Sensitivity Analysis of Missing Data: A Delta-Based Approach Using Multiple Imputation in Clinical Trials

Abstract

Missing data in clinical trials can significantly impact treatment effect estimates and potentially bias conclusions. This paper provides an expository review of δ-based sensitivity analysis approaches for handling missing continuous outcomes, comparing them with complete case analysis and multiple imputation. The simulation study revealed that complete case analysis underestimated the true treatment effect (0.38 vs 0.50), while δ-based approaches with values of 0, -5, and -10 produced estimates ranging from 0.40 to 0.53. Application to the HELP study (Kaplan and Horton (2019)) showed treatment effects varying from -0.70 to -0.49 across δ values (-10, -5, 0), with widening confidence intervals reflecting increased uncertainty under stronger assumptions. These findings highlight the importance of sensitivity analyses in assessing the robustness of conclusions drawn from clinical trials with missing data.

Introduction

Missing data in clinical trials presents a fundamental challenge that can significantly impact treatment evaluation and patient care decisions. When participants drop out or fail to complete scheduled assessments, we must carefully consider how to handle these missing values to avoid biased conclusions (Buuren (2018)). Consider a clinical trial testing a new depression treatment over 24 months. Some participants complete all assessments, while others stop attending after a few months. How should researchers analyze this incomplete data to draw valid conclusions about the treatment's effectiveness?

This missing data problem extends beyond simply having incomplete information. For instance participants might drop out because the treatment is not working for them; they experienced side effects; they moved to a different city or they felt better and saw no need to continue. Notice each scenario has different meaning for analyzing our treatment effectiveness.

To help researchers deal with these different scenarios, statisticians have developed frameworks for categorizing missing data which was first established by Rubin (1987). The first category, is Missing Completely at Random (MCAR), occurs when the probability of missing values does not depend on any variables in the study. For example, participants missing assessments due to random events, such as transportation issues or clinic scheduling conflicts. The second category, Missing at Random (MAR), occurs when the probability of missing data depends only on information we have observed. For example, participants with more severe baseline symptoms being more likely to miss follow-ups which is related to data we have measured. The most challenging situation, Missing Not at Random (MNAR), occurs when the missingness depends on unobserved information. For example, participants miss assessments because their condition got worse beyond what we last measured.

Early statistical methods mainly focused on MAR scenarios, but this does not always hold. Participants might stop attending because their condition is changing which is information we do not have. Recognizing this limitation, researchers have developed methods that let us test how our conclusions might change under different assumptions about missing data.

One approach is the δ-based method that was developed by Ratitch, O'Kelly, and Tosiello (2013). It allows us to explore what would happen if missing values were systematically different than what we would expect based on the information we have. This method uses maximum likelihood estimation to find optimal parameter values, as we do in regression.

This paper provides an expository review of δ-based approaches compared with complete case analysis and multiple imputation with predictive mean matching. Using the `HELP data` from the `mosaicData` package (Kaplan and Horton (2019)), we demonstrate how this sensitivity analysis methods work in practice.

As Bell et al. (2014) found, only around a third of trials with missing data report sensitivity analyses which shows a huge gap between statistical methods and how they are being used in

practice. The aim is to promote wider adoption of the method among students and clinical researchers.

The remainder of the paper is organized as follows. Section 2 introduces our motivating example using the HELP study (Kaplan and Horton (2019)). Section 3 presents the methodological foundations, including the multiple imputation framework and δ-based approach. Section 4 evaluates these methods through simulation studies comparing complete case analysis, standard multiple imputation under MAR, and δ-based sensitivity analyses. Section 5 demonstrates practical application of these methods to the HELP study (Kaplan and Horton (2019)). Section 6 discusses implications future research directions. Finally section 7 discusses limitations of the paper.

2. Motivating example

The HELP study (Kaplan and Horton (2019)) is a clinical trial that recruited adult inpatients from a detoxification unit with the goal of linking patients without primary care physicians to primary medical care. Eligible subjects were adults who spoke Spanish or English, reported alcohol, heroin, or cocaine as their first or second drug of choice, resided near the primary care clinic to which they would be referred, or were homeless.

