



Published in final edited form as:

Alcohol Clin Exp Res. 2013 December ; 37(12): 2152–2160. doi:10.1111/acer.12205.

Missing Data in Alcohol Clinical Trials: A Comparison of Methods

Kevin A. Hallgren, MS and Katie Witkiewitz, PhD

Department of Psychology, University of New Mexico

Abstract

Background—The rate of participant attrition in alcohol clinical trials is often substantial and can cause significant issues with regard to the handling of missing data in statistical analyses of treatment effects. It is common for researchers to assume that missing data is indicative of participant relapse and under that assumption many researchers have relied on setting all missing values to the worst case scenario for the outcome (e.g., missing=heavy drinking). This sort of single imputation method has been criticized for producing biased results in other areas of clinical research, but has not been evaluated within the context of alcohol clinical trials and many alcohol researchers continue to use the missing=heavy drinking assumption.

Methods—Data from the COMBINE study, a multisite randomized clinical trial, were used to generate simulated situations of missing data under a variety of conditions and assumptions. We manipulated the sample size ($n = 200$, $n = 500$, and $n = 1000$) and dropout rate (5%, 10%, 25%, 30%) under three missing data assumptions (missing completely at random, missing at random, missing not at random). We then examined the association between receiving naltrexone and heavy drinking during the first 10 weeks following treatment using five methods for treating missing data (complete case analysis, last observation carried forward, missing=heavy drinking, multiple imputation, and full information maximum likelihood).

Results—Complete case analysis, last observation carried forward, and missing=heavy drinking produced the most biased naltrexone effect estimates and standard errors under conditions that are likely to exist in randomized clinical trials. Multiple imputation and maximum likelihood produced the least biased naltrexone effect estimates and standard errors.

Conclusions—Assuming that missing=heavy drinking produces biased results of the treatment effect and should not be used to evaluate treatment effects in alcohol clinical trials.

Keywords

missing data; alcohol use disorder; relapse; treatment; clinical trials

Introduction

Participant attrition (i.e., dropout) in alcohol clinical trials can be substantial, with typical attrition rates ranging between 10% and 35% (e.g., Anton et al., 2006; Fertig et al., 2012; Johnson et al., 2007). The reasons for dropout are usually unknown, although numerous studies have evaluated predictors of dropout (Mackenzie et al., 1987; Postel et al., 2011; Prisciandaro et al., 2011; Sobell et al., 1984; Suh et al., 2008). Regardless of the reason for dropping out, participant attrition (or participant non-response at a specific assessment) can cause significant issues with regard to the handling of missing data in statistical analyses of treatment effects.

Missing Data: Mechanisms and Methods

The impact of missing data is often largely dependent on the process by which the data are missing. Rubin and colleagues (Little and Rubin, 2002; Rubin, 1976) developed a taxonomy of missing value mechanisms and simulation studies have provided guidance on the best analytic techniques for specific missing data mechanisms (Collins et al., 2001; Enders, 2011; Hedeker et al., 2007; Schafer and Graham, 2002). The recommendations from these studies have also been supported by guidelines for handling missing data in clinical trials from the National Research Council (2010).

According to Rubin and colleagues (Little and Rubin, 2002; Rubin, 1976), when the mechanism of the missing data is *missing completely at random* (MCAR) all missing values are unrelated to the observed values of the studied variables and unrelated to outcomes that were unobserved. In other words, participant attrition is completely unrelated to any of the constructs being studied. When the missing values are *missing at random* (MAR) then the missing values may be related to the observed values of the studied variables, but are completely unrelated to outcomes that were unobserved. Finally, when the missing values are *missing not at random* (MNAR) then the missing values are related to outcomes that were unobserved.

Unfortunately the mechanism of missingness is almost always unknown, thus the researcher can only assume that data are MCAR, MAR, or MNAR, and utilize an analytic approach that is best suited for the assumed missing data mechanism. Many methods have been developed for analyzing missing data and the interested reader is referred to numerous articles and books on the topic (Allison, 2001; Collins et al., 2001; Enders, 2011; Graham, 2009; Little et al., 2012; Schafer and Graham, 2002). In the current paper we focus on five methods: complete case analyses, single imputation techniques (e.g., last observation carried forward and the worst-case scenario of missing=heavy drinking), multiple imputation, and full information maximum likelihood. The performance of each of these approaches has been studied in numerous simulations and the findings consistently show that complete case techniques only produce unbiased results when data are MCAR and perform poorly when data are MAR or MNAR, while single imputation techniques are almost always problematic (see Hedden et al., 2009; Lane, 2008; Liu and Gould, 2002; Mallinckrodt et al., 2001). Multiple imputation and maximum likelihood have been shown to be statistically valid when data are MCAR and MAR (Barnes et al., 2010; Hedden et al., 2009; Lane, 2008; Mallinckrodt et al., 2001). As might be expected, all approaches generally perform poorly

when data are MNAR, in which case sensitivity analyses (e.g., pattern mixture modeling and/or selection models; Enders, 2011) are recommended (Little et al., 2012; Mallinckrodt et al 2008; Molenberghs et al., 2004).

