (and shared parameter) models add survival and multinomial missing data codes (Table 1, columns 10-15) as additional response variables assuming that missing response variable data is conditional on observed data. Both pattern mixture and selection (and shared parameter) models offer very different plausible assumptions for why response variable data are missing, but those assumptions cannot be tested or proven definitively. Research (e.g., Gomer & Yuan, 2021) shows that, among other possibilities (discussed below), MNAR can also be suspected in a longitudinal design due to one or more trajectory random effects (shared parameter models; Wu & Carroll, 1988; e.g., Yang & Maxwell, 2014; Yang, et al., 2015, as cited in Gomer & Yuan, 2021). Readers interested in the mathematical formulations and definitions of how pattern mixture models and selection models factorize the joint distributional relationship between observed data and missing data can consult several available sources (e.g., see Enders, 2010, pp. 290-291; 2022, pp. 348-351; Gomer & Yuan, 2021, p. 562; Little & Rubin, 2002; pp. 351-355; Muthén et al., 2011, p. 18).

Gomer and Yuan (2021; see also Enders, 2022, pp. 349-353) further distinguish between suspected focal MNAR and diffuse MNAR scenarios. Focal MNAR is defined as

missing response variable data being wholly conditional on the unobserved values themselves and not on any other observed variable. Previously mentioned published studies investigating severe depression, substance abuse treatment, and marital satisfaction are examples of suspected focal MNAR where missing response variable values are completely due to the unobserved values themselves. In contrast, diffuse MNAR is defined as missing response variable values being dependent upon unobserved values even after conditioning on observed response variable or predictor variable values. For example, returning to the BINGE drinking study, if fraternity or sorority membership was known to increase the likelihood of BINGE drinking, missing BINGE drinking data would still be dependent on the unobserved values even after controlling for college Greek society participation. Pattern mixture, shared parameter, and selection models will be specified to address both focal and diffuse MNAR where applicable.

Further, as noted previously, there are two ways in which suspected MNAR can arise: missing response variable data can be dependent upon the missing values themselves or can depend on one or more random trajectory effects from a longitudinal growth model. A third way that a MNAR process could arise is by excluding one or more relevant auxiliary variables (Graham, 2003) from the analysis model needed to answer the research question. Briefly, effective auxiliary variables are defined as: 1) not needed to answer the research question but are both related to missing response variable values and available for analysis model inclusion, 2) showing a response variable mean difference effect size of $d > 0.20$ when comparing dummy-coded observed data and missing data groups, and 3) are correlated with data analysis model response variable residual variances at $r > 0.30$ (see e.g., Enders, 2022, pp. 21-23). BINGE dataset variables such as gender, age, Greek fraternity or sorority membership, parent income,



Table 2 ■ Estimates for Traditional MNAR Models

| | MAR | Pattern Mixture | | Shared Parameter | | | |
|----------------------|----------------|-----------------|----------------|------------------|----------------|----------------|----------------|
| | | Focal | Diffuse | MAR | Diffuse | DiffuseIV | Multinomial |
| RT → Intercept | -1.77 (0.47)** | -1.78 (0.47)** | -1.82 (0.49)** | -1.77 (0.47)** | -1.74 (0.47)** | -1.76 (0.47)** | -1.80 (0.46)** |
| SFAS → Slope | -1.24 (0.51)* | -1.24 (0.51)* | -1.31 (0.52)* | -1.24 (0.51)* | -1.25 (0.51)* | -1.25 (0.50)* | -1.25 (0.42)** |
| Estimated Parameters | 14 | 16 | 20 | 19 | 19 | 21 | 23 |
| LogL | -3820.60 | -3820.27 | -3823.47 | -4018.52 | -4019.36 | -4018.25 | -4220.59 |
| AIC | 7669.19 | 7672.55 | 7686.94 | 8075.05 | 8076.72 | 8078.49 | 8487.18 |
| BIC | 7724.82 | 7736.13 | 7766.41 | 8150.55 | 8152.22 | 8161.94 | 8578.58 |
| ΔAIC | | -3.35 | -17.75 | | -1.68 | -3.45 | NA |
| ΔBIC | | -11.30 | -41.59 | | -1.68 | -11.40 | NA |

Note. **: $p < .01$; * : $p < .05$; NA = not applicable

participant income, family history of problematic drinking, grade point average, marijuana use, depression, anxiety, alcohol consequences, self-regulation, protective behaviors, and delay discounting were tested, but none of the additional variables collected in the BINGE drinking trial met criteria for auxiliary inclusion, so the possibility of a MNAR by omission (Collins et al., 2001) scenario will be discussed later.

Importantly, it is well-known in the missing data literature that there is no definitive test of whether MAR or MNAR is more plausible in a sample of data (e.g., Little & Rubin, 2002). However, it has been shown how traditional MNAR models can be re-specified as models that as-

sume MAR (e.g., Molenberghs et al., 2008), and comparisons can be made between a MNAR model and a properly respecified and comparable MAR model (Molenberghs et al., 1997, 2008; Molenberghs & Kenward, 2007; Verbeke et al., 2001; Verbeke & Molenberghs, 2000; Verbeke et al., 2008). Although traditional model comparison techniques such as likelihood ratio nested model tests cannot be used (e.g., Molenberghs & Kenward, 2007), Kuha (2004) and others have recommended relative model fit comparisons using AIC and BIC values. Sterba and Gottfredson (2015, p. 297) have advocated for the use of AIC and BIC difference statistics (ΔAIC & ΔBIC) between MNAR models and their respective appropriate MAR counterparts:

$$\Delta AIC = -2(\text{Log}L_{MAR} - \text{Log}L_{MNAR}) + 2(P_{MAR} - P_{MNAR}) \tag{1}$$

$$\Delta BIC = -2(\text{Log}L_{MAR} - \text{Log}L_{MNAR}) + (\ln[N]) * (P_{MAR} - P_{MNAR}), \tag{2}$$

where $\text{Log}L$ represents the model log-likelihood value and P represents the number of parameters estimated. Negative ΔAIC and ΔBIC values suggest support for the MAR model, positive values suggest support for the MNAR model. However, ΔAIC and ΔBIC values can be biased

based on participants that show aberrant or extreme response variable values (e.g., Kenward, 1998). Sterba and Pek (2012) provide computations for individual ΔAIC and ΔBIC influence values as follows:

$$\Delta AIC_i^{INF} = -2(\ln[\text{Log}L_i^{MAR} - \text{Log}L_i^{MNAR}]) \tag{3}$$

$$\Delta BIC_i^{INF} = -2(\ln[\text{Log}L_i^{MAR} - \text{Log}L_i^{MNAR}]) + (P_{MAR} - P_{MNAR}) \ln(N/N - 1). \tag{4}$$

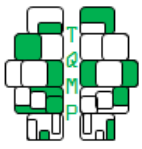


Table 3 ■ Estimates for Traditional MNAR Models (contd.)

| | Hybrid | Selection | | | | |
|----------------------|----------------|----------------|----------------|----------------|--------------------|----------------|
| | Diffuse | Focal | Diffuse (Lag) | Diffuse (IV) | Diffuse (Lag & IV) | Multinomial |
| RT → Intercept | -1.80 (0.46)** | -1.78 (0.47)** | -1.78 (0.47)** | -1.77 (0.47)** | -1.78 (0.47)** | -1.75 (0.47)** |
| SFAS → Slope | -1.22 (0.46)** | -1.25 (0.51)* | -1.26 (0.51)* | -1.25 (0.51)* | -1.26 (0.51)* | -1.29 (0.46)* |
| Estimated Parameters | 23 | 18 | 19 | 20 | 21 | 20 |
| LogL | -3823.51 | -4019.78 | -4019.43 | -4018.86 | -4018.59 | -4264.22 |
| AIC | 7693.02 | 8075.56 | 8076.87 | 8077.72 | 8079.19 | 8568.44 |
| BIC | 7784.42 | 8147.08 | 8152.37 | 8157.19 | 8162.64 | 8647.92 |
| ΔAIC | NA | -2.48 | -2.88 | -2.67 | -3.29 | NA |
| ΔBIC | NA | -6.45 | -6.86 | -6.65 | -7.26 | NA |

Note. **: $p < .01$; * : $p < .05$; NA = not applicable

An individual case is considered influential if $\Delta AIC_i^{INF} > \Delta AIC$ or $\Delta BIC_i^{INF} > \Delta BIC$. If either condition is observed, excluding the influential case will change the magnitude, sign, and interpretation of the ΔAIC or ΔBIC values (Sterba & Gottfredson, 2015). Although none of these values can definitively diagnose MAR or MNAR at the sample or individual levels, model comparisons and individual influence values can contribute useful information to an extensive MNAR sensitivity analysis (Enders, 2022; Molenberghs & Verbeke, 2001; Rubin, 1977; Sterba & Gottfredson, 2015). Individual influence diagnostics were computed for all eight traditional MNAR models shown in Tables 2 and 3, but none of the values exceeded the ΔAIC or ΔBIC values for any given model.

In the examples that follow, twelve traditional and eight trajectory mixture MNAR models will be presented not only to describe their different model assumptions and specifications, but also to underscore an important point: an extensive MNAR sensitivity analysis is necessary and should never be performed either on an arbitrarily selected single MNAR model or an arbitrarily selected subset of MNAR modeling possibilities. However, and somewhat counter-intuitively, investigating a MNAR suspicion begins with estimating a model that assumes MAR.

MAR Model

A longitudinal structural equation model (SEM), by default in most statistical analysis software packages, will assume all missing data to be missing at random (MAR). For the Murphy et al. (2019) study data, MAR assumes that observed BINGE scores at an earlier time point, random assignment, or both are related to missing BINGE data at a later assessment timepoint. Preliminary unconditional longitudinal model analyses of the BINGE data conducted for this tutorial showed significant fixed effects for the intercept ($\alpha_0 = 5.61$; $p < .001$), linear slope ($\alpha_1 = -6.60$; $p < .001$), and quadratic change ($\alpha_2 = 5.47$; $p < .001$), and sig-

nificant random effects for the intercept ($\text{VAR}[\zeta_{0i}] = \Psi_{00} = 11.08$; $p < .001$) and linear slope ($\text{VAR}[\zeta_{1i}] = \Psi_{11} = 4.30$; $p < .05$) best modeled average BINGE drinking behavior changes over time. Note that the quadratic random effect was non-significant, meaning the variation in non-linear changes in BINGE drinking across participants over time was not significantly different from zero. The quadratic random effect (i.e., quadratic variance) was fixed to zero in all analyses; $\Psi_{22} = 0$.

