HHS Public Access

Author manuscript

Alcohol Clin Exp Res. Author manuscript; available in PMC 2017 July 24.

Published in final edited form as:

Alcohol Clin Exp Res. 2016 July ; 40(7): 1548–1557. doi:10.1111/acer.13106.

Missing Data in Alcohol Clinical Trials with Binary Outcomes

Kevin A. Hallgren, PhD,
University of Washington, Department of Psychiatry and Behavioral Sciences

Katie Witkiewitz, PhD,
University of New Mexico, Department of Psychology, Center on Alcoholism, Substance Abuse, and Addictions

Henry R. Kranzler, MD,
Department of Psychiatry and Center for Studies of Addiction, University of Pennsylvania Perelman School of Medicine and VISN 4 MIRECC, Crescenz Philadelphia VAMC

Daniel E. Falk, PhD,
National Institute on Alcohol Abuse and Alcoholism

Raye Z. Litten, PhD,
National Institute on Alcohol Abuse and Alcoholism

Stephanie S. O'Malley, PhD, and
Department of Psychiatry, Yale School of Medicine

Raymond F. Anton, MD
Department of Psychiatry, Addiction Sciences Division, Medical University of South Carolina

In conjunction with the Alcohol Clinical Trials Initiative (ACTIVE) Workgroup

Abstract

The corresponding author, Kevin Hallgren, may be contacted at 1959 NE Pacific Street, Box 356560, Seattle, WA, 98195, United States; khallgre@uw.edu; phone: 206-616-2906; fax: 206-543-9520.

Author note:
Dr. Kranzler has been a consultant, advisory group member, or CME lecturer for Alkermes, Lundbeck, Otsuka, and Indivior. Dr. Witkiewitz has been a consultant for Alkermes. Drs. Witkiewitz, Kranzler, Falk, Litten, O'Malley, and Anton are members of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials (ACTIVE) Group, which is supported by Abbvie, Alkermes, Ethypharm, Lilly, Lundbeck, and Pfizer. In addition to the authors, the following individuals are or were members of the ACTIVE Workgroup and provided intellectual input into this manuscript during attendance at workgroup meetings.
Deborah Hasin Ph.D., Columbia University
Joann Fertig, Ph.D., National Institute on Alcohol Abuse and Alcoholism
Karl Mann, M.D., Central Institute of Mental Health, Mannheim, Germany
David McCann, Ph.D., National Institute on Drug Abuse
Roger Meyer, M.D., Penn State Hershey Medical Center
Charles O'Brien, M.D., Ph.D., University of Pennsylvania
Thomas Permutt, Ph.D., Food and Drug Administration
Rebecca Robinson, M.S., Eli Lilly & Company
Francoise Trinquet, M.D., Ethypharm
Celia Winchell, M.D., U.S. Food and Drug Administration
Conrad Wong, Ph.D., Eli Lilly & Company
Bernard Silverman, M.D., Alkermes
Benjamin Zakine, M.D., Ethypharm
Sarah Timm (ASCP) provided important administrative support to the ACTIVE workgroup.

Background—Missing data are common in alcohol clinical trials for both continuous and binary endpoints. Approaches to handle missing data have been explored for continuous outcomes, yet no studies have compared missing data approaches for binary outcomes (e.g., abstinence, no heavy drinking days). The present study compares approaches to modeling binary outcomes with missing data in the COMBINE study.

Method—We included participants in the COMBINE Study who had complete drinking data during treatment and who were assigned to active medication or placebo conditions (*N*=1146). Using simulation methods, missing data were introduced under common scenarios with varying sample sizes and amounts of missing data. Logistic regression was used to estimate the effect of naltrexone (vs. placebo) in predicting any drinking and any heavy drinking outcomes at the end of treatment using four analytic approaches: complete case analysis (CCA), last observation carried forward (LOCF), the worst-case scenario of missing equals any drinking or heavy drinking (WCS), and multiple imputation (MI). In separate analyses, these approaches were compared when drinking data were manually deleted for those participants who discontinued treatment but continued to provide drinking data.

Results—WCS produced the greatest amount of bias in treatment effect estimates. MI usually yielded less biased estimates than WCS and CCA in the simulated data, and performed considerably better than LOCF when estimating treatment effects among individuals who discontinued treatment.

Conclusions—Missing data can introduce bias in treatment effect estimates in alcohol clinical trials. Researchers should utilize modern missing data methods, including MI, and avoid WCS and CCA when analyzing binary alcohol clinical trial outcomes.

Keywords

Alcohol clinical trials; missing data; multiple imputation; simulation study; naltrexone

Missing data are a common and substantial problem in the conduct of clinical trials for alcohol use disorder (AUD). Missing data most often result from participant attrition (i.e., dropping out of the trial), the rate of which ranges from roughly 10% to more than 50% of the allocated patients in some AUD clinical trials (Anton et al. 2006, Johnson et al. 2007, Mason et al. 2014, Witte et al. 2012). Reasons for dropping out of a trial are highly variable and may not be known to the researcher. While some individuals drop out due to their continued alcohol use, others may drop out because they have decreased their alcohol use and no longer desire treatment. Participants may also drop out due to the adverse effects of a medication, because they have moved away from the research study site, or because they have been incarcerated or changed jobs (Ball et al. 2006). Some individuals also report dropping out for psychosocial concerns (e.g., lack of support from family or friends), treatment program issues (e.g., cost, inadequate alliance with clinicians), and practical challenges (e.g., transportation, childcare; Palmer et al. 2009).