Patients with established primary care relationships they planned to continue, significant dementia, specific plans to leave the Boston area that would prevent research participation, failure to provide contact information for tracking purposes, or pregnancy were excluded. Subjects were interviewed at baseline during their detoxification stay, and follow-up interviews were undertaken every 6 months for 2 years.(Kaplan and Horton (2019))

The full dataset contains 788 variables containing a variety of continuous, count, discrete, and survival time predictors and outcomes variables. For simplicity, we selected four variables: Center for Epidemiologic Studies Depression Scale `CES-D` which is score measure of depressive symptoms and high scores are worse, Gender `A1` (1=Male, 2=Female), Mental component score `MCS`, and treatment group assignment `GROUP` (0=Control, 1=Clinic). Subjects were interviewed at baseline during their detoxification stay, and follow-up interviews were undertaken every 6 months for 2 years. Gender was measured at baseline only and carried forward for each participant. We replicate the baseline gender value across all time points for each participant.

Our primary analysis aimed to assess treatment impact on depression outcomes at 24 months. Comparing at baseline participants who did and did not complete had significant differences. Those missing 24-month outcomes had lower baseline depression (`CES-D` 7.0 vs 32.9) and higher mental health scores (`MCS` 43.2 vs 37.3). The dropout rates for the `CES_D` are 0% missing data at 6 months, 1.38% at 12 months, 0.39% at 18 months, and 1.8% at 24 months. As shown in **?@fig-1**, the missing data percentage increases over time, peaking at 24 months, though the missingness remains relatively low overall. The number of participants also decreased over time, with 470 participants assessed at baseline, 254 at the 6-month follow-up, 217 at 12 months, and 277 at 24 months.

3

These non-response rates raise questions about the representativeness of participants who stayed in the study vs who did not. Those with better mental health scores may have been more likely to drop out due to recovery, while others may have dropped out due to treatment failure. This means the people who stayed in the study might be different from those who left, and these differences could affect how we interpret our results. These relationships between dropout and participants' outcomes, including unobserved changes we could not measure, suggest our missing data might follow a Missing Not at Random (MNAR) mechanism.

3. Methods

3.1 The Multiple Imputation Framework

Multiple imputation (MI) is one of the most widely used approaches for handling missing data in clinical trials. It creates multiple complete datasets by filling in missing values with plausible estimates. This helps us maintain statistical power while accounting for the uncertainty in our missing value predictions.

For continuous outcomes, the regression model:

$Y = X\beta + \epsilon$

where Y is outcome, X includes covariates, β contains the coefficients, and ϵ represents random error.

The MI process has three main steps:

1. Imputation Phase: In this step, we create multiple complete datasets (usually 20-50) to fill in missing values. Similar to how we use predictive models in regression, MI uses relationships between variables to make informed predictions of missing values. One approach for imputation is predictive mean matching (PMM), which fills in missing values by borrowing real observed values from similar participants. For example, in our HELP study (Kaplan and Horton (2019)), if a participant missed their 24 month depression score, we look at other participants with similar baseline scores and treatment responses who did complete the study. Based on these patterns, we make several reasonable predictions about what the missing score might have been. We do this 50 times to create 50 slightly different complete datasets, each representing a possible version of what our complete data could have looked like.

2. Analysis Phase: We then analyze each of these 50 datasets using our planned analysis method. In HELP study (Kaplan and Horton (2019)), this means running a regression of depression scores on treatment group and other important factors like gender, group and mental health scores. This gives us 50 slightly different estimates of our treatment effect. This is similar to bootstrapping, where we analyze multiple versions of our data to understand uncertainty.

3. Pooling Phase: Finally, we combine results across all analyses using Rubin's rules. (Rubin (1987)) This means when we get 50 different treatment effect estimates (one from each imputed dataset), we then take their average to get our final estimate and look at how much these estimates vary across our 50 datasets. The mathematical formula is:

$$\bar{Q} = \frac{1}{M} \sum \hat{Q}_m$$

Where:

- $\bar{Q}$ is our final combined estimate.
- M is the number of imputations (in our case, 50).
- $\hat{Q}_m$ is the estimate from each imputed dataset.

The uncertainty in this estimate comes from two sources:

1. Within-imputation variance (W): This is the uncertainty we would have even with complete data. For example, in HELP study (Kaplan and Horton (2019)) some participants' scores might improve a lot while others improve just a little as everyone responds to treatment in a different way. Hence this captures natural variation.