The conditional longitudinal MAR analysis model that included RT and SFAS as predictors of intercept and linear slope variance for the BINGE trial data is shown in Figure 1 (see supplemental appendix A, p. 1 for linear model Equations 1-4 and the relevant Mplus input syntax). Conditional longitudinal model analysis results showed that RT participants had significantly lower BINGE scores at baseline ($\gamma_{01} = -1.77$; $p < .001$) and SFAS participants showed significantly lower BINGE scores over time ($\gamma_{12} = -1.24$; $p < .05$; see Table 2). If the missing response variable data at a given assessment time point can be explained by observed response variable data at a previous time point, randomization, or both consistent with a MAR assumption, RT and SFAS significantly reduce binge drinking behavior.

It is important to note that trialists have long been encouraged to validate treatment effect conclusions with sensitivity analyses that test the robustness of results to MAR assumption violations, but research shows that only $\approx 30\%$ do so (Bell et al., 2014; Powney et al., 2014) despite several available published resources (e.g., Creemers et al., 2010, 2011; Cro et al., 2020; Iddrisu & Gumedze, 2020; Kenward, 2015; Leacy et al., 2017; Mallinckrodt et al., 2014). Two popular and easily implemented techniques that leverage multiple imputation missing data handling procedures are briefly illustrated here: “jump to reference” and the “delta method”.

MAR assumes that the response variable distribution for treatment group participants who drop out at any time

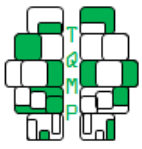
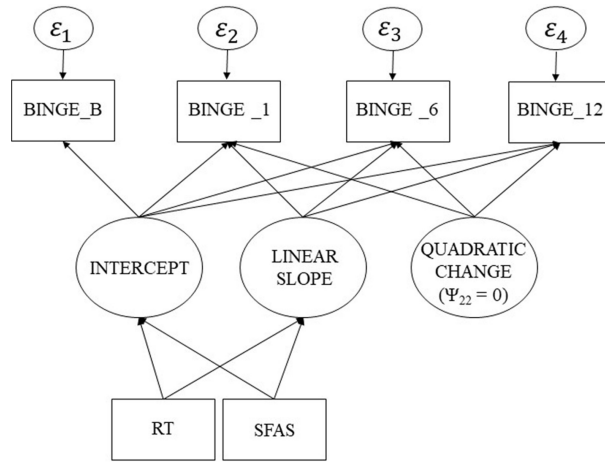


Figure 1 ■ MAR Model for Murphy et al. (2019) BINGE



point is more like the distribution for treatment group participants who completed the study (Carpenter & Kenward, 2007) consistent with an “intent to treat” (ITT) data analysis perspective (Little & Yau, 1996). In contrast, an “as-treated” perspective assumes, by virtue of removing themselves from exposure to treatment, that the response variable distribution for treatment group participants who drop out is more like the distribution for control group participants (Little & Yau, 1996; Ratitch et al., 2013). The “Jump-to-reference” procedure uses multiple imputation to test MAR and intent-to-treat assumptions by handling missing data from an “as-treated” perspective in four steps: 1) Identify participants with complete data in the two treatment groups (RT & SFAS; e.g., pattern 1 in Table 1) and save their data to a separate file; 2) with the remaining data, use the MAR model (Figure 1) to conduct model-based (or H_0 : model) multiple imputation to obtain m imputed datasets (see Graham et al., 2007, for recommendations); 3) merge the treatment group complete data file from Step 1 back into each of the m imputed data files created in Step 2; 4) analyze the merged m multiple imputation data files using the MAR analysis model and obtain pooled estimates (Ratitch et al., 2013; Kenward, 2015; Mallinckrodt et al., 2014). Jump-to-reference follows an as-treated perspective because the treatment group missing data in the Step 2 file will be imputed based heavily on control group participants with no missing data. Results from applying the jump-to-reference procedure ($m = 20$) to the BINGE data showed significant treatment group effects: RT ($\gamma_{01} = -1.71, p < .05$), SFAS ($\gamma_{12} = -1.11, p < .05$), suggesting treatment group effects are robust to deviations from specific MAR and intent-to-treat assumptions.

In contrast to the jump-to-reference method that tests

the ITT assumption, the delta method tests a very different and hypothetical scenario. Specifically, if treatment group participants who dropped out had remained in the study and their BINGE drinking scores could have been collected, then a reasonable question could be posed: How much of an increase in BINGE drinking scores would need to have been observed before SFAS treatment group effects (recall that RT had no effect on binge drinking scores over time) would no longer be significant? The delta method refers to a quantity (delta [δ]) that is computed in Cohen’s d response variable units and added to the multiply imputed BINGE response variable scores of treatment participants who dropped out (Cro et al., 2020). Specifically, the observed average post-baseline BINGE standard deviation was 4.15, so a value of 4.15 defines a Cohen’s d value of one. Further, a delta value that indicates an increase in BINGE drinking consistent with a “small” effect size ($d = 0.25$) would be $\delta = (4.15 \times 0.25) = 1.04$. After using the MAR analysis model (Figure 1) to conduct model-based multiple imputation for all BINGE missing data, but prior to analyzing the imputed data, the value of $\delta = 1.04$ is added to: 1) the 1-, 6-, and 12-month BINGE scores for treatment group participants who dropped out post-baseline (i.e., Table 1, pattern 8); 2) the 6- and 12-month follow-up scores for treatment group participants who dropped out after the 1-month follow-up (i.e., Table 1, pattern 4); and 3) to the 12-month follow-up scores for treatment participants who dropped out after the 6-month assessment (i.e., Table 1, patterns 2 and 6). Delta method results showed that if the binge drinking scores of treatment group participants who dropped out could have been collected and showed an increase by only a small ($d = 0.25, \delta = 1.04$) effect size amount, the treatment effect for SFAS (i.e., recall BINGE

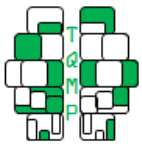
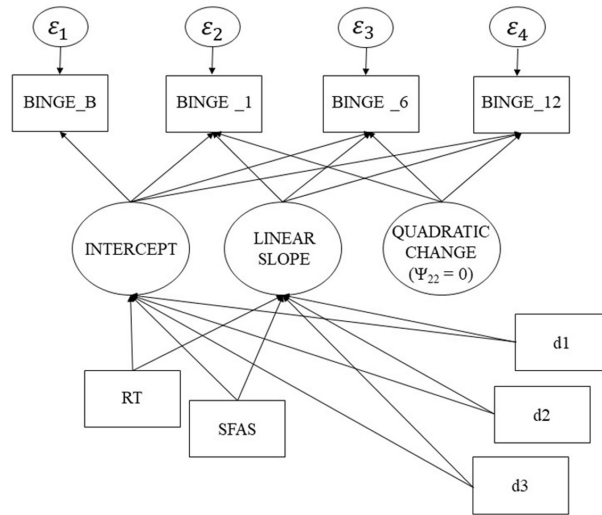


Figure 2 ■ A Focal Pattern Mixture MNAR Model



drinking scores significantly decreased over time for SFAS participants) would no longer be significant ($\gamma_{12} = -0.97$, $p > .05$). Jump to reference and delta method Mplus code examples are provided in the supplemental appendix (see Appendices A1-A4, pp. 2-5).

At this point, many researchers might understandably be experiencing confusion. On the one hand, researchers would be satisfied with both significant treatment effect estimates as well as the additional information gained from further “jump to reference” and delta method exploration. On the other hand, researchers could reasonably be wondering: “If MNAR is suspected, why bother starting with a MAR analysis and pursuing subsequent sensitivity testing?” Three answers can be given to this question, all of which are important to keep in mind when undertaking a MNAR investigation. First, results from MAR stress tests show the conditions under which treatment effect estimates remain (“jump to reference”) or do not remain (delta method) significant. Second, although a longitudinal MAR model does nothing to appropriately address suspected MNAR data (Carpenter & Kenward, 2007), it is the most parsimonious data analytic model that handles missing data correctly (e.g., to avoid listwise deletion). Third, and most importantly, parsimonious MAR treatment effect estimates provide an acceptable starting point from which to determine whether additional MNAR analyses corroborate or contradict those results (Enders, 2022; Muthén et al., 2011).

Traditional MNAR Models

A sensitivity analysis to test suspected MNAR begins with three traditional MNAR models: pattern mixture, shared

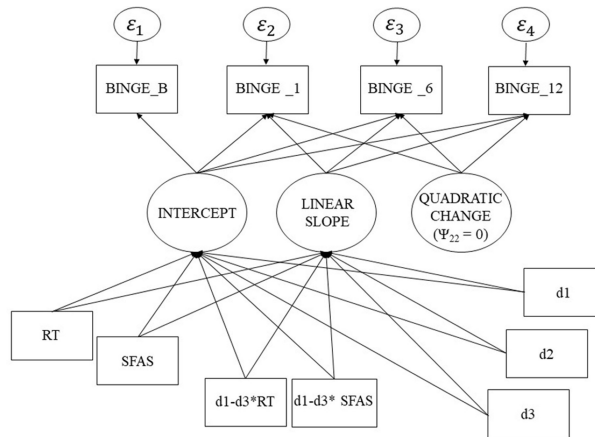
parameter, and selection. The three models will be specified to assess both focal and diffuse MNAR possibilities for the BINGE data. A fourth hybrid model is introduced and included as a test of the conditional independence assumption, as will be shown.

Pattern Mixture Model (Hedeker & Gibbons, 1997; Little, 1995, 2009).

A focal pattern mixture MNAR model (Demirtas & Schafer, 2003; Hedeker & Gibbons, 1997; Little, 1995, 2009; Verbeke & Molenberghs, 2000) specified for the BINGE trial data analysis model is shown in Figure 2 (see supplemental Appendix B, p. 6, for linear model Equations 5-8 and relevant Mplus input syntax). Including dummy-coded missing data indicators as additional predictors of trajectory random effect variance (γ_{03} & γ_{13}) implies both the presence of substantively different subgroups with unique expected trajectories, and that treatment effects ($\gamma_{01} - \gamma_{12}$; see supplemental Appendix B, p. 6, for linear model Equations 5-8 and the relevant Mplus input syntax) need to be estimated conditional on the effects of dropout (d1-d3). Conditional independence in the pattern mixture model assumes that observed and missing response variable data are unrelated conditional on estimating dropout-specific trajectory fixed effects (Little, 1995, 2009). In addition, a pattern mixture MNAR model further assumes that: 1) BINGE response variables are normally distributed, 2) missing data is MAR within each dropout pattern, and 3) all dropout patterns have the same residual covariance matrix (Enders, 2011b, 2011a; Fitzmaurice et al., 2001; Hedeker & Gibbons, 1997, 2006; Hogan & Laird, 1997; Muthén et al., 2011).