Missing Data in Alcohol Clinical Trials

In the alcohol and substance abuse fields it is common for researchers to assume that missing data is indicative of relapse (Arndt, 2009). Under that assumption many researchers have relied on setting all missing values to the worst case scenario for the outcome (e.g., missing=heavy drinking; Falk et al., 2010). It has been suggested that assuming the worst case scenario will provide a conservative estimate of treatment effects (Papp et al., 2008), yet numerous simulation studies have provided evidence that using a single value (e.g., heavy drinking) to replace, or impute, missing data leads to severely biased treatment effect and standard error estimates (e.g., Barnes et al., 2010; Hedden et al., 2009; Hedeker et al., 2007; Lane, 2008; Mallinckrodt et al., 2001). Importantly, the treatment effect bias can favor the control or treatment group, and standard errors can be overestimated (increasing type-II errors) or underestimated (increasing type-I errors). The degree of bias from these different methods often depends on the patterns of missing data, rates of missing data in the treatment and control groups, and the mechanism of the missing data.

Researchers have found the worst case scenario missing data assumption in smoking trials (i.e., missing=smoking) to be strongly biased (Barnes et al., 2010; Hedeker et al., 2007), but no studies have examined the worst case scenario assumption in alcohol treatment trials. Importantly, alcohol and smoking clinical trials often differ, with studies of smoking often using a dichotomous outcome (e.g., smoking or not) and many alcohol trials using a continuous outcome (e.g., percent heavy drinking days). Previous research has consistently shown that multiple imputation and full information maximum likelihood produce more accurate results than complete case analysis and single-imputation methods, yet no studies have examined how the assumption of missing=heavy drinking impacts the results from alcohol clinical trials. Based on the prior research we anticipate that the assumption of missing=heavy drinking, as a single imputation method, will produce extremely biased results and that other methods of missing data analyses, such as multiple imputation and full information maximum likelihood will produce more accurate results in the context of missing data in alcohol clinical trials. It is particularly important to study the missing=heavy drinking assumption because the Food and Drug Administration (FDA) considers percent of subject with no heavy drinking days as the primary endpoint for Phase 3 Alcohol Clinical trials and has used the missing=heavy drinking assumption in evaluating the efficacy of alcohol treatment medications (FDA, 2006; also see Falk et al., 2010).

The goal of the current study was to examine various missing data assumptions and methods for analyzing treatment outcomes with missing data in a randomized trial for alcohol dependence. Specifically, we conducted simulations using real data from the COMBINE study (Anton et al., 2006) to examine the effect of naltrexone on percent heavy drinking days 10 weeks after treatment. Our primary focus was on assessing the performance of five methods for handling missing data (complete case analysis, last observation carried forward, missing=heavy drinking, multiple imputation, and full information maximum likelihood)

when data were generated by MCAR, MAR, and MNAR mechanisms. We started with observed data so that the observed distributions and covariance structures were maintained. The generated data and resulting analyses were then evaluated with respect to the known values observed in the complete dataset.

Materials and Methods

Participants

The data for this study are from the COMBINE study (COMBINE Study Research Group, 2003), a multi-site randomized trial. A total of 1383 subjects across 11 research sites were randomized into nine treatment groups. Treatment was provided for 16 weeks and participants were followed for one year following treatment.

The sample was recruited from inpatient and outpatient referrals at study sites and throughout the community. The final sample included 1,383 participants, 31% were female, 23% of the study sample were ethnic minorities (76.3% Non-Hispanic White, 11.6% Hispanic, 7.8% African American, and 4.1% Other). The subjects' median age was 44 years, 71% had at least 12 years of education, and 42% were married. Within the treatment period, 94% completed all drinking assessments, while 82.1% ($N = 1136$) completed the drinking assessment at ten weeks post-treatment. Only participants who completed the drinking assessment at ten weeks post-treatment were retained for the simulation.

Upon meeting the study criteria, subjects completed a baseline assessment and were randomly assigned to one of nine treatment groups. The Medical Management groups ($n=607$) included: Naltrexone, Acamprosate, Naltrexone + Acamprosate, and Placebo. The Combined Behavioral Intervention (CBI) groups ($n=776$) consisted of: Naltrexone + CBI, Acamprosate + CBI, Naltrexone + Acamprosate + CBI, Placebo + CBI, and CBI-only.

Assessments

Outcome measure—Percent heavy drinking days (PHD) was used as the primary outcome variable because it combines both frequency and intensity of drinking and is commonly used in alcohol clinical trials (Falk et al., 2010). The Form-90 interview (Miller and Del Boca, 1994) was used to calculate PHD. Heavy drinking was defined as 4 or more drinks per day for women and 5 or more drinks per day for men. Drinking measures were derived for the prior 30 days at each assessment point. In the current study we examined heavy drinking over the 30 days prior to the 10 week post-treatment assessment.

Auxiliary measures—Auxiliary variables were used to generate data that were MAR and to assist with the estimation of the multiple imputation and full information maximum likelihood models. Previous simulations have shown that using many auxiliary variables produces less biased results when multiple imputation and full information maximum likelihood are used (Collins et al., 2001; van Buuren et al., 1999). In the current study, auxiliary variables included alcohol dependence symptoms at baseline, treatment condition, gender, age, naltrexone adherence, and measures of drinking frequency and PHD derived at baseline and during the 16 weeks of treatment. Number of alcohol dependence symptoms was determined by Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition

(*DSM-IV*; American Psychiatric Association, 1994) criteria using the Structured Clinical Interview for *DSM-IV* (SCID; First et al., 1997).