2. Between-imputation variance (B): This additional uncertainty comes from having missing data. It tells us how much our treatment effect estimates vary across our 50 imputed datasets. If these estimates vary a lot, it suggests the missing data is creating more uncertainty in our conclusions.

The total variance combines both:

$$T = W + \left(1 + \frac{1}{M}\right) B$$

However, in clinical trials, there is often concern that data are Missing Not at Random (MNAR). Standard MI assuming MAR will not capture this mechanism. This is where δ-based approaches become valuable which allow us to explore departures from MAR assumptions (Ratitch, O'Kelly, and Tosiello (2013)).

3.2 The δ-based Approach

After imputing values under MAR, we add a fixed adjustment δ to assess how results might change if missing values were systematically different from what we would predict based on observed data (Carpenter, Roger, and Kenward (2013)).

The magnitude of δ represents how much worse we think the unobserved outcomes might be. For example, if we think patients who drop out tend to have worse outcomes, we might add a positive δ to the imputed values.

The δ-based approach has three main steps:

1. Imputation phase: We start by creating multiple complete datasets using the same imputation approach as standard multiple imputation under MAR assumptions.

2. Modification phase: We then modify each imputed dataset by adding our chosen δ value. For instance, in HELP study (Kaplan and Horton (2019)), we explored δ values (0, -5, and -10) to see what would happen if dropouts had different depression scores than predicted. For example δ = -5 means we are assuming participants who dropped out had depression scores 5 points better than what we would predict from their observed data.

3. Analysis and pooling phase: Finally, we analyze each modified dataset using our planned analysis model and combine all results using Rubin's rules.

4. Simulation Study

The objective of the simulation study was to evaluate the performance of δ-based multiple imputation under different scenarios. We generated $n = 1000$ independent datasets with MAR and MNAR to compare our approach against complete case analysis, multiple imputation under MAR.

4.1 Data-generating process

Let us generate two independent variables and a continuous outcome for $n = 1000$ observations: a binary treatment indicator X ~ Bernoulli(0.5) that randomly assigns participants to treatment (1) or control (0) with equal probability and a continuous covariate Z ~ N(0,1). These variables are associated with Y through $(\beta_0, \beta_1, \beta_2)$, a vector of coefficients in the linear model:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon$$

where ϵ ~ N(0,1). The parameter values were fixed at $\beta_0 = -1$, $\beta_1 = 0.5$, and $\beta_2 = 1$ in the outcome equation.

To explore different missing data scenarios, we created missing values according to Missing at Random (MAR) and Missing Not at Random (MNAR) mechanisms. For MAR, whether data is missing $(R = 1)$ depends on both treatment status (X) and baseline measures (Z) through the formula $P(R = 1|X, Z) = 0.2 * (0.5X + 0.5|Z|)$, creating about 20% missing data overall.

This means participants in the treatment group and those with more extreme baseline values are more likely to have missing data.

For MNAR, missing data depends on the outcome itself through $P(R = 1|Y) = 0.3 * I(Y > Y_{75})$, creating about 30% missingness overall. This means participants with outcomes above the 75th percentile are more likely to have missing data. This can be a situations where participants might drop out because they are feeling better (or worse).

The complete simulation procedure was implemented in R (version 4.2.2) using the mice package (vanBuuren and Groothuis-Oudshoorn (2011)) for multiple imputation.

4.2 Analysis Methods

The analysis outcome model was a linear regression including treatment (X) and covariate (Z) as predictors:

$$Y \sim X + Z$$

The incomplete data were then analyzed using the following approaches:

- The complete case analysis (CCA) method consisted of estimating the parameters of the outcome model based only on complete observation.

- The standard multiple imputation under MAR used predictive mean matching (PMM) with X and Z in the imputation model, with $m = 50$ imputations combined via Rubin's rules.

- The δ-based sensitivity analyses applied systematic adjustments ($\delta = 0$, -5, -10) to imputed values to explore departures from MAR.

For each analytical setting, we assessed performance by comparing estimated treatment effects to true values and coverage of confidence intervals.