Figure 3 ■ A Diffuse Pattern Mixture MNAR Model



Prior to estimation, properly identifying a pattern mixture MNAR model requires recalling MAR model results that showed a quadratic fixed effects model best captured BINGE changes over time (results assuming MAR). A minimum of three observed data points is needed to specify a quadratic model, so a quadratic model is not identified for participants dropping out at the first ($d1 = 1$) or second ($d2 = 1$) post-baseline assessment but is identified if a participant dropped out at the third post-baseline assessment ($d3 = 1$). Pattern mixture model identification is achieved by 1) centering linear slope and quadratic change latent variable loadings so that the intercept fixed effect is defined as the expected BINGE drinking value at the baseline assessment (Enders, 2011b), and 2) constraining separately to equality the effects of the dummy indicators as predictors of the intercept and linear slope random effects (γ_{03} & γ_{13} respectively). Results from fitting the focal pattern mixture model shown in Figure 2 and Table 2 to the BINGE data show that, conditional on dropout, RT ($\gamma_{01} = -1.78$; $p < .001$) and SFAS ($\gamma_{12} = -1.24$; $p < .05$) remained significant predictors of the BINGE intercept and slope random effects, respectively. Stated differently, conditional on modeling the predictive effects of different dropout groups on the average BINGE drinking trajectory as a missing data model, RT and SFAS again significantly reduced BINGE drinking behavior.

Further, as originally described in Hedeker and Gibbons (1997), a diffuse pattern mixture MNAR model can be specified by adding interactions between treatment groups and dummy dropout indicators, which is shown in Figure 3 (see supplemental Appendix C, p. 7, for linear model Equations 9-12 and relevant Mp1us input syntax) for the BINGE data. Model identification is achieved by constraining both the main effects of dummy indicators and the dummy in-

dicators by treatment group interaction terms as predictors of intercept and linear slope random effects separately to equality. Results from fitting the diffuse pattern mixture MNAR model in Figure 3 showed that both RT ($\gamma_{01} = -1.82$; $p < .001$) and SFAS ($\gamma_{12} = -1.31$; $p < .05$) again remained significant predictors of the intercept and slope random effects, respectively. Conditional on estimating the differential predictive effects of dropout groups on the average BINGE drinking trajectory as moderated by random assignment as a missing data model, RT and SFAS still significantly reduced BINGE drinking behavior. Further, as shown in Table 2, if the pattern mixture focal model and pattern mixture diffuse model are both compared against the MAR model, the ΔAIC (-3.35, -17.75) and ΔBIC (-11.30, -41.59) values are negative, suggesting support for the MAR model (Although not shown here, a random coefficient pattern mixture model [Little, 1993, 1994, 1995] can also be estimated. Examples can be found in Enders [2010, pp. 306-312; 2011b; 2011a; 2022, pp. 385-388]).

Shared Parameter Model (Wu & Carroll, 1988).

A shared parameter model assumes that response variable score changes over time, as captured by the growth trajectory random effects (assumed to be distributed normally), predict dropout (i.e., the survival indicators $s1 - s3$). Conditional independence for a shared parameter model (Wu & Carroll, 1988) assumes that observed response variable data and missing data indicators are independent conditional on shared parameters: the growth trajectory random effects. Further, because trajectory random effects predict dropout indicators, and random effects are estimated based on all available response variable data, shared parameter models are diffuse only; there is no focal shared



Figure 4 ■ Diffuse Shared Parameter MNAR Model #1

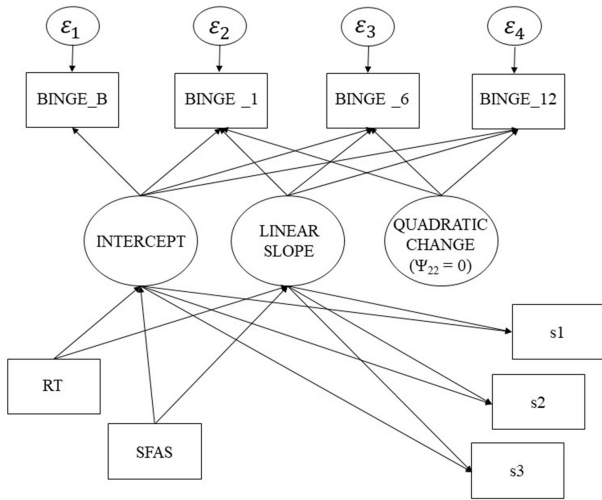
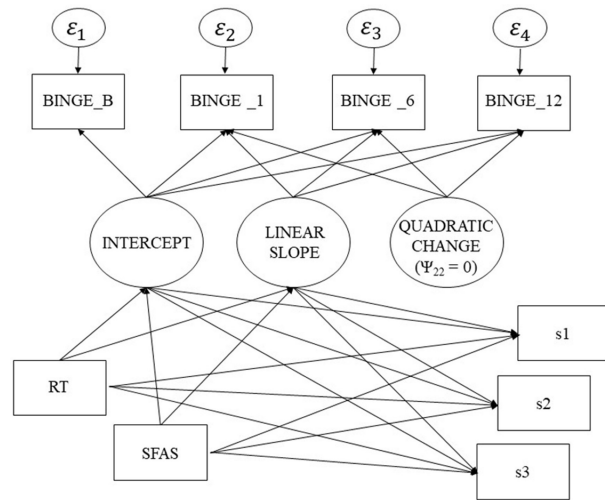


Figure 5 ■ Diffuse Shared Parameter MNAR Model #2



parameter model. A diffuse shared parameter model that specifies only random trajectory effects as predictors of dropout indicators is shown in Figure 4 (see supplemental Appendix D, p. 8, for linear model Equations 13-19 and the relevant Mplus input syntax) for the BINGE data. Shared parameter model identification is achieved by constraining the effects of intercept and slope random effects as predictors of dropout separately to equality. Results from fitting the first diffuse shared parameter MNAR model (see Table 2) to the BINGE data showed that both RT ($\gamma_{01} = -1.74$; $p < .001$) and SFAS ($\gamma_{12} = -1.25$; $p < .05$) remained significant predictors of intercept and slope random effects. Conditional on the inclusion of a missing data model that specifies BINGE drinking response variable trajectory components as predictors of dropout missing data, RT and SFAS again significantly predicted decreases in BINGE drinking behavior.

Further, a second diffuse shared parameter model that includes independent variable groups as predictors of missing data is shown in Figure 5 (see supplemental Appendix E, p. 9, for linear model Equations 20-26 and the relevant Mplus input syntax). As shown in Table 2, results from estimating the additionally diffuse shared parameter model again indicated that both RT ($\gamma_{01} = -1.76$; $p < .001$) and SFAS ($\gamma_{12} = -1.25$; $p < .05$) remained significant predictors of intercept and slope random effects, respectively. Conditional on estimating the effects of both randomization and BINGE drinking trajectories as a missing data model, RT and SFAS still significantly reduced BINGE drinking behavior. Additionally, if the second diffuse shared parameter model is estimated by replacing the survival dropout indicators (s1-s3) with multinomial missing data indicators (m1-m3 in Table 3) as a sensitivity anal-

ysis (see supplemental Appendix F, p. 10, for the relevant Mplus input syntax), the treatment effect estimates both remained significant and changed negligibly. Finally, a shared parameter MAR model can be estimated by fixing the effects of random trajectory latent variables as missing data indicator predictors to zero. If such a shared parameter MAR model is compared to both diffuse models, negative ΔAIC (-1.68, -3.45) and ΔBIC (-1.68, -11.40) values again suggested support for the MAR model.

Hybrid MNAR Model (Yuan & Little, 2009).

A hybrid MNAR model proposed by Yuan and Little (2009) that expands upon the second diffuse shared parameter model shown previously to include droptime as a random effects predictor consistent with pattern mixture specifications is shown Figure 6 (see supplemental Appendix G, pp. 11-12, for linear model Equations 27-33 and the relevant Mplus input syntax) for the BINGE data. The shared parameter model in Figure 5 is nested within the hybrid model, so that the hybrid model can be viewed as a test of the shared parameter model conditional independence assumption using nested model comparisons (Yuan & Little, 2009). Results from fitting the hybrid model to the BINGE data (Table 3) again showed significant treatment effect estimates: RT ($\gamma_{01} = -1.80$, $p < .001$), SFAS ($\gamma_{12} = -1.22$, $p < .01$). The hybrid model ($LogL = -3,823.51$; $BIC = 7,784.42$) was also shown to be a significantly better fit to the BINGE data versus the second diffuse shared parameter model ($LogL = -4,018.25$; $BIC = 8,161.94$; $\chi^2_2 = 389.48$, $p < .001$), which calls into question the shared parameter model conditional independence assumption. Even if both different dropout groups and observed BINGE drinking trajectories are included in a model of dropout missing data



Figure 6 ■ A Hybrid MNAR Model

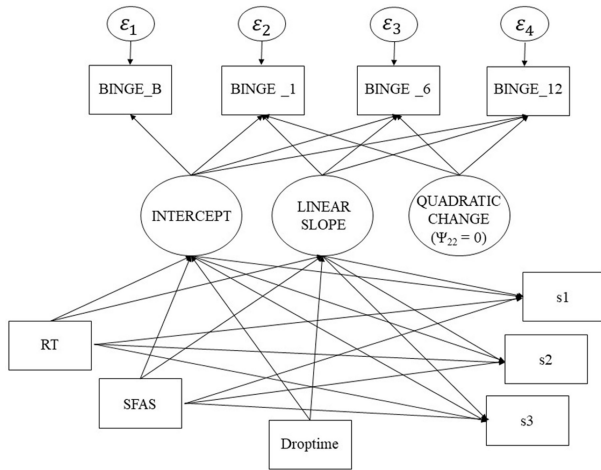
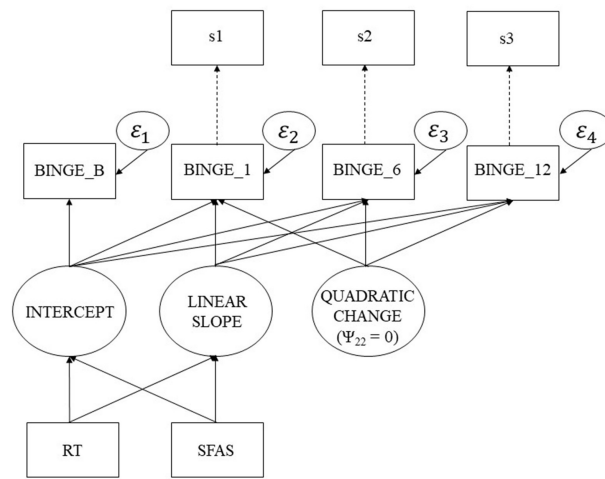


Figure 7 ■ A Focal Selection MNAR Model



and given that the conditional independence assumption was shown to be questionable as a result, RT and SFAS were still shown to significantly reduce BINGE drinking behavior.