Simulation Design

To evaluate the effects of various methods for handling missing data, participant dropout was simulated under various conditions (Burton et al., 2006; Hallgren, in press). In the MCAR condition of the simulation, participants in each naltrexone treatment condition (received naltrexone vs. did not receive naltrexone) dropped out at rates of 5% and 10% or 25% and 30%, with the probability of dropout unrelated to baseline or follow-up variables. In the MAR condition, dropout rates were likewise manipulated at 5% and 10% or 25% and 30% for each naltrexone condition, and the proportion of participants who dropped out within both conditions was also dependent on baseline alcohol dependence symptoms. Specifically, the probability of dropout was manipulated such that either 25% or 75% of participants who dropped out were above the median level of baseline alcohol dependence symptoms. In the MNAR condition, the proportion of participants who dropped out within both conditions was dependent on 10-week post-treatment PHD (the dependent variable in this study), such that either 25% or 75% of participants who dropped out were above the median post-treatment PHD. Sample size was manipulated at three levels by creating datasets that randomly sampled 200, 500, and 1000 participants with complete data at the 10 week post-treatment follow-up. Two hundred simulations of participant dropout were created within each of the 120 resulting conditions.

Missing Data Handling Methods

Each simulated dataset was analyzed using five methods for handling missing data. For the complete case analysis (CCA) condition, all participants with missing data at 10 weeks post-treatment were dropped from the analysis. For the last observation carried forward (LOCF) condition, values from the most recently available time point, typically the last four weeks during the treatment period, were carried forward to replace the missing values at the 10-week post-treatment time point. For the worst-case scenario (WCS) condition where it is assumed that missing=heavy drinking, participants with missing data were assumed to have relapsed to daily heavy drinking (100% PHD).

For the multiple imputation (MI) condition, missing data were imputed using chained regression equations in the *mice* package (van Buuren and Groothuis-Oudshoorn, 2011) in the R programming environment (R Development Core Team, 2012). *mice* estimated plausible values for missing data based on the mean and covariance structures among the missing data variable and the set of auxiliary variables listed above. For each simulated dataset, five imputation datasets were created with missing values imputed within their plausible ranges (e.g., 0 to 100 for PHD). Regression parameter estimates were then computed for each of the imputation datasets and were pooled according to Rubin's rules. A tutorial and syntax for using MI in the *mice* package are provided by van Buuren et al. (2001), and an example model with syntax is included as supplementary material.

For the full information maximum likelihood (FIML) condition the variance-covariance matrix for all available data was analyzed using Mplus version 7 (Muthén and Muthén,

2012). Through an expectation maximization algorithm, maximum likelihood uses all available data to identify the values of the model parameters that maximize the fit of the model to the observed data. Consistent with the idea of ordinary least square regression (Baraldi and Enders, 2010), maximum likelihood seeks those values for the parameters that minimize the distance between the observed data and the predicted data. The auxiliary variables listed above were included in each model. Example Mplus syntax for using FIML with auxiliary variables is included as supplementary material.

Analytic Plan

Linear regression models were analyzed for each simulated dataset with PHD as the dependent variable and naltrexone condition (0 = did not receive naltrexone, 1 = received naltrexone) as the independent variable. The *average naltrexone effect* (β) was computed within each simulation condition to determine whether the missing data handling method produced results that were systematically different from the dataset with no missing data. The *average standard error* (SE) was also computed for each condition to determine the effect of missing data on confidence of the parameter estimates and statistical power. Larger standard errors would indicate lower confidence in the treatment effect estimate (i.e., wider confidence intervals) and reduced statistical power. Smaller standard errors would indicate inflated confidence in the treatment effect estimate (i.e., smaller confidence intervals) and inflated statistical power. Finally, we calculated the *root mean square error* (RMSE) as a measure of efficiency and bias (Collins et al., 2001). The RMSE quantifies the overall degree of bias and inaccuracy in treatment effect estimates within each simulation condition. It is calculated as the average squared difference between the observed effect and the true effect. RMSE values closer to zero indicate less bias and greater accuracy, while larger values indicate greater bias and less accuracy.

Results

Figure 1 displays the average naltrexone effect estimates (top figures), average SE estimates (middle figures), and RMSE values (bottom figures) for all conditions with dropout rates of 5% and 10% and $n = 1000$. Missing data mechanisms are represented across the horizontal axis of each figure, e.g., MAR-high represents the group of conditions with higher dropout for participants with higher baseline dependence symptoms and MAR-low represents the group of conditions with higher dropout for participants with lower baseline dependence symptoms; MNAR-high represents conditions with higher dropout for participants with higher follow-up PHD and MNAR-low represents conditions with higher dropout for participants with lower follow-up PHD. Different rates of dropout within each condition of the missing data mechanism are also presented on the horizontal axis (“a” = 5% dropout in naltrexone and no-naltrexone conditions; “b” = 5% naltrexone dropout and 10% no-naltrexone dropout; “c” = 10% naltrexone dropout and 5% no-naltrexone dropout; “d” = 10% dropout in both conditions). CCA, LOCF, and WCS results are presented on the left side of the figures, and MI and FIML results are presented on the right side of the figures. Figure 2 displays results for the conditions with dropout rates of 25% and 30% and $n = 1000$ in a similar manner. The patterns of results in Figures 1 and 2 were similar for the smaller

sample sizes with $n = 200$ and 500 and are not presented here but are available in the supplementary tables to this manuscript.