4.3 Results

The results of our simulation study, which compared multiple approaches for handling missing data, are shown in Table 1 and visualized in Figure 1. Complete case analysis (CCA) yielded a treatment effect estimate of 0.38 with confidence interval ranged from 0.25 to 0.51, underestimating the true treatment effect of 0.5. The primary analysis using multiple imputation under MAR produced a slightly improved estimate of 0.40 with confidence interval ranged from 0.28 to 0.53.

For sensitivity analyses, at $\delta = 0$, assuming missing values would have been similar to what we would predict, the treatment effect stayed stable at 0.40. As we decreased δ to -5, the

estimate increased to 0.47 which is closest to the true value of 0.5. At δ = -10, it went up to 0.53 slightly overestimating the true effect. As illustrated in the top left of Figure 1, the confidence intervals got wider with each step from (0.28, 0.52) with δ = 0, to (0.26, 0.68) with δ = -5, and (0.18, 0.89) with δ = -10, shown by the expanding shaded area around the treatment effect estimate line.

Looking at the covariate effect, where the true value was 1, we observed overestimation as δ decreased. The estimates ranged from 1.02 in complete case analysis to 1.66 with δ = -10. The top right of Figure 1 clearly shows this strong linear trend of increasing covariate estimates as δ becomes more negative, with the confidence bands widening moderately. The intercept showed similar sensitivity, moving from -0.87 under complete case analysis to -1.81 under δ = -10. The bottom left of Figure 1 illustrates this negative trend in the intercept estimates, with expanding confidence bands at more extreme δ values. This pattern suggests that as we make stronger assumptions about how missing data differs from observed data, we become less certain about our conclusions.

Table 1: Comparison of parameter estimates and confidence intervals for treatment (X), co-variate (Z), and intercept CCA, MAR, and δ-based sensitivity analyses