Given the varied reasons individuals drop out from clinical trials (Ball et al. 2006, Palmer et al. 2009) and the importance of drawing valid conclusions from clinical trial findings, a potential for bias exists if researchers assume a particular outcome when follow-up data are missing. Assuming that missing data are equivalent to treatment failure (e.g., any drinking or

heavy drinking in an AUD clinical trial) carries strong assumptions about the outcomes that have not been observed and may substantially bias estimates of the treatment being tested. Because missing data in AUD clinical trials are common, researchers should consider approaches for analyzing AUD clinical trial data that minimize the bias associated with attrition and other causes of nonresponse (National Research Council 2010, Witkiewitz et al. 2015).

Three decades of research on missing data for continuous outcomes in clinical trials has provided fairly conclusive guidance that model-based approaches, such as maximum likelihood estimation and multiple imputation, typically yield the least biased estimates when data are “missing completely at random” (MCAR) or “missing at random” (MAR; Enders 2010, Graham 2009, Little & Rubin 2002, Rubin 1976, Schafer & Graham 2002). MCAR and MAR refer to the mechanism of the missing data and whether there is a systematic relationship between the data that are missing and the outcomes of interest. Data are MCAR when there is no association between the missing values, other studied variables, and the outcomes of interest. Data are MAR when there is an association between the missing values and other studied variables that are known (e.g., dropout related to baseline alcohol dependence severity). Data are missing not at random (MNAR; also called non-ignorable missingness) when the missing values are systematically related to outcomes that were not observed (e.g., missing values are related to drinking after the individual dropped out of the trial and the drinking outcomes are not known to the researcher). Although it is impossible to know the actual reasons for all missing data, sensitivity analyses can provide a test of whether the assumptions of MCAR, MAR, or MNAR are likely for a given set of analyses (Enders 2011).

Recently, two studies examined the effects of various missing data assumptions and missing data analysis approaches to estimate treatment effects in alcohol clinical trials (Hallgren & Witkiewitz 2013, Witkiewitz et al. 2014). Hallgren and Witkiewitz (2013) conducted a simulation study with real data from the COMBINE study (Anton et al. 2006) and simulated participant dropout from random subsets of participants to examine the effect of missing data under several conditions (e.g., varying sample size and dropout rate) and missing data scenarios (e.g., MCAR, MAR, MNAR). The simulation focused on the naltrexone versus placebo treatment effect estimates on the continuous outcome of percentage of heavy drinking days. Results indicated that multiple imputation (MI) and maximum likelihood (ML) produced similar results with the least amount of bias in estimating the naltrexone effect and yielded more accurate standard errors than complete case analysis (CCA), last observation carried forward (LOCF), or missing equals heavy drinking. MI and ML also performed somewhat better than other approaches when data were MNAR. The worst performing missing data approach was assuming the worst case scenario (WCS) of missing equals heavy drinking, which produced the most inflated standard errors and the most biased treatment effect estimates across simulation conditions.

Witkiewitz and colleagues (2014) focused on continuous drinking outcomes among individuals in the COMBINE study (Anton et al. 2006) who dropped out of treatment (i.e., discontinued medication), but continued to provide drinking data for the duration of the study period, a group that may be similar to those who were lost to follow-up. Rather than

randomly assigning dropout, Witkiewitz and colleagues (2014) manually deleted follow-up drinking data from the 185 participants who discontinued treatment but provided drinking data. The missing data approaches were applied and compared with the "true results" based on the reported data. Consistent with Hallgren and Witkiewitz (2013), the findings indicated that MI and ML yielded similar results and were most likely to recover the true effect size estimates and standard errors that were observed in the original sample. The WCS approach assuming missing equals heavy drinking also performed the worst by greatly underestimating the observed treatment effect and overestimating standard errors.

The findings from these two studies are consistent with those of prior studies using continuous outcomes with simulated datasets (Ayele et al. 2014) and pharmacotherapy trials (Siddiqui et al. 2009). However, it remains unclear how these methods fare when the outcome of interest is binary (e.g., any drinking vs. no drinking or any heavy drinking vs. no heavy drinking). Given that the Food and Drug Administration (FDA) and others have recently recommended percent of subjects with total abstinence or percent of subjects with no heavy drinking days as the primary endpoints for Phase III alcohol clinical trials (Falk et al. 2010, FDA 2015), we examined approaches for handling missing data for these binary outcomes. Notably, approaches for handling missing binary outcomes differ from those for continuous outcomes. For example, imputation procedures must incorporate binary outcomes for missing values (i.e., 0's and 1's). Further, ML methods that handle missing data with binary outcomes are unavailable in many widely used software packages. Some studies have suggested that binary outcomes are impacted by the missing data approach only under extreme conditions (e.g., high rates of missingness; Caille et al., in press, Jackson et al. 2014, Ma et al. 2011). However, the effects of different approaches for handling missing binary data have not been studied in the context of alcohol clinical trials.

The present study compared the effects of different approaches for handling missing binary outcomes with the goal of informing recommendations for alcohol clinical trials. This extends previous work that tested similar approaches with continuous outcomes under simulated conditions of dropout (Hallgren & Witkiewitz 2013) and among individuals who dropped out of treatment but continued to provide drinking data (Witkiewitz et al. 2014).

Materials and Methods

Participants

The data and methods for this study are similar to those described previously for missing data with continuous outcomes (Hallgren & Witkiewitz 2013, Witkiewitz et al. 2014). Data were obtained from the COMBINE study, a randomized clinical trial of combination pharmacotherapy and behavioral interventions for alcohol dependence conducted at 11 research sites (see Anton et al. 2006). Alcohol dependent adults (age 18 and older; mean age = 44.6 years (SD=10.2), 76.8% non-Hispanic white, 69.0% men) were randomized to one of nine outpatient treatment conditions in which they received combinations of naltrexone (vs. placebo), acamprosate (vs. placebo), and medication management with or without a combined behavioral intervention (CBI) or CBI only (no pills) for up to 16 weeks. All participants received medical management—a brief intervention focusing on medication adherence, lab results, side effects, and goals of treatment— up to 9 times over the 16 week

trial. CBI consisted of up to 20 weekly individual sessions over 16 weeks and incorporated content from motivational interviewing, cognitive behavioral treatment, and twelve-step facilitation (Miller et al. 2004).