Treatment effect (β) estimates

Mean naltrexone effects (top rows of Figures 1 and 2) estimated with CCA (left figure, solid line) had little deviation from the observed naltrexone effect in the dataset with no dropout (i.e., $\beta = -3.46$ in complete dataset; indicated by the horizontal dashed line) when data were MCAR. However, CCA results deviated more substantially from the observed naltrexone effect with no dropout when missing data were MAR or MNAR. Naltrexone effects estimated using LOCF (left figure, dashed line) deviated from the effect estimates with no dropout across missing data mechanisms, and typically overestimated the magnitude of the treatment effect (i.e., suggested a greater reduction in PHD due to receiving naltrexone than what actually occurred), especially when overall dropout rates were higher or when a greater proportion of participants who dropped out were in the naltrexone condition. Naltrexone effects estimated using WCS (left figure, dotted line) had the most substantial deviations from the estimates obtained with no dropout across all missing data mechanisms. Treatment effects estimated using WCS were highly inaccurate, at times doubling the magnitude of treatment effects and at other times reducing it to zero. Naltrexone effects estimated using MI and FIML (right figure, solid and dashed lines, respectively) were closest to the treatment effect estimated with no dropout under MCAR and MAR conditions. When data were MNAR, MI and FIML treatment effect estimates had some deviation, but the magnitude of the deviation was smaller compared to the CCA, LOCF, and WCS conditions (represented in the left figure). Treatment effect estimates obtained using MI and FIML were nearly identical.

Standard error (SE)

Mean SE estimates (see middle rows of Figures 1 and 2) estimated using CCA and WCS were always larger compared to the SE estimate from the dataset with no dropout (i.e., $SE = 2.00$ in complete dataset). As dropout rates increased, the magnitude of the SE inflation created by using CCA and WCS become larger, corresponding with larger reductions in statistical power and increased type-II errors. Alternatively, using LOCF slightly underestimated the magnitude of SE values, corresponding with artificially increased statistical power and increased type-I errors. Higher rates of dropout increased the magnitude by which SE values were underestimated when LOCF was used. Using MI and FIML produced SE estimates that were slightly inflated, corresponding with some loss of statistical power compared to the dataset with no dropout. However, the magnitude of inflation was smaller than when CCA and WCS were used, corresponding with smaller reductions in statistical power and type-II errors. SE estimates obtained using MI and FIML were nearly identical.

Root mean square error (RMSE)

RMSE values, which indicate inaccuracy in treatment effect estimates due to bias and variability of treatment effect estimates, were substantially higher when WCS was used compared to any other method. RMSE values also were higher for CCA and LOCF

compared to MI and FIML. For all methods, RMSE values were higher when data were MNAR. Averaged across all conditions, RMSE values were lowest in the MI (RMSE = 1.10) and FIML (1.01) conditions, and were highest in the CCA (1.48), LOCF (1.44), and WCS (3.06) conditions.

Discussion

The present study examined the effect of five methods for handling missing data under a variety of conditions on treatment effect estimates in a multisite randomized clinical trial for alcohol dependence. The results indicated that even with a modest amount of participant dropout (e.g., 5–10%), treatment effect estimates and standard errors computed from the same dataset can vary substantially based on the method one uses to handle missing data. As dropout rates become higher (e.g., 25–30%), the variability in treatment effect estimates and standard errors increases even more.

Across simulation conditions, the amount of bias in treatment effect estimates was highest when WCS was used to handle missing data. Even when participant dropout was equal between treatment groups, WCS often resulted in more biased treatment effect estimates than the other missing-data handling methods. Based on average estimates of the RMSE, using WCS produced results that were more than three-times more biased than the FIML approach and more than two-times more biased than the MI, CCA, and LOCF approaches. In some cases when participant dropout was uneven between groups, the amount of bias due to using WCS was large enough to cause the size of the treatment effect to be doubled or reduced to be effectively zero. Importantly, larger sample sizes (i.e., $n = 500$; $n = 1000$) did not protect against the bias caused by WCS. Further, WCS also resulted in the largest standard errors across simulation conditions due to the increased standard deviation caused by imputing missing = 100% heavy drinking. This indicates that WCS creates substantially larger confidence intervals of treatment effect estimates and reduces statistical power.

The use of CCA also resulted in biased treatment effect estimates except in the case when data were MCAR, in which case CCA produced unbiased effect estimates. Unfortunately, the mechanism of missingness is unknown in many clinical trials and the MCAR assumption often may be unreasonable, thus CCA is likely to produce biased results in the situations that are typical of most clinical trials. Further, due to decreased sample size, the use of CCA increased the size of standard errors and reduced statistical power.