| Analysis	Parameter	Est.	95% CI
CCA	(Intercept)	-0.87	(-0.96, -0.78)
CCA	X	0.38	(0.25, 0.51)
CCA	Z	1.02	(0.95, 1.08)
Primary (MI MAR)	(Intercept)	-0.94	(-1.03, -0.86)
Primary (MI MAR)	X	0.40	(0.28, 0.53)
Primary (MI MAR)	Z	1.07	(1.00, 1.14)
Sensitivity (MI $\delta = 0$)	(Intercept)	-0.88	(-0.96, -0.79)
Sensitivity (MI $\delta = 0$)	X	0.40	(0.28, 0.52)
Sensitivity (MI $\delta = 0$)	Z	1.00	(0.93, 1.06)
Sensitivity (MI $\delta = -5$)	(Intercept)	-1.34	(-1.49, -1.20)
Sensitivity (MI $\delta = -5$)	X	0.47	(0.26, 0.68)
Sensitivity (MI $\delta = -5$)	Z	1.33	(1.22, 1.43)
Sensitivity (MI $\delta = -10$)	(Intercept)	-1.81	(-2.06, -1.56)
Sensitivity (MI $\delta = -10$)	X	0.53	(0.18, 0.89)
Sensitivity (MI $\delta = -10$)	Z	1.66	(1.48, 1.84)
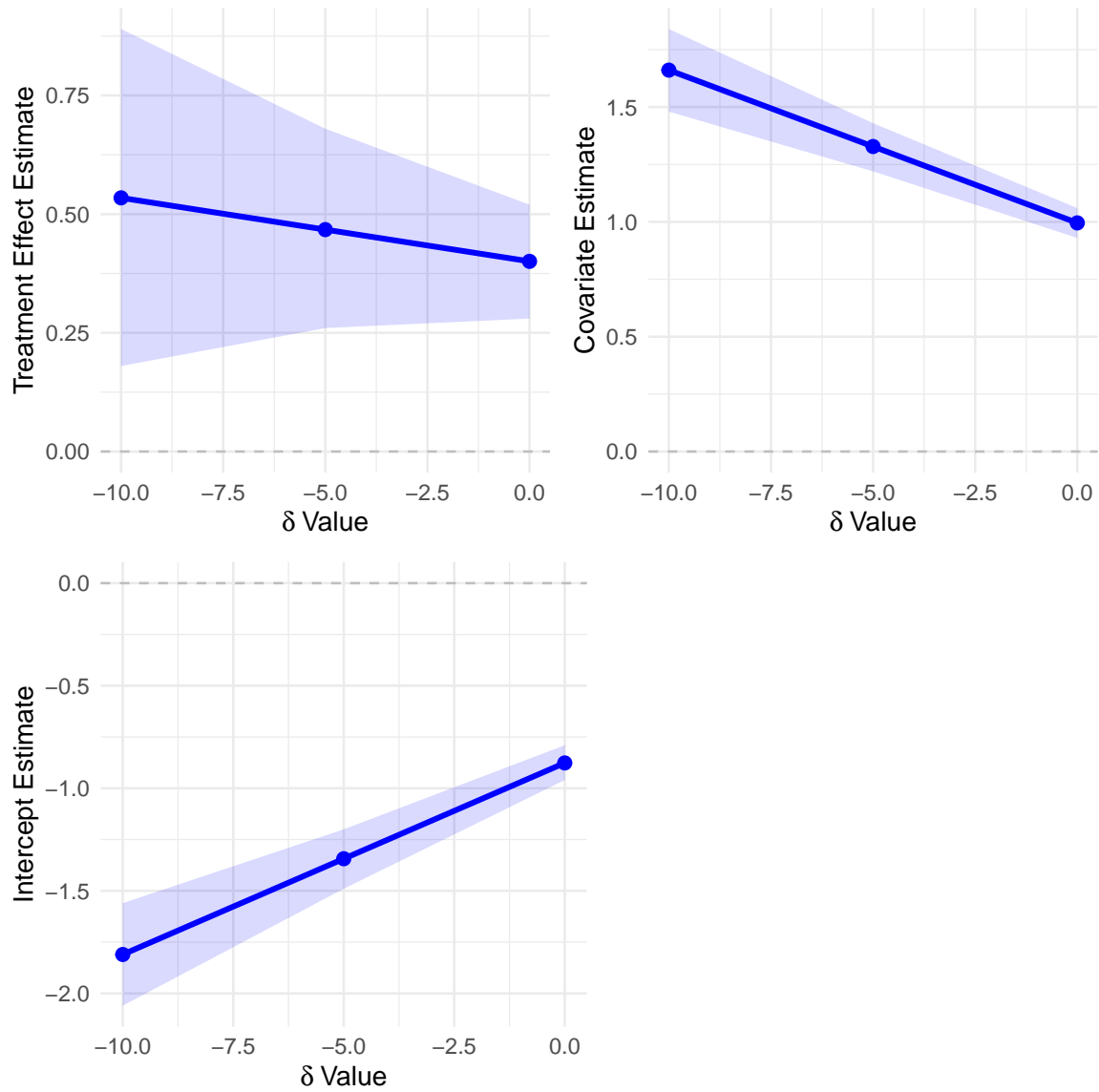
Figure 1: The plots illustrate the sensitivity of treatment effect, covariate, and intercept estimates at delta values of -10, -5, and 0. Shaded areas represent 95% confidence intervals.

5. Application to HELP data

5.1 Analysis Methods

5.1.1 Complete Case Analysis

As an initial approach, we conducted a complete case analysis (CCA) that included only participants with observed 24-month outcomes. The model is:

$$CESD_{24} = \beta_0 + \beta_1 GROUP + \beta_2 A1 + \epsilon$$

where β_1 tells us about the treatment effect on depression scores, β_2 tells us about the relationship between gender and depression, and ϵ represents random error. While this approach seems straightforward, it could bias our results if participants who dropped out differed systematically from those who completed the study.

5.1.2 Standard Multiple Imputation

We implemented multiple imputation using predictive mean matching (`PMM`) through the mice package in R (vanBuuren and Groothuis-Oudshoorn (2011)). We created $m = 50$ complete datasets by imputing missing values using `GROUP`, `MCS` and `A1` as predictors. For each imputed dataset, we fit our primary analysis model examining treatment effects on 24 month depression scores.

5.1.3 δ-based sensitivity analysis

Building on our MAR imputation, we conducted sensitivity analyses using $\delta = 0, -5, -10$ to the imputed values. We chose negative δ values assuming participants who dropped out might have had better mental health outcomes than predicted under MAR. This could be true if some participants left the study because they were feeling better and no longer felt they needed treatment.