Selection MNAR Model (Diggle & Kenward, 1994).

In direct contrast to assuming conditional independence, a selection MNAR model (Heckman, 1976, 1979) assumes survival dropout indicators at specific time points are predictable by observed BINGE scores at the same time points because the missing data itself contains information about the unobserved response variable values. However, selection MNAR models are inestimable unless a strong assumption of multivariate normally distributed data is invoked because response variable data at each post-baseline assessment time point are observed if $s_1-s_3 = 0$ but are missing if $s_1-s_3 = 1$. A focal selection MNAR model is shown in Figure 7 (see supplemental Appendix H, p. 13, for linear model Equations 34-40 and the relevant Mplus input syntax) for the BINGE data. Selection model identification is achieved by constraining the effects of BINGE scores at a given time point as predictors of missing data at the same time point to equality. Results from fitting the focal selection MNAR model to the BINGE data (Table 3) again showed significant treatment group effects: RT ($\gamma_{01} = -1.78, p < .001$), SFAS ($\gamma_{12} = -1.25, p < .05$). Stated differently, if the unobserved values contain information about why BINGE drinking data is missing, and indicators of missing data at a given assessment time point are predicted by assessment time point-specific observed BINGE drinking data as a missing data model, RT and SFAS remained significant predictors of reduced BINGE drinking behavior,

Also shown in Table 3 are three possible diffuse se-

lection MNAR models for the BINGE drinking data. The first diffuse selection model adds observed BINGE drinking scores at the previous assessment time point, or lagged missing data predictions, to the missing data model as shown in Figure 8 (see also supplemental Appendix I, p. 14 for linear model Equations 41-47 and relevant Mplus input syntax). The effects of the lagged predictions are constrained to equality over time to facilitate model identification. Results from fitting the diffuse selection MNAR model with lagged predictions to the BINGE data (Table 3) again showed significant treatment group effects: RT ($\gamma_{01} = -1.78, p < .001$), SFAS ($\gamma_{12} = -1.26, p < .05$).

A second diffuse selection MNAR model replaces lagged predictions with independent variable groups as predictors in the missing data model as shown in Figure 9 (see supplemental Appendix J, p. 15, for linear model Equations 48-54 and the relevant Mplus input syntax), where the effects of the independent variable groups as missing data predictors are also separately constrained to equality. Results from fitting the diffuse selection MNAR model with independent variable group missing data predictions to the BINGE data (Table 3) also again showed significant treatment group effects: RT ($\gamma_{01} = -1.77, p < .001$), SFAS ($\gamma_{12} = -1.25, p < .05$).

The third diffuse selection MNAR model combines the previous two by adding both lagged predictors and independent variable groups to the missing data model, with all effects separately constrained to equality, as shown in Figure 10 (see supplemental Appendix K, p. 16, for linear model Equations 55-61 and relevant Mplus input syntax). Results again showed significant treatment effects: RT ($\gamma_{01} = -1.78, p < .001$), SFAS ($\gamma_{12} = -1.26, p < .05$). Further, if the focal selection MNAR model with both lagged and independent variable predictors is estimated by re-



Figure 8 ■ Diffuse (Lag Prediction) Selection MNAR Model

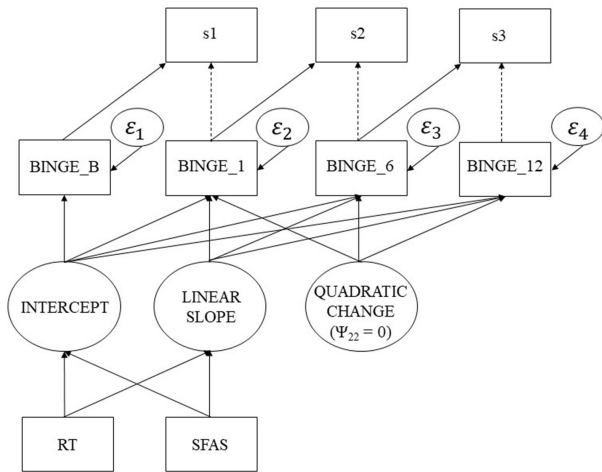
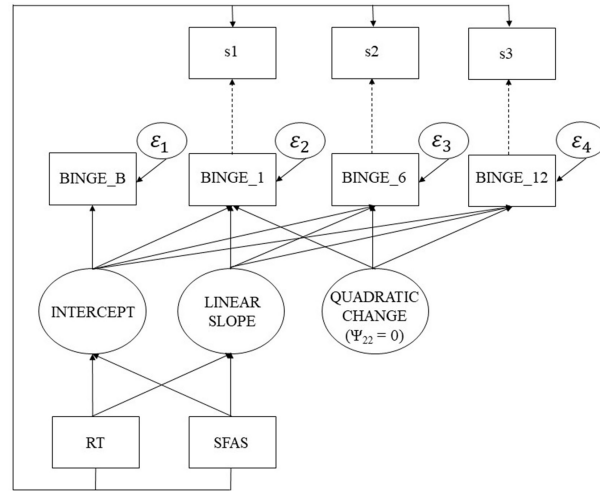


Figure 9 ■ Diffuse (IV Group Prediction) Selection MNAR Model



placing the survival dropout indicators (s1-s3) with multinomial missing data indicators (m1-m3; see supplemental Appendix L, p. 17, for the relevant Mplus input syntax) in Figure 10 as a sensitivity analysis (Table 3), the treatment group effect estimates again remain significant and changed negligibly. Taken together, diffuse selection model results suggested that conditional on estimating missing data explanatory models that include same time point response variable data, previous time point response variable data, and random assignment as predictors of dropout, RT and SFAS remained significant predictors of decreased BINGE drinking behavior.

At this point, a researcher might wonder why five different selection models are needed. Two answers can be offered. First, both focal and diffuse MNAR are plausible possibilities for missing BINGE data, so both warrant investigation. Second, the purpose of a MNAR sensitivity analysis is to test the robustness of treatment effect estimates as model specification (e.g., focal and diffuse) and missing data code (survival or multinomial) changes are implemented to reflect different missing data assumptions. Finally, two conclusions can be drawn from examining all selection MNAR model results in Table 3. First, treatment effects remained significant across all models and the estimates varied negligibly. Said differently, testing different MNAR assumptions did not change initial MAR treatment effect analysis results. Second, in contrast to the pattern mixture and shared parameter models, there is no single appropriate comparison MAR selection model. The appropriate MAR model for each selection MNAR model is obtained by constraining the effects of predicting missing data indicators (s1-s3) at each time point by observed re-

sponse variable data at that time point (shown as dashed arrows in Figures 7-10) to zero. When each selection MNAR model is compared to its appropriate MAR counterpart, all ΔAIC (range: -2.48, -3.29) and ΔBIC (range: -6.45, -7.26) values were negative, suggesting support for the MAR models.

Mixture Trajectory MNAR Models

Traditional MNAR models have expanded to include finite mixture modeling specifications for two reasons. First, modeling the unknown relationship between observed response variable data and missing response variable data is more easily accomplished within mixture models that both allow trajectory random effect means and missing data indicator point estimates to vary across unobserved mixture trajectories. This allows the unknown observed data and missing data distributional relationship to be modeled via semi-parametric “support” or “pillar” points defined by the separate response variable trajectories (e.g., Nagin, 1999) which reduces the risk of model misspecification. Second, all traditional MNAR models assume either normally distributed response variable or normally distributed latent trajectory variable data. Semi-parametric longitudinal mixture modeling can better handle both non-normal response variable (see Micceri, 1989) data (which violates pattern mixture and selection MNAR model assumptions) and non-normal trajectory random effects (which violate shared parameter model assumptions; e.g., see Bauer & Curran, 2003). As such, longitudinal finite mixture MNAR models can be viewed as additional sensitivity tests of traditional MNAR model normality assumptions (Enders, 2011b, 2011a; Muthén et al., 2011). However, it is important



Figure 10 ■ A Diffuse (Lag & IV Group Prediction) Selection MNAR Model **Figure 11** ■ MAR Mixture Model

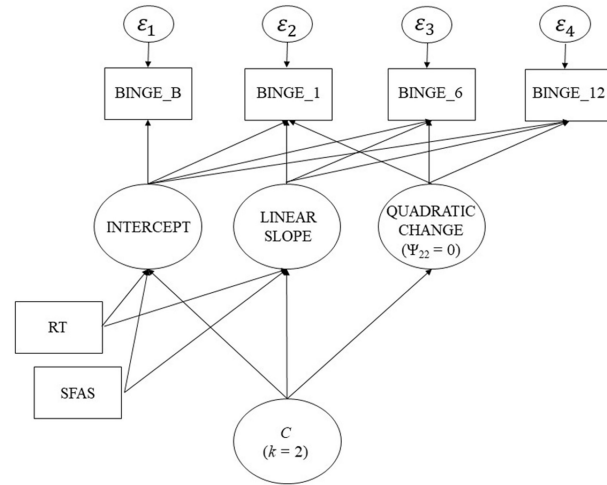
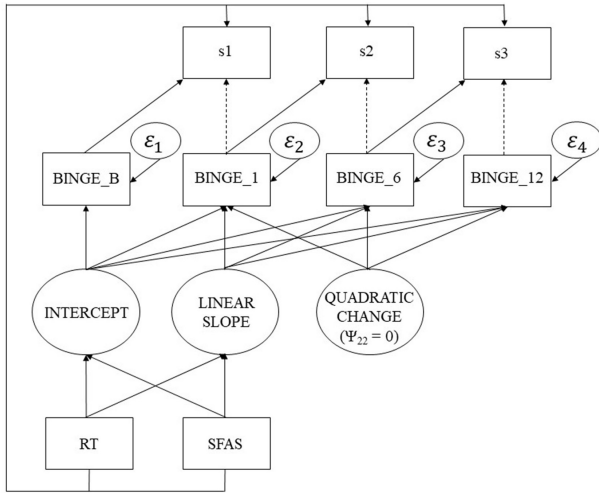