The use of LOCF resulted in biased treatment effect estimates that typically overestimated the size of the naltrexone effect on reducing PHD. As the overall rate of dropout increased, the degree to which LOCF over-estimated the treatment effect also increased. The use of LOCF also underestimated standard error values, implying greater confidence in the treatment effect estimate than is actually warranted. Together, the tendency for LOCF to over-estimate treatment effects and yield smaller standard errors can lead to increased type-I errors, where ineffective treatments are perceived to be more effective than they actually are, and the degree of confidence in the treatment effect estimate is inflated.

In contrast, the use of MI and FIML produced the most accurate treatment effect and standard error estimates, indicating that these methods produce the most accurate results and retain most of the statistical power that would be present in a complete dataset without increasing type-I error rates. Even though all conditions produced biased results when data were MNAR, the deviation of treatment effects from their true values was still typically lower when MI or FIML were used compared to other methods.

The current findings are consistent with prior simulation studies of missing data in clinical trials (e.g., Lane, 2008; Mallinckrodt et al., 2001) and we echo the recommendations of many others to use MI, FIML, or other recommended modern missing data analytic tools (e.g., weighted estimating equations) in the analysis of alcohol clinical trial data when there are missing data (Little et al., 2012; Mallinckrodt et al., 2008; Molenberghs et al., 2004). We also caution against the use of CCA or LOCF, and especially discourage researchers from using the WCS (missing=heavy drinking) assumption.

Many people continue to use CCA because it is the default in many statistical programs (e.g., SPSS) or because modern missing data methods (e.g., MI, FIML) may be perceived as inaccessible. Numerous software programs (including Stata, SAS, SPSS, Mplus, and R) now incorporate MI and FIML, and there are many tutorials and reviews that describe the proper use of these methods in various software programs (e.g., Asparouhov and Muthén, 2010; Graham, 2009; van Buuren and Groothuis-Oudshoort, 2011; Zhang and Yiu-Fai, 2011). Many people continue to use LOCF or WCS because it is assumed that these approaches are more “conservative” and/or because single imputation is easy to implement. The results from the current study and many prior studies (e.g., Lane, 2008; Hedeker et al., 2007; Siddiqui 2011; Siddiqui et al., 2009) soundly refute the notion that LOCF and WCS produce conservative estimates of the treatment effect.

Limitations and Strengths

The primary limitation of the current study, and other simulation studies of missing data, is that reasons for missingness are rarely known by the researcher. The simulated conditions were developed to reflect possible real world scenarios, but in the real world the analyst does not know the missingness mechanism. Methods have been developed for examining the sensitivity of models to various missing data assumptions (Enders, 2011; Hedeker & Gibbons, 1997; Wu & Carroll, 1988)) and researchers are encouraged to consider using these methods to analyze longitudinal data (see Witkiewitz et al., 2012 for an example).

This study was also limited to only five approaches for handling missing data even though other approaches exist and are used. For example, there are numerous other single imputation approaches (e.g., group-mean imputation, baseline carried forward, missing=50% PHD) that were not considered, although based on prior studies we presume that these approaches would have also performed poorly in comparison to MI and FIML. There are also alternatives to MI and FIML that have proven useful and less biased than LOCF, CCA, and other approaches. For example, weighted estimating equations, which have also been supported as a favored method by the National Research Council (2010; Little et al., 2012), were not tested in the current study and similar approaches have been useful in studying missing data in substance abuse clinical trials (Hedden, 2009).

Finally, we view the use of real data in the current study as both a limitation and strength. The use of real complete data is limiting because it includes the drinking data of participants who did not dropout and their data could be dissimilar to those participants who dropped out of the study. An analysis of the full COMBINE dataset with those with missing data at 10 weeks posttreatment compared to those with complete data indicated no significant differences between groups in baseline drinking rates, alcohol dependence severity, demographic variables, readiness to change, social support, social network drinking, or quality of life. The groups were different on baseline drinking consequences ($t(1379) = -2.17, p = 0.03$) with individuals with missing data reporting approximately 3 more consequences, on average.

The use of real data is also a major strength of the current study because we retained the real-world features of clinical trial data. The use of real data also allowed us to compare the treatment effect that was estimated under various conditions to the treatment effect observed in the complete dataset. A related strength is the use of the COMBINE data, which is being analyzed in many secondary data analytic studies. Finally, the current study adds to the literature on approaches for handling missing data by examining the effect of the WCS (missing=heavy drinking) assumption on non-dichotomous treatment outcomes.

Summary

Participant attrition is common in alcohol clinical trials and there are several methods for handling missing data. The results from the current study, in addition to several prior studies, provide evidence that using CCA, LOCF, or WCS will bias results when there are missing data, particularly if data are MAR or MNAR. These techniques do not bring us any closer to identifying reliably effective treatments for alcohol use disorders and, in the case of assuming missing=heavy drinking, also discredit those clients who are successful following treatment by stigmatizing and stereotyping them as treatment failures for dropping out of a clinical trial (Arndt, 2009). Researchers are encouraged to consider missing data techniques, such as FIML and MI, which are readily available in many popular software programs. Recommendations to not use CCA and single imputation methods (e.g., LOCF, WCS) have been voiced for decades and we urge the alcohol treatment research community to consider these recommendations when conducting analyses.