5.2 Results

Results are displayed in Table 2. For complete case analysis, we found a treatment effect of -1.73. While this suggests the treatment might reduce depression scores, the wide confidence interval (-5.11 to 1.64) tells us we can not be certain about this effect. The gender effect was at 4.66 (0.83 to 8.49). In multiple imputation under MAR the treatment effect became smaller (-0.70) and the confidence interval narrowed (-2.72 to 1.32). The gender effect remained stable but became more precise at 4.37 (1.92 to 6.83).

In sensitivity analyses when we set $\delta = 0$, we got exactly the same results as our MAR analysis, which served as a good check that our sensitivity analysis was working correctly. When we

assumed better outcomes for people who dropped out (δ = -5), the treatment effect barely changed to -0.74, but our uncertainty increased as the confidence interval ranged from -2.95 to 1.47. For δ = -10, the treatment effect became smaller at -0.49, with confidence interval ranged from -2.72 to 1.74.

What is noteworthy is how stable the gender effect remained across all these analyses (ranging from 4.37 to 4.59), suggesting this finding is robust to different missing data assumptions. The intercept showed dramatic changes, dropping from 19.28 under MAR to 9.45 with δ = -10 which indicates that our assumptions about missing data substantially affect our baseline estimates. While the treatment might help reduce depression which shown by the negative estimates, we can not be statistically significant about this effect. Because the widening confidence intervals in our sensitivity analyses reflect the growing uncertainty when we make different assumptions about the missing data.

Table 2: Estimated 24 month depression (CES-D) scores in the HELP study comparing complete case, multiple imputation, and sensitivity analyses

Analysis	Parameter	Est.	95% CI
CCA	(Intercept)	15.42	NA
CCA	GROUP	-1.73	NA
CCA	A1	4.66	NA
Primary (MI MAR)	(Intercept)	19.28	(15.68, 22.87)
Primary (MI MAR)	GROUP	-0.70	(-2.72, 1.32)
Primary (MI MAR)	A1	4.37	(1.92, 6.83)
Sensitivity (MI δ = 0)	(Intercept)	19.28	(15.68, 22.87)
Sensitivity (MI δ = 0)	GROUP	-0.70	(-2.72, 1.32)
Sensitivity (MI δ = 0)	A1	4.37	(1.92, 6.83)
Sensitivity (MI δ = -5)	(Intercept)	14.59	(10.73, 18.44)
Sensitivity (MI δ = -5)	GROUP	-0.74	(-2.95, 1.47)
Sensitivity (MI δ = -5)	A1	4.49	(1.89, 7.08)
Sensitivity (MI δ = -10)	(Intercept)	9.45	(5.09, 13.81)
Sensitivity (MI δ = -10)	GROUP	-0.49	(-2.72, 1.74)
Sensitivity (MI δ = -10)	A1	4.59	(2.14, 7.04)

6 DISCUSSION

In this expository review paper, we have described and illustrated how sensitivity analysis can be conducted to explore departures from an MAR assumption for unobserved continuous outcome data using δ-based approach. The δ-based approach extends standard MI by allowing us to explore systematic departures from MAR (Ratitch, O'Kelly, and Tosiello (2013)).

Our decision of using different MI approaches depends on the problem at hand. It is important that researchers employ only methods which make plausible assumptions relevant to the clinical

setting and estimand of interest. In some circumstances, the assumptions made by the δ-based MI procedures may not be suitable. We advise researchers to consider carefully the assumptions behind the MI analyses to ensure each analysis undertaken is suitable to the context. We showed how δ-based MI can be implemented using the mice package (vanBuuren and Groothuis-Oudshoorn (2011)) in R. One is not confined to using R and can use MI packages within other software to complete the analysis.

In the presence of missing data, when we do sensitivity analysis, we need to be sure we are not inappropriately injecting or removing information within the analysis regardless of the specific method utilized. As discussed within, it has been shown that the δ method of MI with a fixed δ adjustment preserves the information loss observed under MAR. This provides relevant, accessible, justified inference in the context of missing data sensitivity analysis. Researchers can be confident when utilizing these approaches that they are not unnecessarily losing or gaining any information beyond that observed under MAR. (Bell et al. (2014))

We have focused throughout on the analysis of continuous outcomes. However, implementations of δ-based MI procedures do exist for other types of outcome variables. Through comprehensive analysis, we demonstrated the practical application and value of δ-based sensitivity analyses, complete case analysis, and multiple imputation under MAR assumptions.