Table 4 ■ Estimates for Mixture MNAR Models (all $k = 2$ latent trajectories)

| | | | | | | Selection (Muthén et al., 2011) | | |
|----------------|----------------|----------------------|-------------------------|------------------------|---------------------|---------------------------------|----------------|----------------|
| | MAR | Dropout ¹ | Muthén-Roy ² | Beunckens ³ | Hybrid ⁴ | Model 1 | Model 2 | Model 3 |
| RT → Intercept | -1.23 (0.36)** | -1.22 (0.36)** | -1.06 (0.36)** | -1.18 (0.35)** | -1.76 (0.47)** | -0.73 (0.35)* | -1.05 (0.35)** | -1.12 (0.36)** |
| SFAS → Slope | -1.26 (0.51)* | -1.26 (0.50)* | -1.16 (0.46)* | -1.28 (0.49)** | -1.26 (0.51)* | -1.05 (0.47)* | -1.24 (0.52)* | -1.21 (0.52)* |
| Estimates | 18 | 18 | 24 | 23 | 25 | 29 | 30 | 40 |
| LogL | -3763.16 | -3763.99 | -3728.11 | -3961.55 | -4093.74 | -3829.84 | -3881.57 | -3850.1 |
| AIC | 7562.32 | 7563.97 | 7504.22 | 7969.09 | 8237.48 | 7717.67 | 7823.14 | 7780.20 |
| BIC | 7633.85 | 635.50 | 7599.59 | 8060.49 | 8336.82 | 7832.91 | 7942.35 | 7812.23 |

Note. **: $p < .01$; *: $p < .05$; NA = not applicable. ¹: Roy (2003); ²: Muthén et al. (2011); ³: Beunckens et al. (2008); ⁴: Gottfredson et al. (2014).

to note that if no evidence for heterogeneity is found (i.e., if a MAR mixture model does not fit better than a MAR model assuming homogeneity), mixture modeling need not be pursued.

A longitudinal mixture model assuming MAR is shown in Figure 11 (see also supplemental Appendix M, p. 18, for linear model Equations 62-66 and the relevant Mplus input syntax) for the BINGE data, where C represents a categorical latent variable that contains $k = 2, 3, \dots, K$ latent mixture trajectories. Further, when $k = 2-4$ possible mixture trajectories were tested, a $k = 2$ trajectory mixture MAR best modeled BINGE drinking heterogeneity and better fit the BINGE data ($LogL = -3,763.16$, $BIC = 7,633.85$) versus the conditional MAR ($k = 1$) model (Table 2; $LogL = -3,820.60$, $BIC = 7,724.82$) and treatment group effect estimates remained significant: RT ($\gamma_{01} = -1.23$, $p = .001$), SFAS ($\gamma_{12} = -1.26$, $p < .05$). This suggests that the BINGE data treatment group effects could be further tested with additional MNAR mixture models. For all mixture MNAR analyses that follow, $k = 2-4$ mixture models were tested, but models that resulted in estimation errors or extracted mixture trajectories containing less than 10%

of the sample (Nagin, 1999) were not considered. As a result, a $k = 2$ trajectory model best fit the BINGE data for all mixture MNAR models that follow. The mixture MAR analysis results showed that if BINGE sample data heterogeneity is modeled by estimating two unobserved subgroups with distinctly different drinking trajectories, RT and SFAS remained significant predictors of decreased BINGE drinking.

Roy (2003) Dropout Model.

The traditional pattern mixture model assumes that participants who drop out at the same time point have the same response variable distribution. Roy's (2003) mixture model extracts unobserved trajectories based on response variable distributional similarities, which is assumed to be a better dropout model and avoids the miss-categorization that could occur in pattern mixture models. Roy's model further assumes that, conditional on mixture trajectory extraction with dummy dropout indicators as predictors, response variable distributions are uncorrelated with dummy dropout indicators (conditional independence; Dantan et al., 2008) and missing data within

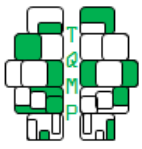


Figure 12 ■ Roy's (2003) Dropout Model

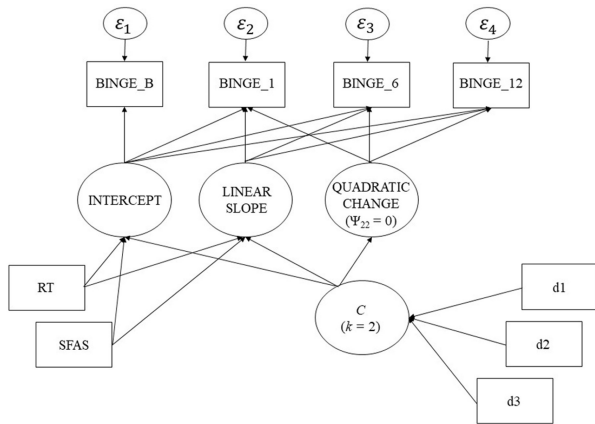
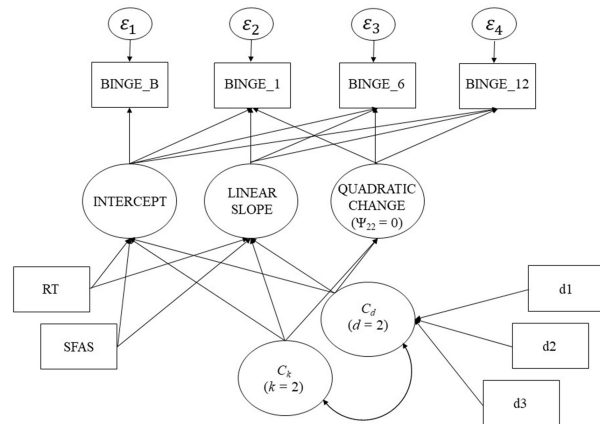


Figure 13 ■ Muthén-Roy Model



each mixture trajectory is MAR. Roy's dropout model is shown in Figure 12 (see supplemental Appendix N, p. 19, for linear model Equations 67-71 and the relevant Mplus input syntax). Results from fitting the Roy's ($k = 2$) model to the BINGE data (Table 4) shows that significant treatment group effects remained: RT ($\gamma_{01} = -1.22, p < .001$), SFAS ($\gamma_{12} = -1.26, p < .05$). Conditional on estimating a missing data model that assumes dropout missing data is better modeled via mixture trajectory extraction based on dropout group prediction rather than modeled with pattern mixture model assumptions and possible misclassification, results again showed RT and SFAS to significantly reduce BINGE drinking behavior. However, critics of the Roy (2003) dropout model (see Muthén et al., 2011) note that mixture trajectories are extracted based only on observed response variable data, not missing data. As a result, mixture trajectory distributions may be confounded with, rather than independent of, dropout.

Muthén et al. (2011) Model.

The Muthén-Roy (Muthén et al., 2011; see also Power et al., 2012) model addresses the Roy model criticism of finite mixture trajectories being confounded with dropout missing data by estimating two latent categorical variables: one that contains $k = 2, 3, \dots, K$ continuous response variable mixture trajectories, the other that contains $d = 2, 3, \dots, D$ binary dropout missing data mixture classes. Rather than assuming independence between the mixture trajectories and dropout indicators, the Muthén-Roy model assumes that latent response variable finite mixture trajectories are moderated by latent dropout classes. The Muthén-Roy model is shown in Figure 13 (see supplemental Appendix O, pp. 20-21, for linear model Equations 72-77 and the relevant Mplus input syntax) for the BINGE data.

Recall that specifying and estimating unidentified pa-

rameters in the pattern mixture model was aided by centering slope loadings such that the intercept mean was defined as the expected BINGE value at baseline. This specification is extended to the mixture Muthén-Roy model by separately constraining mixture trajectory intercept means to equality across dropout missing data classes (i.e., see supplemental Appendix O, p. 20; α_{0k} in Equation 73 has no d subscript; Muthén et al., 2011). Results from fitting the Muthén-Roy model ($k = 2, d = 2$) to the BINGE data again (Table 4) showed significant treatment group effects: RT ($\gamma_{01} = -1.06, p < .01$), SFAS ($\gamma_{12} = -1.16, p < .05$). Conditional on estimating a dropout missing data model that specifies latent response variable mixture trajectories moderated by latent dropout classes, Muthén-Roy model results still showed that RT and SFAS significantly reduced BINGE drinking behaviors.

Beunckens et al. (2008) Model.

Beunckens et al. (2008) developed a mixture extension of a shared parameter (Wu & Carroll, 1988) model (see Equation 7, p. 101 in Beunckens et al., 2008) that was designed to minimize the number of estimated parameters without a loss of information. The Beunckens model is shown in Figure 14 (see supplemental Appendix P, pp. 22-23, for linear model Equations 78-85 and the relevant Mplus input syntax) for the BINGE data. The intercept random effect predicts survival dropout indicators at all post-baseline time points, and those prediction slopes are constrained to equality for parsimony and model identification purposes. This specification illustrates a key assumption of the Beunckens et al. (2008) model: that the response variable model and the missing data model are now indirectly related (conditional independence) through the shared intercept (η_{0i} ; Beunckens et al., 2008; Muthén et al., 2011). The effects of RT and SFAS as predictors of dropout miss-

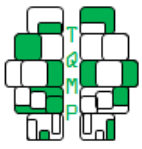


Figure 14 ■ Beunckens et al. (2008) Model. Gray stars indicate logit intercepts re-scaled to follow a quadratic trend.