References

- Allison, PD. Missing data. Sage; Thousand Oaks, CA: 2001.
- American Psychiatric Association. Diagnostic and Statistical Manual (DSM-IV). 4. Washington DC: 1994.
- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, Gastfriend DR, Hosking JD, Johnson BA, LoCastro JS, Longabaugh R, Mason BJ, Mattson ME, Miller WR, Pettinati HM, Randall CL, Swift R, Weiss RD, Williams LD, Zweben A. Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. *JAMA*. 2006; 295:2003–2017. [PubMed: 16670409]
- Arndt S. Stereotyping and the treatment of missing data for drug and alcohol clinical trials. *Subst Abuse Treat Prev Policy*. 2009; 4:2. [PubMed: 19226454]
- Asparouhov, T.; Muthén, B. [Accessed Mar 28, 2013] Multiple imputation with Mplus version 2. 2010 Sep 29. 2010. Available at: www.statmodel.com/download/Imputations7.pdf

- Baraldi AN, Enders CK. An introduction to modern missing data analyses. *J Sch Psychol.* 2010; 48:5–37. [PubMed: 20006986]
- Barnes SA, Larsen MD, Schroeder D, Hanson A, Decker PA. Missing data assumptions and methods in a smoking cessation study. *Addiction.* 2010; 105:431–437. [PubMed: 20402986]
- Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Statist in Med.* 2006; 25:4279–4292.
- Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods.* 2001; 6:330–351. [PubMed: 11778676]
- COMBINE Study Research Group. Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: rationale and methods. *Alcohol Clin Exp Res.* 2003; 27:1107–1122. [PubMed: 12878917]
- Enders CK. Missing not at random models for latent growth curve analyses. *Psychol Methods.* 2011; 16:1–16. [PubMed: 21381816]
- Falk D, Wang XQ, Liu L, Fertig J, Mattson M, Ryan M, Johnson B, Stout R, Litten RZ. Percentage of subjects with no heavy drinking days: evaluation as an efficacy endpoint for alcohol clinical trials. *Alcohol Clin Exp Res.* 2010; 34:2022–2034. [PubMed: 20659066]
- Fertig JB, Ryan ML, Falk DE, Litten RZ, Mattson ME, Ransom J, Rickman WJ, Scott C, Ciraulo D, Green AI, Tiouririne NA, Johnson B, Pettinati H, Strain EC, Devine E, Brunette MF, Kampman KA, Tompkins D, Stout R. A double-blind, placebo-controlled trial assessing the efficacy of levetiracetam extended-release in very heavy drinking alcohol-dependent patients. *Alcohol Clin Exp Res.* 2012; 36:1421–1430. [PubMed: 22324516]
- First, MB.; Gibbon, M.; Spitzer, RL.; Williams, JBW.; Benjamin, LS. *Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II)*. Washington, D.C: American Psychiatric Press, Inc; 1997.
- FDA. Medical Review of Vivitrol: 21-897. US Government; Rockville, Maryland: 2006.
- Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol.* 2009; 60:549–576. [PubMed: 18652544]
- Hallgren, K. *Tutorials in Quant Method for Psychol. Conducting simulation studies in the R programming environment.* (in press)
- Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychol Methods.* 1997; 2:64–78.
- Hedeker D, Mermelstein RJ, Demirtas H. Analysis of binary outcomes with missing data: missing = smoking, last observation carried forward, and a little multiple imputation. *Addiction.* 2007; 102:1564–1573. [PubMed: 17854333]
- Hedden SL, Woolson RF, Carter RE, Palesch Y, Upadhyaya HP, Malcolm RJ. The impact of loss to follow-up on hypothesis tests of the treatment effect for several statistical methods in substance abuse clinical trials. *J Subst Abuse Treat.* 2009; 37:54–63. [PubMed: 19008067]
- Johnson BA, Rosenthal N, Capece JA, Wiegand F, Mao L, Beyers K, McKay A, Ait-Daoud N, Anton RF, Ciraulo DA, Kranzler HR, Mann K, O'Malley SS, Swift RM. Topiramate for treating alcohol dependence: a randomized controlled trial. *JAMA.* 2007; 298:1641–1651. [PubMed: 17925516]
- Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharm Stat.* 2008; 7:93–106. [PubMed: 17351897]
- Little RA, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H. The prevention and treatment of missing data in clinical trials. *N Engl J Med.* 2012; 367:1355–1360. [PubMed: 23034025]
- Little, RA.; Rubin, DB. *Statistical Analysis with Missing Data.* 2. Wiley; New York: 2002.
- Liu G, Gould AL. Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *J Biopharm Stat.* 2002; 12:207–226. [PubMed: 12413241]
- Mackenzie A, Funderburk FR, Allen RP, Stefan RL. The characteristics of alcoholics frequently lost to follow-up. *J Stud Alcohol.* 1987; 48:119–123. [PubMed: 3560947]
- Mallinckrodt CH, Clark SW, David SR. Accounting for dropout bias using mixed-effects models. *J Biopharm Stat.* 2001; 11:9–21. [PubMed: 11459446]

- Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Inf J*. 2008; 42:303–319.
- Miller WR, Del Boca FK. Measurement of drinking behavior using the form 90 family of instruments. *J Stud Alcohol Suppl*. 1994; 12:112–118. [PubMed: 7722987]
- Molenberghs G, Thijs H, Jansen I, Beunckens C. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*. 2004; 5:445–464. [PubMed: 15208205]
- Muthén, LK.; Muthén, B. *Mplus Users Guide*. 7. Muthén & Muthén; Los Angeles, CA: 2012.
- National Research Council. *The prevention and treatment of missing data in clinical trials*. Washington, DC: National Academies Press; 2010. <http://www.nap.edu/catalog.php?recordid=12955>
- Papp KA, Fonjallaz P, Casset-Semanaz F, Krueger JG, Mittkowsky KM. Approaches to reporting long-term data. *Curr Med Res Opin*. 2008; 24:2001–2008. [PubMed: 18534049]
- Postel MG, De Haan HA, Ter Huurne ED, Van der Palen J, Becker ES, De Jong CAJ. Attrition in web-based treatment for problem drinkers. *J Med Internet Res*. 2011; 13:e117. [PubMed: 22201703]
- Prisciandaro JJ, Rembold J, Brown DG, Brady KT, Tolliver BK. Predictors of clinical trial dropout in individuals with co-occurring bipolar disorder and alcohol dependence. *Drug Alcohol Depend*. 2011; 118:493–496. [PubMed: 21549529]
- Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *J Stud Alcohol*. 1997; 58:7–29. [PubMed: 8979210]
- R Development Core Team. *R: a language and environment for statistical computing [computer software] version 2.15.0*. 2012.
- Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–592.
- Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002; 7:147–177. [PubMed: 12090408]
- Siddiqui O. MMRM versus MI in dealing with missing data--a comparison based on 25 NDA data sets. *J Biopharm Stat*. 2011; 21:423–436. [PubMed: 21442517]
- Siddiqui O, Hung HMJ, O'Neill R. MMRM vs LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *J Biopharm Stat*. 2009; 19:227–246. [PubMed: 19212876]
- Sobell LC, Sobell MB, Maisto SA. Follow-up attrition in alcohol treatment studies: is “no news” bad news, good news or no news? *Drug Alcohol Depend*. 1984; 13:1–7. [PubMed: 6734420]
- Suh JJ, Pettinati HM, Kampman KM, O'Brien CP. Gender differences in predictors of treatment attrition with high dose naltrexone in cocaine and alcohol dependence. *Am J Addict*. 2008; 17:463–468. [PubMed: 19034737]
- van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*. 1999; 18:681–694. [PubMed: 10204197]
- van Buuren S, Groothuis-Oudshoort K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011; 45:1–67.
- Witkiewitz K, Bush T, Magnusson LB, Carlini BH, Zbikowski SM. Trajectories of cigarettes per day during the course of telephone tobacco cessation counseling services: a comparison of missing data models. *Nicotine Tob Res*. 2012; 14:1100–1104. [PubMed: 22180594]
- Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*. 1988; 44:175–188.
- Zhang, W.; Yiu-Fai, Y. [Accessed March 28, 2013] A tutorial on structural equation modeling with incomplete observations: multiple imputation and FIML methods using SAS. 2011 Jul 21. 2011. Available at: support.sas.com/rnd/app/stat/papers/imps2011_FIML.pdf