The original COMBINE Study sample included 1383 participants. In the present study, we excluded participants who received CBI with no pills (n=157) because the main analysis contrasted naltrexone vs. placebo and participants who were missing any within-treatment drinking data to ensure that missing data were controlled completely through the simulation procedures (n=80). This left a final sample of 1146, including 961 participants who completed treatment and 185 who dropped out of treatment early but who had complete within-treatment (full 16 week) drinking data.

Measures

Outcome measures—Two primary, dichotomous outcome variables were utilized: (1) the presence of any drinking versus abstinence and (2) the presence of any heavy drinking days versus no heavy drinking days. Both outcomes were measured over the 28 days prior to the 16-week follow-up assessment, corresponding approximately to the last month of the treatment period. Drinking was assessed using a Timeline Follow-Back method via the Form-90 interview (Miller 1996). Heavy drinking was defined as consuming 4 or more drinks within a single day for women and 5 or more drinks within a single day for men.

Auxiliary measures—Additional variables were used to stratify the likelihood of missing data in the simulation study and to aid in the estimation of multiple imputation models. The inclusion of a greater number of auxiliary variables in multiple imputation is generally associated with reduced bias and greater information recovery (Collins et al. 2001, van Buuren et al. 1999). We chose the following auxiliary variables based on their association with end-of-treatment drinking and the likelihood that they are available in other alcohol clinical trials: alcohol dependence severity at baseline, treatment condition, sex, age, medication (naltrexone) adherence rate, percentage of drinking days and percentage of heavy drinking days measured at baseline and during each month of treatment, and binary indicators of any drinking and any heavy drinking for each month during treatment. Baseline alcohol dependence severity was a sum of the seven alcohol dependence criteria from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV; American Psychiatric Association 1994), measured using Module E of the Structured Clinical Interview for DSM-IV (SCID; First et al. 1996).

Approaches to Handle Missing Drinking Data

The effects of four approaches for handling missing drinking data were compared. Each approach aimed to test the same effect: namely, the impact of receiving naltrexone vs. placebo on binary outcomes (any drinking or any heavy drinking) at the end of treatment via logistic regression. Complete case analysis (CCA; i.e., listwise deletion) utilized data only from participants without missing drinking data during the last 28 days of the 16-week follow-up by dropping all participants with missing data from the statistical analysis. For the last observation carried forward (LOCF) approach, missing outcomes for the last month of the trial (week 16) were replaced with the last-available measurement of the same outcome,

obtained at baseline, week 4, week 8 or week 12. The time periods that last observations were sampled from were based on the real rates of missing data (i.e., “last observations”) in the COMBINE study and reflected the true rate of dropout at varying points during treatment. Specifically, the last observations were carried forward from week 12, week 8, week 4, and baseline for 20%, 22%, 49%, and 8% of the participants with missing data, respectively. These percentages are proportional to the rates that missing data first occurred among participants who did not provide week-16 data. Because the timing of dropout can vary across studies, we also tested LOCF (see online supplementary materials) with 50% of the missing observations carried forward from baseline (i.e., dropout occurring between baseline and the first follow-up) and the other 50% carried forward from week 8 (i.e., dropout occurring between mid-treatment and end-of-treatment). For the worst-case scenario (WCS) approach, missing data were set to values indicating that drinking or heavy drinking had occurred.

For the multiple imputation (MI) condition, missing drinking outcomes were imputed with plausible values (i.e., 0=not drinking or 1=any drinking) that were predicted by chained regression equations using the mice package (van Buuren & Groothuis-Oudshoorn 2011) in R (R Core Development Team 2015). Estimates of missing values were obtained from a logistic regression model that included the auxiliary variables listed above. Twenty imputation datasets were created for each simulated dataset. Regression parameter estimates were computed for each imputation dataset and then pooled using Rubin’s rules (Rubin 1976). A tutorial and syntax for using MI in the *mice* package are provided by van Buuren et al. (2011).

Maximum likelihood (ML) approaches were originally considered and tested using common statistical software (R, Mplus version 7.3). Because pilot testing indicated that ML approaches could not recover information from missing binary outcomes in a logistic regression framework using these software packages, we did not test ML in the present study.

Simulation Design

Simulated dropout study—Dropout was simulated by manipulating the mechanism of missingness, rates of dropout, and sample size (Hallgren 2013, Hallgren & Witkiewitz 2013). Each simulation condition modeled the effects of random assignment to the naltrexone condition (with or without additional acamprosate or CBI) vs. placebo naltrexone (with our without additional acamprosate or CBI) on binary drinking outcomes using logistic regression. The effects of assignment to the CBI and acamprosate conditions were not tested. In the simulated MCAR conditions, dropout in each treatment condition (naltrexone vs. placebo) was simulated at rates of 5% and 10% or 25% and 30% (an upper rate not uncommon in AUD clinical trials), with dropout being unrelated to baseline or follow-up variables. In the MAR condition, dropout was also set at rates of 5% and 10% or 25% and 30% for each treatment condition; however, consistent with the MAR assumption (that missing data are dependent on other studied variables that are known), dropout was made contingent on baseline alcohol dependence symptom severity such that either 25% or 75% of participants who dropped out were above the median level of dependence severity.

These values were selected to represent a 3-to-1 imbalance in dropout rates by dependence severity and to mirror prior work (Hallgren & Witkiewitz 2013). In the MNAR condition, the proportion of participants who dropped out within both conditions was conditional on post-treatment heavy drinking, such that either 25% or 75% of the missing follow-up data were for heavy drinkers. Sample size was manipulated at three levels by creating datasets with random samples of 200, 500, and 1000 participants with complete follow-up data. Participant dropout was simulated 1000 times within each combination of conditions.