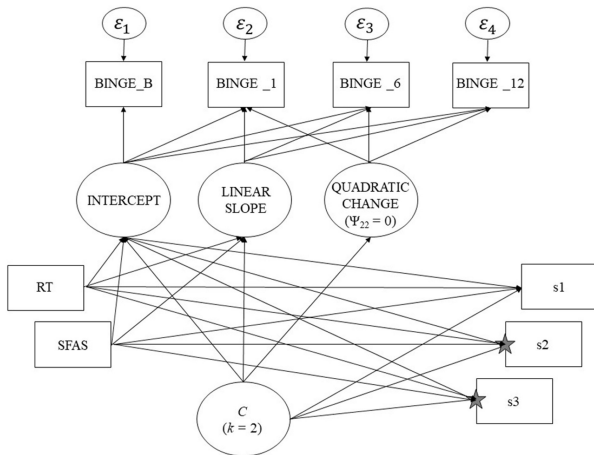
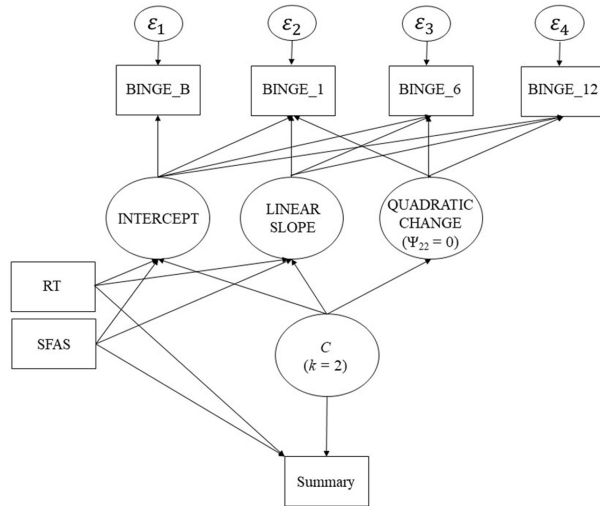


Figure 15 ■ Gottfredson et al. (2014) Shared Parameter Hybrid Mixture Model (SPMM)



ing data are also constrained to equality. Further, as indicated in Figure 14 (gray stars), the s2 and s3 survival dropout logit intercepts are re-scaled to follow a quadratic fixed effect trend similar to the response variable expected trajectory that best describes average BINGE drinking over time. Fitting the Beunckens et al. model ($k = 2$) to the BINGE data (Table 4) again showed significant treatment group effects: RT ($\gamma_{01} = -1.18, p < .01$), SFAS ($\gamma_{12} = -1.28, p < .01$). Conditional on the estimation of a missing data model that includes mixture trajectory extraction and assumes that a shared trajectory intercept, randomization, and latent mixture trajectory membership adequately predict dropout missing data indicators whose logit intercepts have been constrained to follow a quadratic trend, RT and SFAS remained significant predictors of decreased BINGE drinking.

Gottfredson et al. (2014) Hybrid Model.

Similar to the Yuan and Little (2009) hybrid MNAR model, Gottfredson et al. (2014) developed a hybrid mixture MNAR model: a shared parameter mixture model (SPMM). Like the shared parameter model, the SPMM also assumes that longitudinal trajectory random effects are the mechanism behind missing data and the SPMM will produce biased results if this assumption is false. The SPMM, like the shared parameter model, also assumes conditional independence: BINGE trajectory random effects and survival missing data codes are independent conditional on finite mixture trajectories. The mixture trajectories serve as the new shared parameters to model the relationship between trajectory random effects and missing data indicators. This is assumed to reduce parameter estimate bias due to non-normal trajec-

tory random effects or missing data model misspecification (Gottfredson et al., 2014). A Gottfredson SPMM is shown in Figure 15 (see supplemental Appendix Q, pp. 24-25, for linear model Equations 86-91 and the relevant Mplus input syntax) for the BINGE data.

In contrast to previous MNAR models that required dummy or survival missing data models, SPMM assumes that including a single missing data indicator variable (summary) is sufficient as a missing data model. Further, the SPMM specifies that each trajectory-specific mean estimate is multiplied by the proportion of the sample each mixture trajectory contains, then pooled across mixture trajectories (see supplemental Appendix Q, p. 24; the p superscripts shown in Equations 87-89 represent pooling of the three trajectory means [$\alpha_{0k}^p, \alpha_{1k}^p, \& \alpha_{2k}^p$]). Results from fitting the Gottfredson SPMM ($k = 2$) to the BINGE data (Table 4) again showed significant treatment group effects: RT ($\gamma_{01} = -1.76, p < .001$), SFAS ($\gamma_{12} = -1.26, p < .05$). Conditional on the estimation of a missing data model that assumes mixture trajectory membership and random assignment as predictors of a single ‘summary’ missing data indicator, RT and SFAS were again shown to be significant predictors of decreased BINGE drinking,

Selection Mixture Model (Muthén et al., 2011).

A selection mixture MNAR model is created by expanding the traditional selection model to allow for mixture trajectory extraction. The selection model for the BINGE data is shown in Figure 16 (see supplemental Appendices R-T, pp. 26-31, and linear model Equations 92-115). Also shown in Figure 16 (gray stars) is the specification that survival missing data logit intercepts are again constrained to

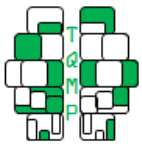
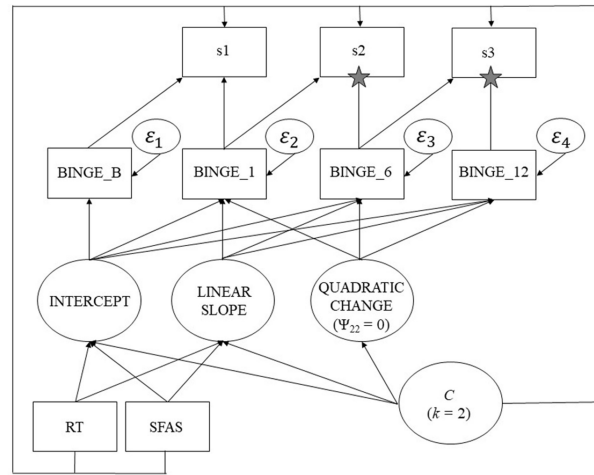


Figure 16 ■ Selection Mixture Model. Gray stars indicate logit intercepts re-scaled to follow a quadratic trend.



follow a quadratic trend consistent with the average expected change in BINGE scores over time as was specified in the Beunckens et al. (2008) model. As shown in Muthén et al. (2011), an internal sensitivity analysis can be performed by estimating three selection mixture models (Model 1 – Model 3) with each successive model increasing in the number of estimated parameters. Model 1 is specified by constraining missing data indicator prediction slopes to equality both across time and between mixture trajectories, while the missing data indicator logit intercepts are estimated freely over time but constrained to equality between mixture trajectories. Model 2 specifications differ from Model 1 in that both missing data indicator prediction slopes and missing data indicator logit intercepts are estimated freely across mixture trajectories, but not over time. Model 3 specifications differ from Model 2 in that missing data indicator prediction slopes are freely estimated both over time and across mixture trajectories.

Results from fitting selection mixture ($k = 2$) Models 1-3 to the BINGE data (Table 4) again showed significant treatment effects across models: Model 1: RT ($\gamma_{01} = -0.73, p < .05$), SFAS ($\gamma_{12} = -1.05, p < .05$); Model 2: RT ($\gamma_{01} = -1.05, p < .01$), SFAS ($\gamma_{12} = -1.24, p < .05$); Model 3: RT ($\gamma_{01} = -1.12, p < .01$), SFAS ($\gamma_{12} = -1.21, p < .05$). Conditional on estimating a missing data model that extends the assumptions of the focal and diffuse selection models to a mixture model that includes constraints on the missing data indicator logit intercepts consistent with Beunckens et al.'s (2008) model specification, RT and SFAS still remained significant predictors of decreased BINGE drinking behavior.

Summary Discussion and Specific Recommendations

Murphy et al. (2019) conducted a longitudinal trial to test if brief MI paired with either relaxation training (RT) or substance-free activity sessions (SFAS) effectively reduced binge drinking in a sample of undergraduate college students. MNAR was suspected because participants that dropped out of the BINGE trial could have done so due to consequences associated with binge drinking behavior. Recall how an MNAR investigation should begin with the inclusion of effective auxiliary correlates (Enders, 2022), but none of the additional data collected in the BINGE trial met criteria for auxiliary correlate inclusion. On the one hand, a MNAR-by omission process remains a possibility due to a lack of auxiliary variables. On the other hand, Thoemmes and Rose (2014) have shown that, under specific conditions, auxiliary correlates can increase, rather than decrease, bias and complicate the MNAR issue in unknown ways.

At this point, research trialists could ask two practical questions regarding suspected MNAR investigations. Given that MNAR: 1) cannot be handled with any MAR missing data handling method, 2) can only be suspected and for which no definitive test exists, 3) requires the addition of a model that describes the possible relationship between observed and missing response variable data, and 4) all possible models rely on assumptions that cannot be tested, a researcher would understandably wonder, first, what results should be reported from an MNAR sensitivity test. There are two pitfalls to be avoided. As stated previously, researchers should conduct as thorough a MNAR sensitivity testing procedure as is possible and avoid the temptation to



test MNAR suspicions with a single model or arbitrarily selected subset of models. Such temptations assume that the selected model optimally describes the observed response data and missing response data relationship, which cannot be verified. Second, researchers should avoid the temptation to present only the subset of models showing the most favorable treatment effects.

Muthén et al. (2011) assert that the dispositive question is whether treatment effect estimates from MAR analysis results are corroborated or contradicted by an MNAR sensitivity analysis. Enders (2022) further asserts that all MNAR models are possible plausible explanations for why response variable data might be missing. Rather than an “either/or” approach to Muthén et al. (2011) and Enders (2022), we recommend a “both/and” approach to be most helpful to trialists who are tasked with answering the question of whether a treatment is beneficial. For example, regarding the BINGE trial results, and in keeping with both Enders (2022) and Muthén et al. (2011), the results of the MAR analysis model are corroborated by an extensive MNAR sensitivity investigation: RT and SFAS significantly reduce binge drinking behavior. Further, although not definitive, a researcher could also cite: 1) Δ AIC and Δ BIC statistics that show, when each MNAR model is compared against its appropriate MAR counterpart, results suggested support for the MAR model, 2) “jump to reference” results showed that MAR BINGE results were robust to deviations from an intent to treat analytic perspective, but 3) delta modeling results showed that the significance for substance-free activity sessions (SFAS) reducing binge drinking are questionable if dropouts hypothetically increased their binge drinking by only a relatively small ($d = 0.25$) effect size. Further, consistent with Enders (2022), researchers can present the results of all models, such as in Tables 2-4, in the interests of transparency and further state that, although very different in magnitude, the treatment effects for RT and SFAS remained significant and in the expected direction, and that such estimate variation is to be expected because the different MNAR models make very different assumptions about the observed and missing response variable data relationship.