From both the simulation study and HELP study (Kaplan and Horton (2019)), we observed the limitations of complete case analysis in the presence of missing data, as it consistently underestimated the true treatment effect. This demonstrates how excluding incomplete cases can lead to biased conclusions. In the the simulation study multiple imputation under MAR assumptions improved the estimates comapred to CCA. As we strengthened our assumptions of departure from MAR by decreasing the δ values, the estimated values moved further away from the true value and the confidence intervals became wider. This emphasizes the importance of making wise decisions when selecting δ values and consulting a clinician to ensure that the chosen values are clinically plausible.

While our analysis demonstrated the value of δ-based approaches, future work could explore the integration of Reference-Based Multiple Imputation (MI) methods, which offer an alternative framework for handling missing data under nonignorable dropout mechanisms. Reference-based MI assumes that participants' post-dropout data follow the behavior of a predefined reference group. This includes Copy-Reference (CR), where participants who discontinue treatment follow patterns similar to the control group, and Jump-to-Reference (J2R), which assumes immediate loss of treatment benefits after dropout (Tang (2018)).

Another future direction is combining δ-based and reference-based methods. For example, some participants could follow control patterns while others have fixed δ adjustments based on their reason for dropout. As Tang (2018) suggests, this hybrid approach would allow for different statistical behaviors of outcomes after dropout, potentially better reflecting the complex reality of clinical trials.

7 Limitations

While this paper demonstrates the use of δ-based sensitivity analysis for handling missing data, it is not without limitations. The δ-based approach assumes a constant δ adjustment across all missing values, which may oversimplify the diverse reasons for dropout or nonresponse in clinical trials. The fixed δ values introduces subjectivity, as the choice of δ is often guided by expert judgment or hypothetical scenarios rather than empirical evidence. Finally, the simulation scenarios and application to the HELP dataset may not generalize to all clinical trial contexts. The findings presented here are contingent on the specific data-generating processes and missingness mechanisms assumed. Further research could evaluate the robustness of δ-based MI under alternative missing data patterns or more complex real-world datasets. Despite these limitations, δ-based sensitivity analysis remains a valuable tool for exploring the robustness of clinical trial findings to departures from MAR assumptions.

References

Bell, Michelle L., Michael Fiero, Nicholas J. Horton, and Chih-Hung Hsu. 2014. "Handling Missing Data in RCTs: A Review of the Top Medical Journals." *BMC Medical Research Methodology* 14 (1): 118. https://doi.org/10.1186/1471-2288-14-118.

Buuren, Stef van. 2018. *Flexible Imputation of Missing Data.* 2nd ed. Boca Raton, FL: Chapman; Hall/CRC. https://stefvanbuuren.name/fimd/.

Carpenter, James R., James H. Roger, and Michael G. Kenward. 2013. "Analysis of Longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible and Reproducible Analysis." *Journal of Biopharmaceutical Statistics* 23 (6): 1352–71. https://doi.org/10.1080/10543406.2013.834911.

Kaplan, Daniel T., and Nicholas J. Horton. 2019. "mosaicData: Project MOSAIC Data Sets." *R Package Version 0.20.0.* https://CRAN.R-project.org/package=mosaicData.

Ratitch, Boris, Michael O'Kelly, and Robert Tosiello. 2013. "Missing Data in Clinical Trials: From Clinical Assumptions to Statistical Analysis Using Pattern Mixture Models." *Pharmaceutical Statistics* 12 (6): 337–47. https://doi.org/10.1002/pst.1549.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons. https://doi.org/10.1002/9780470316696.

Tang, Yuanyuan. 2018. "On the Multiple Imputation Variance Estimator for Control-based and Delta-adjusted Pattern Mixture Models." *Biometrics* 74 (4): 1379–87. https://doi.org/10.1111/biom.12889.

vanBuuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "MICE: Multivariate Imputation by Chained Equations in r." *Journal of Statistical Software* 45 (3): 1–67. https://doi.org/10.18637/jss.v045.i03.

Acknowledgements