Treatment non-completer study—Because the true mechanism of missing data can never be known, a set of follow-up analyses aimed to mimic missing data among a subset of participants who may be most similar to individuals likely to be lost to follow-up in clinical trials. Specifically, for the subset of participants who discontinued treatment but continued to provide follow-up assessment data, we set drinking data values to missing. Thus, this dataset included the 961 participants who completed treatment (and had complete drinking data) and the 185 participants who discontinued treatment (and had drinking data set to missing). The effects of naltrexone on binary drinking and heavy-drinking outcomes were then examined using logistic regression for each missing data approach, the results of which were compared to the complete dataset. The analysis was performed across the full dataset (N = 1146) and within the subsample of participants who discontinued treatment (N = 185).

Analytic Plan

The performance of each approach for handling missing data was evaluated by comparing regression coefficient estimates and standard errors of treatment effects against the “true” logistic regression parameter estimates that were obtained in the absence of missing data. Logistic regression analyses followed the intention-to-treat principle and were not adjusted to control for medication adherence. The logistic regression coefficient of the naltrexone effect (β), which is equal to the log-odds ratio of random assignment to naltrexone on drinking outcomes, was compared across simulation conditions to test whether a method systematically biased estimates of treatment effects relative to the dataset with no missing data. The sign of the regression coefficients reflects the direction of the difference between groups, with a negative coefficient indicating that drinking or heavy-drinking rates were lower in the naltrexone condition than in the placebo condition, and the size of coefficient reflects the magnitude of the difference between the treatment conditions. Standard errors (SE) were compared to test whether a method over- or under-estimated the precision or confidence of the estimate.

Results

Simulated Dropout Study

Figure 1 shows the average treatment effect estimates of naltrexone on binary drinking outcomes (top figure) and average SE estimates (bottom figure) for the simulated dropout conditions with n = 1000 and dropout rates of 25% and 30%. The mechanism of missing data is shown across the horizontal axis of each figure. Missing data scenarios labeled “MAR-high” indicate the conditions with higher dropout for participants with higher baseline dependence symptoms and “MAR-low” labels the conditions with higher dropout

for participants with lower baseline dependence symptoms. Scenarios labeled “MNAR-high” and “MNAR-low” show results for scenarios with higher dropout rates for participants with higher and lower follow-up heavy drinking rates, respectively. Different rates of dropout within each scenario are represented on the horizontal axis above each missing data assumption (scenario “a” = 25% missing data in naltrexone and placebo conditions; “b” = 25% missing data in naltrexone and 30% missing data in placebo conditions; “c” = 30% missing data in naltrexone and 25% missing data in placebo conditions, and “d” = 30% missing data in both conditions). Figure 2 displays the results in an identical format for binary heavy-drinking outcomes. Although not displayed here, the patterns of results in Figures 1 and 2 were similar for the other tested sample sizes (n = 200 or 500) and similar in pattern but larger in magnitude than those for smaller rates of missing data (5% or 10% missing); these results are available in online supplementary materials.

Treatment effect (β) estimates—Mean logistic regression coefficients represent the estimated log odds ratios of the effect of naltrexone on binary drinking (top of Figure 1) and heavy-drinking outcomes (top of Figure 2). Estimates based on complete case analysis (CCA; dotted line) deviated only slightly from the “true” estimate obtained with no missing data (horizontal dashed line) when data were MCAR or MAR, but had greater deviation under most conditions when data were MNAR. Estimates based on the worst-case scenario assumption of missing equals heavy drinking (WCS; line with alternating dots/dashes) were often the most biased except in the MNAR-high conditions, where dropout rates were simulated to be much higher among heavier drinkers. WCS was especially biased when dropout rates were not equal between groups, but still produced bias when dropout rates were equal between groups and even when data were MCAR. Last observation carried forward (LOCF; dashed line) had less extreme bias than WCS but often had more bias than CCA and MI under MCAR and MAR conditions. However, LOCF also had relatively less bias than other approaches in several MNAR conditions. Notably, when last observations were carried forward at different rates from previous assessments (i.e., 50% from baseline data, 50% from week-8 data), the level of bias was often more pronounced (see online supplemental materials), indicating that the LOCF performs differently depending on when most of the last observations were observed. Estimates based on multiple imputation (MI; solid line) had relatively low bias when data were MCAR or MAR. MI performed less well under most MNAR conditions. Although MI was consistently less biased than CCA under MNAR conditions, the methods with the least amount of bias often alternated between MI, LOCF, and WCS, depending on the missing data scenario.

Standard Errors (SE)—Mean SE estimates were always highest using CCA. As expected, MI produced lower SE estimates than CCA but higher SE estimates than the true SE in the complete data. Mean SE estimates using WCS were usually lower than those of MI and closer to the true SE for the analyses of heavy drinking, but were much closer to the complete data values for the analyses of heavy drinking. In contrast, mean SE’s were almost uniformly estimated at their “true” values with complete data using LOCF, indicating that LOCF likely underestimates the uncertainty in treatment effect estimates despite the higher degree of uncertainty in the outcome due to missing data (Little & Rubin 2002). In other

words, LOCF results often suggested an inflated level of confidence in a potentially biased estimate of the treatment effect.

Treatment Non-Completer Study

Among participants who dropped out of treatment but continued to provide drinking data (n = 185), 97 received naltrexone and 88 received placebo; the dropout rates between the naltrexone and placebo groups were not statistically different (p = 0.50). Compared to participants who completed treatment, participants who discontinued treatment had significantly higher rates of drinking (77.3% vs. 60.0%, respectively, p = 0.006) and heavy drinking (68.1% vs. 41.6%, respectively, p < .001).