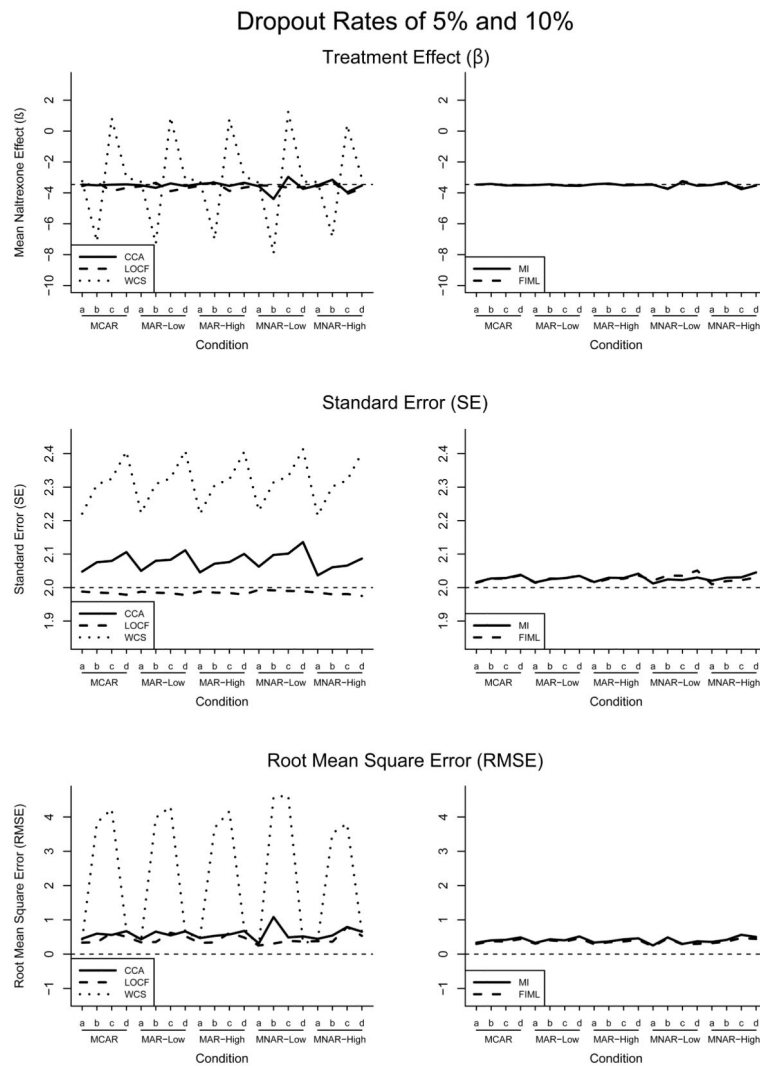


Figure 1. Model results for percentage of heavy drinking days (PHD) predicted by naltrexone condition. CCA = complete case analysis; LOCF = last observation carried forward; WCS = worst case scenario; MI = multiple imputation; FIML = full information maximum likelihood; MCAR = missing completely at random; MAR-high = missing at random with higher dropout rates in high-baseline dependence group; MAR-low = missing at random with higher dropout rates in low-baseline dependence group; MNAR-high missing not at random with higher dropout rates in high-post-treatment PHD group; MNAR-low missing not at random with higher dropout rates in low-post-treatment PHD group; a = 5% dropout in both naltrexone and no-naltrexone groups; b = 5% dropout in naltrexone group and 10% dropout in no-naltrexone group; c = 10% dropout in naltrexone group and 5% dropout in no-naltrexone group; d = 10% dropout in both groups.

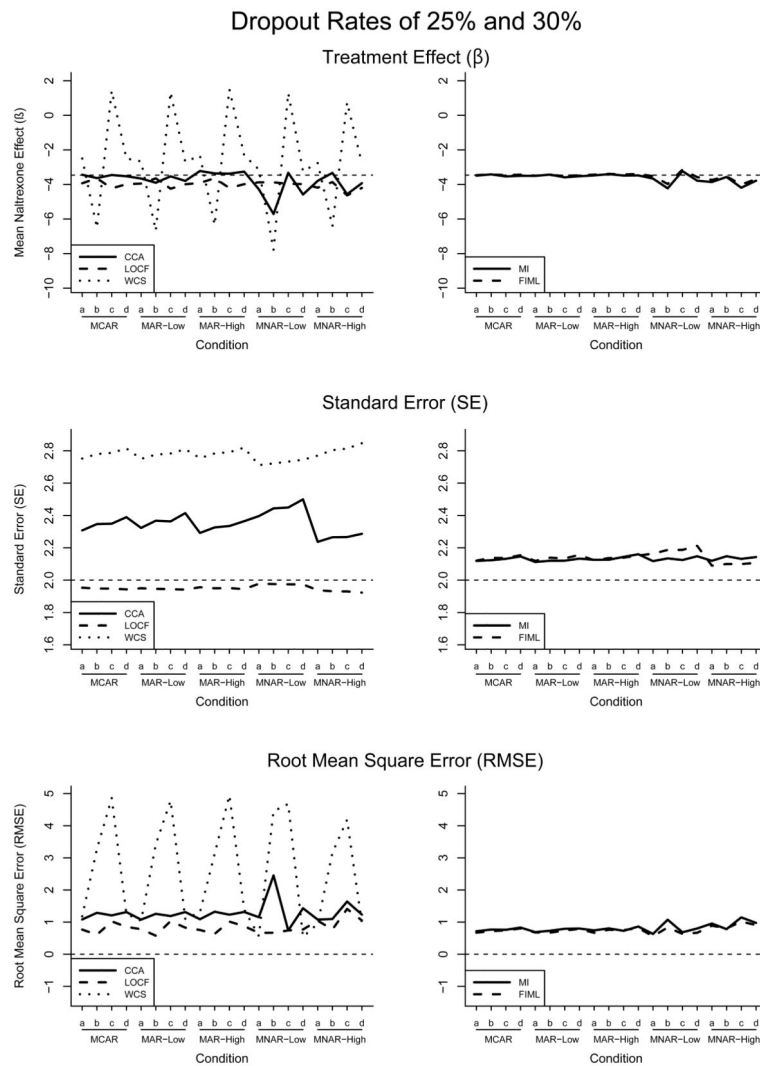


Figure 2. Model results for percentage of heavy drinking days (PHD) predicted by naltrexone condition. CCA = complete case analysis; LOCF = last observation carried forward; WCS = worst case scenario; MI = multiple imputation; FIML = full information maximum likelihood; MCAR = missing completely at random; MAR-high = missing at random with higher dropout rates in high-baseline dependence group; MAR-low = missing at random with higher dropout rates in low-baseline dependence group; MNAR-high missing not at random with higher dropout rates in high-post-treatment PHD group; MNAR-low missing not at random with higher dropout rates in low-post-treatment PHD group; a = 25% dropout in both naltrexone and no-naltrexone groups; b = 25% dropout in naltrexone group and 30% dropout in no-naltrexone group; c = 30% dropout in naltrexone group and 25% dropout in no-naltrexone group; d = 30% dropout in both groups.