These suggestions beg an obvious second question: What should a researcher do if the results of an extensive MNAR sensitivity analysis produced a “mixed bag” of statistically significant and statistically non-significant treatment effect results? Treatment effect questions are central to longitudinal trials, and treatment effects that show statistical significance for some models and non-significance for others can be problematic. If available, two additional sources of information could prove helpful. First, Muthén et al. (2011) showed how a binary distal outcome variable can be added to MNAR models to further test treatment ef-

fect results. For example, in the BINGE trial, if a binary indicator of whether participants were subsequently placed on academic probation or suspension were available for inclusion, it could serve as a validity check on MNAR results. Second, researchers have long advocated for follow-up interviews to be conducted with longitudinal trial participants who drop out. Such interviews could provide specific information as to the relationship between the response variable and why participants left the study. Such information is most critical prior to any attempt to estimate treatment effects. Absent additional distal outcome or interview data, researchers faced with a “mixed-bag” of treatment effects should suspend judgment on the question of efficacy until the trial can be replicated.

Limitations and Additional Possibilities

This paper used a “wide” data format with latent growth models and maximum likelihood estimation to investigate MNAR suspicions. MNAR models can also be estimated using a “long” data format and longitudinal HLM (also called longitudinal multilevel models or longitudinal mixed linear models) specifications that have been shown to possibly provide a modicum of protection against model misspecification and parameter estimate bias (Gottfredson et al., 2017). Further, although ML estimation was the focus of this tutorial, research has shown that Bayesian estimation (e.g., Asparouhov & Muthén, 2021; Du et al., 2021; Lüdtke et al., 2020) and additional multiple imputation techniques (Carreras et al., 2021; Hsu et al., 2020) can also be used for MNAR investigations. In summary, several options are available to researchers, and the goal of this paper is to contribute practical and hands-on application materials to the published literature to enable trialists to investigate suspected MNAR data in a longitudinal trial and draw defensible treatment conclusions with confidence.

Authors' note

The authors extend their gratitude to James Murphy, Ashley Dennhardt, and Katie Witkiewitz for allowing us to use their trial data for the illustrative examples, and to Stephen Becker, Robert Gibler, Kim Klages, Constance Mara, and Maya Marzouk for helpful observations and comments.

References

- Albert, P. S., Follmann, D. A., Wang, S. A., & Suh, E. B. (2002). A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics*, *58*, 631–642. doi: [10.1111/j.0006-341X.2002.00631.x](https://doi.org/10.1111/j.0006-341X.2002.00631.x).
- Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 1–14. doi: [10.1080/10705511.2020.1764360](https://doi.org/10.1080/10705511.2020.1764360).



- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338–363. doi: [10.1037/1082-989X.8.3.338](https://doi.org/10.1037/1082-989X.8.3.338).
- Bell, M. L., Fiero, M., Horton, N. J., & Hsu, C. H. (2014). Handling missing data in trials; a review of the top medical journals. *BMC Medical Research Methodology*, 14, 1–8. doi: [10.1186/1471-2288-14-118](https://doi.org/10.1186/1471-2288-14-118).
- Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., Mosconi, P., Osoba, D., Simes, J., & Hurny, C. (1998). Missing quality of life data in cancer clinical trials: Serious problems and challenges. *Statistics in Medicine*, 17, 517–532. doi: [10.1002/\(SICI\)1097-0258\(19980315/15\)17:5/7<517::AID-SIM799>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0258(19980315/15)17:5/7<517::AID-SIM799>3.0.CO;2-S).
- Beunckens, C., Molenberghs, G., Verbeke, G., & Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, 64, 96–105. doi: [10.1111/j.1541-0420.2007.00837.x](https://doi.org/10.1111/j.1541-0420.2007.00837.x).
- Carpenter, J. R., & Kenward, M. G. (2007, June). *Missing data in randomised controlled trials—a practical guide*. https://missingdata.lshtm.ac.uk/downloads/rm04_jh17_mk.pdf
- Carreras, G., Miccinesi, G., Wilcock, A., Preston, N., Nieboer, D., Deliens, L., Groenvold, M., Lunder, U., van der Heide, A., & Baccini, M. (2021). Missing not at random in end of life care studies: Multiple imputation and sensitivity analysis on data from the action study. *BMC Medical Research Methodology*, 21, 1–12. doi: [10.1186/s12874-020-01180-y](https://doi.org/10.1186/s12874-020-01180-y).
- 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. doi: [10.1037/1082-989X.6.4](https://doi.org/10.1037/1082-989X.6.4).
- Coppock, A., Gerber, A. S., Green, D. P., & Kern, H. L. (2017). Combining double sampling and bounds to address nonignorable missing outcomes in randomized experiments. *Political Analysis*, 25, 188–206. doi: [10.1017/pan.2016.6](https://doi.org/10.1017/pan.2016.6).
- Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., & Kenward, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52, 111–125. doi: [10.1002/bimj.200800235](https://doi.org/10.1002/bimj.200800235).
- Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., & Kenward, M. G. (2011). Generalized shared-parameter models and missingness at random. *Statistical Modelling*, 11, 279–310. doi: [10.1177/1471082X1001100401](https://doi.org/10.1177/1471082X1001100401).
- Cro, S., Morris, T. P., Kenward, M. G., & Carpenter, J. R. (2020). Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Statistics in Medicine*, 39, 2815–2842. doi: [10.1002/sim.8569](https://doi.org/10.1002/sim.8569).
- Dantan, E., Proust-Lima, C., Letenneur, L., & Jacqmin-Gadda, H. (2008). Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics*, 4, 1–26. doi: [10.2202/1557-4679.1088](https://doi.org/10.2202/1557-4679.1088).
- Demirtas, H., & Schafer, J. L. (2003). On the performance of random coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, 22, 2553–2575. doi: [10.1002/sim.1475](https://doi.org/10.1002/sim.1475).
- Dewberry, L. C., Jalivand, A., Gupta, R., Jenkins, T. M., Beamish, A., Courcoulas, A., Helmuth, M., Brandt, M. L., Harmon, C. M., & Chen, M. (2020). Weight loss and health status 5 years after adjustable gastric banding in adolescents. *Obesity Surgery*, 30, 2388–2394. doi: [10.1007/s11695-020-04504-5](https://doi.org/10.1007/s11695-020-04504-5).
- Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43, 49–73. doi: [10.2307/2986113](https://doi.org/10.2307/2986113).
- DiLillo, D., Peugh, J., Walsh, K., Panuzio, J., Trask, E., & Evans, S. (2009). *Child maltreatment history among newlywed couples: A longitudinal study of marital outcomes and mediating pathways*. doi: [10.1037/a0015708](https://doi.org/10.1037/a0015708).
- Du, H., Enders, C., Keller, B. T., Bradbury, T. N., & Karney, B. R. (2021). A Bayesian latent variable selection model for nonignorable missingness. *Multivariate Behavioral Research*, 57, 478–512. doi: [10.1080/00273171.2021.1874259](https://doi.org/10.1080/00273171.2021.1874259).
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Enders, C. K. (2011a). Analyzing longitudinal data with missing values. *Rehabilitation Psychology*, 56, 267–288. doi: [10.1037/a0025579](https://doi.org/10.1037/a0025579).
- Enders, C. K. (2011b). Missing not at random models for latent growth curve analyses. *Psychological Methods*, 16, 1–16. doi: [10.1037/a0022640](https://doi.org/10.1037/a0022640).
- Enders, C. K. (2022). *Applied missing data analysis (2nd ed.)*. Guilford Press.
- Fitzmaurice, G. M., Laird, N. M., & Shneyer, L. (2001). An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable dropouts. *Statistics in Medicine*, 20, 1009–1021. doi: [10.1002/sim.718](https://doi.org/10.1002/sim.718).
- Gachau, S., Quartagno, M., Njagi, E. N., Owuor, N., English, M., & Ayieko, P. (2020). Handling missing data in modelling quality of clinician-prescribed routine care: Sensitivity analysis of departure from missing at random assumption. *Statistical Methods in Medical Research*, 29, 3076–3092. doi: [10.1177/0962280220918279](https://doi.org/10.1177/0962280220918279).



- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. Norton.
- Gomer, B., & Yuan, K.-H. (2021). Subtypes of the missing not at random missing data mechanism. *Psychological Methods, 26*, 559–598. doi: [10.1037/met0000377](https://doi.org/10.1037/met0000377).
- Gomila, R., & Clark, C. S. (2022). Missing data in experiments: Challenges and solutions. *Psychological Methods, 27*, 143–155. doi: [10.1037/met0000361](https://doi.org/10.1037/met0000361).
- Gottfredson, N. C., Bauer, D. J., & Baldwin, S. A. (2014). Modeling change in the presence of nonrandomly missing data: Evaluating a shared parameter mixture model. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 196–209. doi: [10.1080/10705511.2014.882666](https://doi.org/10.1080/10705511.2014.882666).
- 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*, 12–19. doi: [10.1007/s11121-016-0735-3](https://doi.org/10.1007/s11121-016-0735-3).
- Graham, J. W. (2003). Adding missing-data-relevant variables to fiml-based structural equation models. *Structural Equation Modeling, 10*, 80–100. doi: [10.1207/S15328007SEM1001_4](https://doi.org/10.1207/S15328007SEM1001_4).
- Graham, J. W. (2012). *Missing data: Analysis and design*. Springer.
- 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. doi: [10.1007/s11121-007-0070-9](https://doi.org/10.1007/s11121-007-0070-9).
- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*, 475–492. <https://www.nber.org/books/aesm76-4>
- Heckman, J. J. (1979). *Statistical models for discrete panel data*. Department of Economics; Graduate School of Business.
- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*, 64–78. doi: [1082-989X/97/S3.00](https://doi.org/1082-989X/97/S3.00).
- Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Wiley.
- Heitjan, D. F., & Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics, 19*, 2244–2253. doi: [10.1214/aos/1176348396](https://doi.org/10.1214/aos/1176348396).
- Hogan, J. W., & Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine, 16*, 239–257.
- Hsu, C. H., He, Y., Hu, C., & Zhou, W. (2020). A multiple imputation-based sensitivity analysis approach for data subject to missing not at random. *Statistics in Medicine, 39*, 3756–3771. doi: [10.1002/sim.8691](https://doi.org/10.1002/sim.8691).
- Iddrisu, A. K., & Gumedze, F. (2020). Sensitivity analysis for the generalized shared-parameter model framework. *Journal of Biopharmaceutical Statistics, 30*, 197–215. doi: [10.1080/10543406.2019.1632875](https://doi.org/10.1080/10543406.2019.1632875).
- Jung, H., Schafer, J. L., & Seo, B. (2011). A latent class selection model for nonignorably missing data. *Computational Statistics & Data Analysis, 55*, 802–812. doi: [10.1016/j.csda.2010.07.002](https://doi.org/10.1016/j.csda.2010.07.002).
- Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Statistics in Medicine, 17*, 2723–2732. doi: [10.1002/\(SICI\)1097-0258\(19981215\)17:23<2723::AID-SIM38>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1097-0258(19981215)17:23<2723::AID-SIM38>3.0.CO;2-5).
- Kenward, M. G. (2015). Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. *Clinical Investigation, 5*, 311–320. doi: [10.4155/CLL.14.132](https://doi.org/10.4155/CLL.14.132).
- Kuha, J. (2004). Aic and bic: Comparisons of assumptions and performance. *Sociological Methods & Research, 33*, 188–229. doi: [10.1177/0049124103262065](https://doi.org/10.1177/0049124103262065).
- Leacy, F. P., Floyd, S., Yates, T. A., & White, I. R. (2017). Analyses of sensitivity to the missing-at-random assumption using multiple imputation with delta adjustment: Application to a tuberculosis/hiv prevalence survey with incomplete hiv-status data. *American Journal of Epidemiology, 185*, 304–315. doi: [10.1093/aje/kww107](https://doi.org/10.1093/aje/kww107).
- Lin, H., McCulloch, C., & Rosenheck, R. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics, 60*, 295–305. doi: [10.1111/j.0006-341X.2004.00173.x](https://doi.org/10.1111/j.0006-341X.2004.00173.x).
- Little, R., & Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics, 52*, 1324–1333. doi: [10.2307/2532847](https://doi.org/10.2307/2532847).
- Little, R. J., & Rubin, D. B. (2002). Missing data in experiments. In *Statistical analysis with missing data* (pp. 24–40). John Wiley & Sons, Ltd. doi: [10.1002/9781119013563.ch2](https://doi.org/10.1002/9781119013563.ch2).
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association, 88*, 125–134. doi: [10.1080/01621459.1993.10594302](https://doi.org/10.1080/01621459.1993.10594302).
- Little, R. J. A. (1994). A class of pattern-mixture models for normal missing data. *Biometrika, 81*, 471–483. doi: [10.2307/2337120](https://doi.org/10.2307/2337120).
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated measures studies. *Journal of the American Statistical Association, 90*, 1112–1121. doi: [10.2307/2291350](https://doi.org/10.2307/2291350).



- Little, R. J. A. (2009). Selection and pattern-mixture models. In M. D. Fitzmaurice, G. Verbeke, & G. Molenberghs (Eds.), *G* (pp. 409–431). Chapman & Hall/CRC.
- Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation. *Psychological Methods, 25*, 157–181. doi: [10.1037/met0000233](https://doi.org/10.1037/met0000233).
- Mallinckrodt, C., Roger, J., Chuang-Stein, C., Molenberghs, G., O’Kelly, M., Ratitch, B., Janssens, M., & Bunouf, P. (2014). Recent developments in the prevention and treatment of missing data. *Therapeutic Innovation & Regulatory Science, 48*, 68–80. doi: [10.1177/2168479013501310](https://doi.org/10.1177/2168479013501310).
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156–166. doi: [10.1037/0033-2909.105.1.156](https://doi.org/10.1037/0033-2909.105.1.156).
- Molenberghs, G., Beunckens, C., Sotto, C., & Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70*, 371–388. doi: [1369--7412/08/70371](https://doi.org/10.1111/j.1471082X0100100402).
- Molenberghs, G., & Kenward, M. G. (2007). *Missing data in clinical studies*. Wiley.
- Molenberghs, G., Kenward, M. G., & Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom dropout. *Biometrika, 84*, 33–44. doi: [10.1093/biomet/84.1.33](https://doi.org/10.1093/biomet/84.1.33).
- Molenberghs, G., & Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling, 1*, 235–269. doi: [10.1177/1471082X0100100402](https://doi.org/10.1177/1471082X0100100402).
- Murphy, J. G., Dennhardt, A. A., Martens, M. P., Borsari, B., Witkiewitz, K., & Meshesha, L. Z. (2019). A randomized clinical trial evaluating the efficacy of a brief alcohol intervention supplemented with a substance-free activity session or relaxation training. *Journal of Consulting and Clinical Psychology, 87*, 657–669. doi: [10.1037/ccp0000412](https://doi.org/10.1037/ccp0000412).
- 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. doi: [10.1037/a0022634](https://doi.org/10.1037/a0022634).
- Muthén, B., & Brown, H. C. (2009). Estimating drug effects in the presence of placebo response: Causal inference using growth mixture modeling. *Statistics in Medicine, 28*, 3363–3385. doi: [10.1002/sim.3721](https://doi.org/10.1002/sim.3721).
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139–157. doi: [1082-989X/99/](https://doi.org/1082-989X/99/).
- Post, W. J., Buijs, C., Stolk, R. P., de Vries, E. G., & Le Cessie, S. (2010). The analysis of longitudinal quality of life measures with informative drop-out: A pattern mixture approach. *Quality of Life Research, 19*, 137–148. doi: [10.1007/s11136-009-9564-1](https://doi.org/10.1007/s11136-009-9564-1).
- Power, R. A., Muthén, B., Henigsberg, N., Mors, O., Placentino, A., Mendlewicz, J., Maier, W., McGuffin, P., Lewis, C. M., & Uher, R. (2012). Non-random dropout and the relative efficacy of escitalopram and nortriptyline in treating major depressive disorder. *Journal of Psychiatric Research, 46*, 1333–1338. doi: [10.1016/j.jpsychires.2012.06.014](https://doi.org/10.1016/j.jpsychires.2012.06.014).
- Powney, M., Williamson, P., Kirkham, J., & Kolamunnage-Dona, R. (2014). A review of the handling of missing longitudinal outcome data in clinical trials. *Trials, 15*, 1–11. doi: [10.1186/1745-6215-15-237](https://doi.org/10.1186/1745-6215-15-237).
- Ratitch, B., O’Kelly, M., & Tosiello, R. (2013). Missing data in clinical trials: From clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics, 12*, 337–347. doi: [10.1002/pst.1549](https://doi.org/10.1002/pst.1549).
- Raykov, T. (2011). On testability of missing data mechanisms in incomplete data sets. *Structural Equation Modeling: A Multidisciplinary Journal, 18*, 419–429. doi: [10.1080/10705511.2011.582396](https://doi.org/10.1080/10705511.2011.582396).
- Roberts, E. M., & English, P. B. (2016). *Analysis of multiple-variable missing-not-at-random survey data for child lead surveillance using nhanes* (Vol. 35). doi: [10.1002/sim.7067](https://doi.org/10.1002/sim.7067).
- Rose, N., von Davier, M., & Xu, X. (2010). Modeling non-ignorable missing data with item response theory. *Paper presented at the 75th annual meeting of the International Psychometric Society, Athens, GA*.
- Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics, 59*, 829–836. doi: [10.1111/j.0006-341X.2003.00097.x](https://doi.org/10.1111/j.0006-341X.2003.00097.x).
- Roy, J. (2007). Latent class models and their application to missing-data patterns in longitudinal studies. *Statistical Methods in Medical Research, 16*, 441–456. doi: [10.1177/0962280206075311](https://doi.org/10.1177/0962280206075311).
- Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592. doi: [10.1093/biomet/63.3.581](https://doi.org/10.1093/biomet/63.3.581).
- Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association, 72*, 538–543.
- Sharmin, S., Kypri, K., Wadolowski, M., Bruno, R., Khanam, M., Aiken, A., Hutchinson, D., Najman, J. M., Slade, T., McBride, N., & Attia, J. (2019). Parent hazardous drinking and their children’s alcohol use in early and mid-adolescence: Prospective cohort study. *European Journal of Public Health, 29*, 736–740. doi: [10.1093/eurpub/ckz029](https://doi.org/10.1093/eurpub/ckz029).



- Shrive, F. M., Stuart, H., Quan, H., & Ghali, W. A. (2006). Dealing with missing data in a multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology*, 6, 1–10. doi: [10.1186/1471-2288-6-57](https://doi.org/10.1186/1471-2288-6-57).
- Sterba, S. K., & Gottfredson, N. C. (2015). Diagnosing global case influence on mar versus mmar model comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, 22, 294–307. doi: [10.1080/10705511.2014.936082](https://doi.org/10.1080/10705511.2014.936082).
- Sterba, S. K., & Pek, J. (2012). Individual influence on model selection. *Psychological Methods*, 17, 582–599. doi: [10.1037/a0029253](https://doi.org/10.1037/a0029253).
- Thoemmes, F., & Rose, N. (2014). A cautious note on auxiliary variables that can increase bias in missing data problems. *Multivariate Behavioral Research*, 49, 443–459. doi: [10.1080/00273171.2014.931799](https://doi.org/10.1080/00273171.2014.931799).
- Verbeke, G., Lessafre, E., & Spiessens, B. (2001). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, 35, 419–434. doi: [10.1177/009286150103500211](https://doi.org/10.1177/009286150103500211).
- Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer-Verlag.
- Verbeke, G., Molenberghs, G., & Beunckens, C. (2008). Formal and informal model selection with incomplete data. *Statistical Science*, 23, 201–218. doi: [10.1214/07-STS253](https://doi.org/10.1214/07-STS253).
- 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, 175–188. doi: [10.2307/2531905](https://doi.org/10.2307/2531905).
- Yang, M., & Maxwell, S. E. (2014). Treatment effects in randomized longitudinal trials with different types of non-ignorable dropout. *Psychological Methods*, 19, 188–210. doi: [10.1037/a0033804](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, 246–267. doi: [10.1111/bmsp.12043](https://doi.org/10.1111/bmsp.12043).
- Yuan, Y., & Little, R. J. (2009). Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics*, 65, 478–486. doi: [10.1111/j.1541-0420.2008.01102.x](https://doi.org/10.1111/j.1541-0420.2008.01102.x).
- Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, 111, 493–504. doi: [10.1037/pspa0000056](https://doi.org/10.1037/pspa0000056).

Open practices

📄 The *Open Material* badge was earned because supplementary material(s) are available on [the journal's web site](#).

Citation

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

Copyright © 2023, Peugh et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 24/08/2023 ~ Accepted: 27/11/2023