Figure 3 displays the results of logistic regression models predicting any drinking (left panels) and any heavy drinking (right panels) when follow-up drinking data were set to missing for participants who discontinued treatment. Results are displayed for both the full sample (top panels) and only participants who discontinued medication and withdrew from treatment (bottom panels). Treatment effect estimates and ±1 SE intervals are presented for the "true" complete dataset (heavier lines) and for each approach for handling missing data (lighter lines). Gray regions correspond to the ±1 SE interval for the "true" complete dataset.

None of the approaches perfectly replicated the complete-data treatment effect and standard error. In the full sample models (top panels of Figure 3), MI and CCA had the lowest bias for both outcomes, LOCF and WCS both had greater bias in the direction of underestimating the effect of naltrexone, with WCS being the most biased approach. Consistent with the simulation results, CCA had the largest SE intervals, followed by MI, WCS, and LOCF.

Results for logistic regression models that were restricted to the subset of participants who discontinued treatment are shown in the bottom panels of Figure 3[1]. Among this subset of participants, LOCF was notably more biased, for example, with the effect of naltrexone on any heavy drinking being biased by more than one standard error. Despite this bias, standard error estimates were similar for LOCF and the complete data. In contrast, MI had lower bias than LOCF and larger standard errors.

Discussion

The present study tested methods for handling missing data when participants have missing data due to dropout in alcohol clinical trials. While previous research (Hallgren & Witkiewitz 2013, Witkiewitz et al. 2014) examined these issues in the context of continuous drinking outcomes (e.g., percentage of heavy drinking days), the present study extends this work by testing methods for handling missing data with binary outcomes (i.e., any drinking and any heavy drinking). This extension into binary outcomes is important because recent draft guidance by the FDA supports the use of binary measures of any drinking and/or heavy drinking as primary endpoints in clinical trials (FDA 2015). Moreover, with binary outcomes

[1]Treatment effects could not be estimated for this subsample via CCA (due to missing outcomes for the full subsample) or via WCS (due to all missing outcomes being classified as heavy drinking, creating a lack of variance in the dependent variable). MI estimates in this analysis were obtained by including the full sample (N=1146) in the imputation process but including only the subsample of treatment non-completers (N=185) in the actual analysis of treatment effects and pooling of multiple imputation results.

there are several practical, computational differences from continuous measures in modeling (e.g., use of logistic regression, no longer assuming normally-distributed outcomes) and tools available to handle missing data (e.g., use of MI for binary outcomes and the lack of suitable ML estimators for missing binary outcomes in many software packages).

Results of the present study illustrate that treatment effect estimates in alcohol clinical trials can be biased and the choice of analytic technique can affect the degree of bias, potentially resulting in substantively different conclusions about the efficacy of interventions. For example, using complete data, the odds ratio (OR) of any heavy drinking days for naltrexone relative to placebo was 0.76 (Figure 2). In the presence of missing data, the estimated mean ORs varied from 0.57 to 0.92, depending on the method used.

Using WCS (i.e., setting missing values to indicate that drinking or heavy drinking occurred) usually led to the greatest bias in treatment effect estimates, particularly in the simulated dropout scenarios with unequal dropout between conditions and when participants who discontinued treatment were modeled to have missing data. Importantly, the amount of bias associated with WCS was often similar or greater with sample sizes of 200 and 500 (see online supplementary materials), which is important because most alcohol clinical trials have sample sizes less than 500.

When data were MNAR, the magnitude of the bias was often lower when using LOCF compared to other approaches. Nonetheless, LOCF tended to have greater systematic bias than CCA or MI when data were MCAR or MAR and when participants who discontinued treatment were modeled to have missing data. This bias was typically in the direction of underestimating effect sizes of naltrexone on drinking and heavy-drinking outcomes. LOCF yielded standard errors that were approximately equal to the complete-data standard errors. Several of these LOCF findings were unexpected and should be interpreted with caution. For example, although LOCF provided standard errors that were approximately equal to the complete-data results, they may actually reflect overestimated confidence in the precision of the treatment effect estimate. That is, with data missing from up to 30% of the sample, there is good reason to be less confident in the precision of the treatment effect estimate than if there were complete data. This uncertainty is reflected in the larger standard error estimates for MI and CCA, but not LOCF (Rubin 1987). In scenarios where LOCF estimates are unbiased, the lower standard errors may not lead to substantive misinterpretations, but in scenarios where LOCF estimates are biased, the lower standard errors would lead one to have an inflated level of confidence in a biased treatment effect estimate. In contrast, the elevated standard errors in MI may reflect an appropriate increase in the uncertainty of parameter estimates due to missing values (Little & Rubin 2002). In addition, results obtained using LOCF are highly dependent on the time point that the last drinking measures were obtained from participants. For example, in the COMBINE Study, the greatest increase in missing data (49%) occurred between week 4 and week 8, resulting in much of the week-4 data being carried forward. In studies characterized by dropout at different time points or that use different assessment schedules that lead to more data being carried forward from baseline (when drinking is relatively more common) or later in treatment (when drinking is relatively less common), LOCF may yield substantively different conclusions.

This notion was supported by our supplementary analysis, where dropout was assumed to occur at different times (see supplemental materials).

The present findings are consistent with other research that has simulated MCAR and MAR binary data in randomized trials and found minimal bias with MI (Hardt et al. 2013) and little difference between MI and CCA (Caille et al., in press, Ma et al. 2011). However, these studies did not test MNAR conditions, where MI consistently outperformed CCA in the present study, nor did they test LOCF or WCS approaches. Smolkowski et al. (2010) also compared CCA, WCS, LOCF, and MI using data from a tobacco cessation clinical trial and found that WCS and MI yielded similar treatment effect estimates to one another, which were generally smaller than LOCF and CCA estimates. Although their analysis used missing data due to actual participant dropout and the amount of bias in each method could therefore not be determined, sensitivity analyses suggested that MI yielded the most robust results.

The impact of each approach for handling missing data can vary based on the characteristics of the data that were used. For example, consistent with expectations from prior studies, CCA was likely to yield the least biased estimated effect when missingness is completely unrelated to the outcome itself (e.g., MCAR conditions). LOCF produces greater bias when drinking outcomes change considerably between the last observation and the outcome period and less bias when drinking is highly stable over time. WCS has the greatest likelihood of introducing bias in most scenarios, although the amount of bias is relatively smaller when dropout is much higher among participants who continue to drink rather than abstain (e.g., 75% vs. 25% of dropout in MNAR-high conditions reported here). MI typically offers less biased estimates compared to other methods as the amount of missing data increases and as the correlation between auxiliary variables and the missing outcome increases (e.g., in this case, drinking and heavy-drinking measures at baseline and earlier periods of treatment). Unfortunately, many of these factors cannot be known when applied to data from clinical trials. Thus, MI may offer the safest approach for modeling binary drinking outcomes in clinical trial data because missingness is often highly prevalent, data may often be MAR or MNAR, and drinking may change considerably over time.

The present study has noteworthy strengths. Data were drawn from a large, well-known clinical trial with good follow-up rates. Binary outcomes are commonly used in clinical trials, but methods for handling missing binary outcomes have not been studied in relation to alcohol treatment outcomes. The present report addresses this methodological gap. The analytic framework tested here also followed the intention-to-treat principle, which is an approach generally required by the FDA.

The present study also has several limitations. In the simulation study, the mechanisms of missingness were specified but the reasons for missing data are usually unknown in real-world studies. Thus, while the results obtained here reflect possible scenarios for missing data, it would be impossible to know which scenario maps onto actual missing data situations in an alcohol clinical trial. In addition, we were unable to test maximum-likelihood (ML) based estimators because many common statistical software packages do not have ML-based estimators that handle missing data for binary outcomes using logistic regression. This limitation was also present in prior studies of missing binary outcomes.

Although the present study followed the intention-to-treat principle, other analytic approaches that adjust for treatment compliance were not tested, such as complier average causal effect models (Jo 2002) and it is possible that such analyses could be affected differently by missing data. Finally, to be consistent with the FDA's focus on a single binary outcome for a pre-specified period of time (e.g., the last month of treatment) as the preferred approach to the analysis of alcohol clinical trials, the current study focused on logistic regression as the analysis model. However, many studies in the alcohol treatment literature have used mixed models (e.g., Anton et al. 2006) and the results from binary outcomes analyzed in a mixed modeling framework could yield different conclusions regarding the best missing data approaches.

Summary

Participant attrition is common in alcohol clinical trials and methods to reduce bias due to missing data are increasingly used by alcohol treatment researchers. In scenarios that are likely in clinical trials (i.e., MAR or MNAR), WCS and CCA are likely to bias treatment effect estimates and conclusions. MI and LOCF often yield less biased treatment effect estimates, although both may still be substantially biased under some MNAR conditions and LOCF typically fails to account for the loss in precision of estimates due to missing data in estimates of standard errors. Based on previous research and the current findings, we recommend against using CCA, WCS, and LOCF for binary or continuous outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support for this research was provided by NIAAA grants T32AA007455 and R01AA022328.

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: American Psychiatric Press; 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; 295:2003–17. [PubMed: 16670409]

Ayele BT, Lipkovich I, Molenberghs G, Mallinckrodt CH. A multiple-imputation-based approach to sensitivity analyses and effectiveness assessments in longitudinal clinical trials. J Biopharm Stat. 2014; 24:211–28. [PubMed: 24605966]

Ball SA, Carroll KM, Canning-Ball M, Rounsaville BJ. Reasons for dropout from drug abuse treatment: symptoms, personality, and motivation. Addict Behav. 2006; 31:320–30. [PubMed: 15964152]

Caille A, Leyrat C, Giraudeau B. A comparison of imputation strategies in cluster randomized trials with missing binary outcomes. Stat Methods Med Res. in press.

Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods. 2001; 6:330–51. [PubMed: 11778676]

Enders, CK. Applied missing data analysis. New York, NY: Guilford Press; 2010.

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; 34:2022–34. [PubMed: 20659066]

First, MB., Spitzer, RL., Gibbon, M., Williams, JBW. Structured Clinical Interview for DSM-IV Axis I Disorders. Washington, D.C.: American Psychiatric Press, Inc; 1996.

Food and Drug Administration. Alcoholism: Developing drugs for treatment. Silver Spring, MD: Author; 2015.

Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol. 2009; 60:549–76. [PubMed: 18652544]

Hallgren KA. Conducting Simulation Studies in the R Programming Environment. Tutor Quant Methods Psychol. 2013; 9:43–60. [PubMed: 25067989]

Hallgren KA, Witkiewitz K. Missing data in alcohol clinical trials: a comparison of methods. Alcohol Clin Exp Res. 2013; 37:2152–60. [PubMed: 23889334]

Hardt J, Herke M, Brian T, Laubach W. Multiple imputation of missing data: a simulation study on a binary response. Open J Stat. 2013; 3:370–8.

Jackson D, White IR, Mason D, Sutton S. A general method for handling missing binary outcome data in randomized controlled trials. Addiction. 2014; 109:1986–93. [PubMed: 25171441]

Jo B. Model misspecification sensititivity analysis in estimating causal effects of interventions with non-compliance. Stat Med. 2002; 21:3161–81. [PubMed: 12375297]

Johnson BA, Rosenthal N, Capece JA, Wiegand F, Mao L, Beyers K, McKay A, Ait-Daoud N, Anton RF, Ciraulo DA, Kranzler R, Mann K, O'Malley SS, Swift RM. Topiramate for treating alcohol dependence: a randomized controlled trial. JAMA. 2007; 298:1641–51. [PubMed: 17925516]

Little, RA., Rubin, D. Statistical analyses with missing data. New York: Wiley; 2002.

Ma J, Akhtar-Danesh N, Dolovich L, Thabane L. Imputation strategies for missing binary outcomes in cluster randomized trials. BMC Med Res Methodol. 2011; 11:18. [PubMed: 21324148]

Mason BJ, Quello S, Goodell V, Shadan F, Kyle M, Begovic A. Gabapentin treatment for alcohol dependence: a randomized clinical trial. JAMA Intern Med. 2014; 174:70–77. [PubMed: 24190578]

Miller, WR. Form 90: A structured assessment interview for drinking and related behaviors. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1996.

Miller, WR. Combined Behavioral Intervention Manual: A clinical research guide for therapists treatmetn people with Alcohol Abuse and Dependence. Vol. 1. Bethesda, MD: NIAAA; 2004. COMBINE Monograph SeriesDHHS Publication No. 04-5288

National Research Council. The prevention and treatment of missing data in clinical trials. Washington, DC, US: The National Academies Press; 2010.

Palmer RS, Murphy MK, Piselli A, Ball SA. Substance user treatment dropout from client and clinician perspectives: a pilot study. Subst Use Misuse. 2009; 44:1021–38. [PubMed: 19938942]

R Core Development Team. R: A language and environment for statistical computing. 2015

Rubin D. Inference and missing data. Biometrika. 1976; 63:581–92.

Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods. 2002; 7:147–77. [PubMed: 12090408]

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; 19:227–46. [PubMed: 19212876]

Smolkowski K, Danaher BG, Seeley JR, Kosty DB, Severson HH. Modeling missing binary outcome data in a successful web-based smokeless tobacco cessation program. Addict. 2010; 105:1005–15.

Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med. 1999; 18:681–94. [PubMed: 10204197]

Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011; 45:1–67.

Witkiewitz K, Falk DE, Kranzler HR, Litten RZ, Hallgren KA, O'Malley SS, Anton RF. Methods to analyze treatment effects in the presence of missing data for a continuous heavy drinking outcome

measure when participants drop out from treatment in alcohol clinical trials. Alcohol Clin Exp Res. 2014; 38:2826–34. [PubMed: 25421518]

Witkiewitz K, Finney JW, Harris AHS, Kivlahan DR, Kranzler HR. Recommendations for the design and analysis of treatment trials for alcohol use disorders. Alcohol Clin Exp Res. 2015; 39:1557–70. [PubMed: 26250333]

Witte J, Bentley K, Evins AE, Clain AJ, Baer L, Pedrelli P, Fava M, Mischoulon D. A randomized, controlled, pilot study of acamprosate added to escitalopram in adults with major depressive disorder and alcohol use disorder. J Clin Psychopharmacol. 2012; 32:787–96. [PubMed: 23131884]

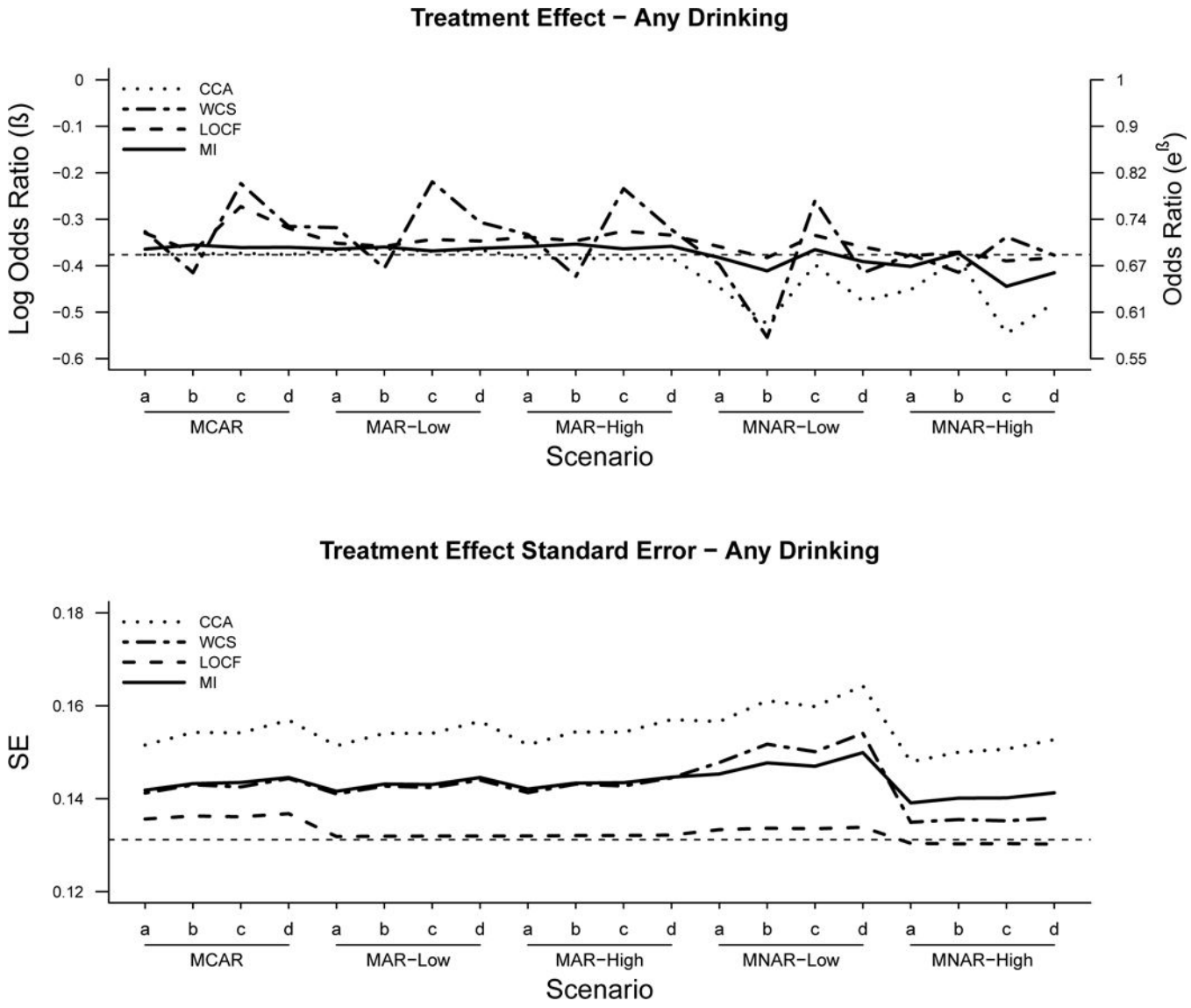

Figure 1.

Model results for effect of naltrexone on any drinking. CCA = complete case analysis; LOCF = last observation carried forward; WCS = worst case scenario; MI = multiple imputation; 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 participants with post-treatment heavy drinking; MNAR-low missing not at random with higher dropout rates participants with no post-treatment heavy drinking; a = 25% dropout in both naltrexone and placebo groups; b = 25% dropout in naltrexone group and 30% dropout in placebo group; c = 30% dropout in naltrexone group and 25% dropout in placebo group; d = 30% dropout in both groups.

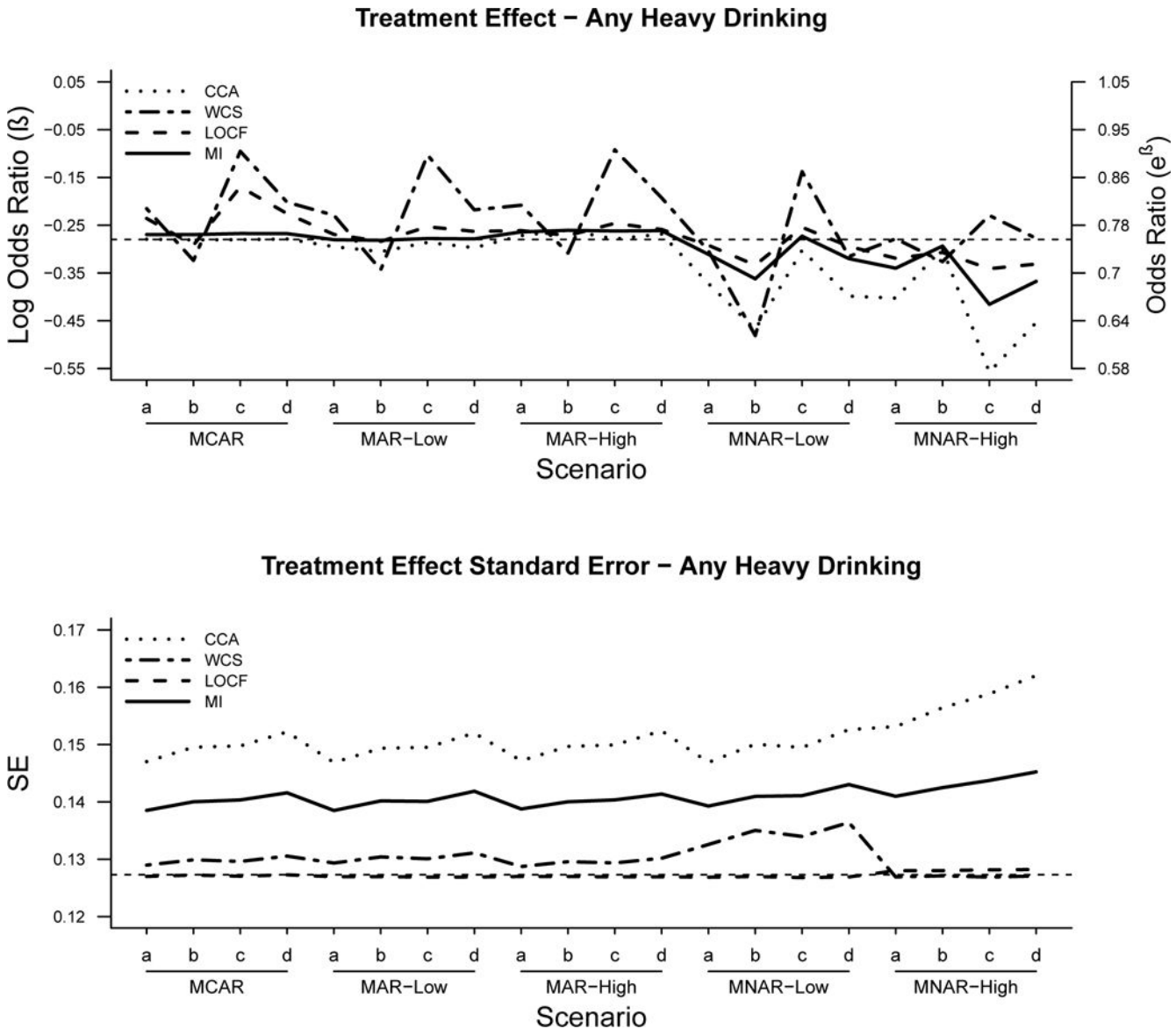

Figure 2.

Model results for effect of naltrexone on any heavy drinking days. CCA = complete case analysis; LOCF = last observation carried forward; WCS = worst case scenario; MI = multiple imputation; 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 participants with post-treatment heavy drinking; MNAR-low missing not at random with higher dropout rates participants with no post-treatment heavy drinking; a = 25% dropout in both naltrexone and placebo groups; b = 25% dropout in naltrexone group and 30% dropout in placebo group; c = 30% dropout in naltrexone group and 25% dropout in placebo group; d = 30% dropout in both groups.

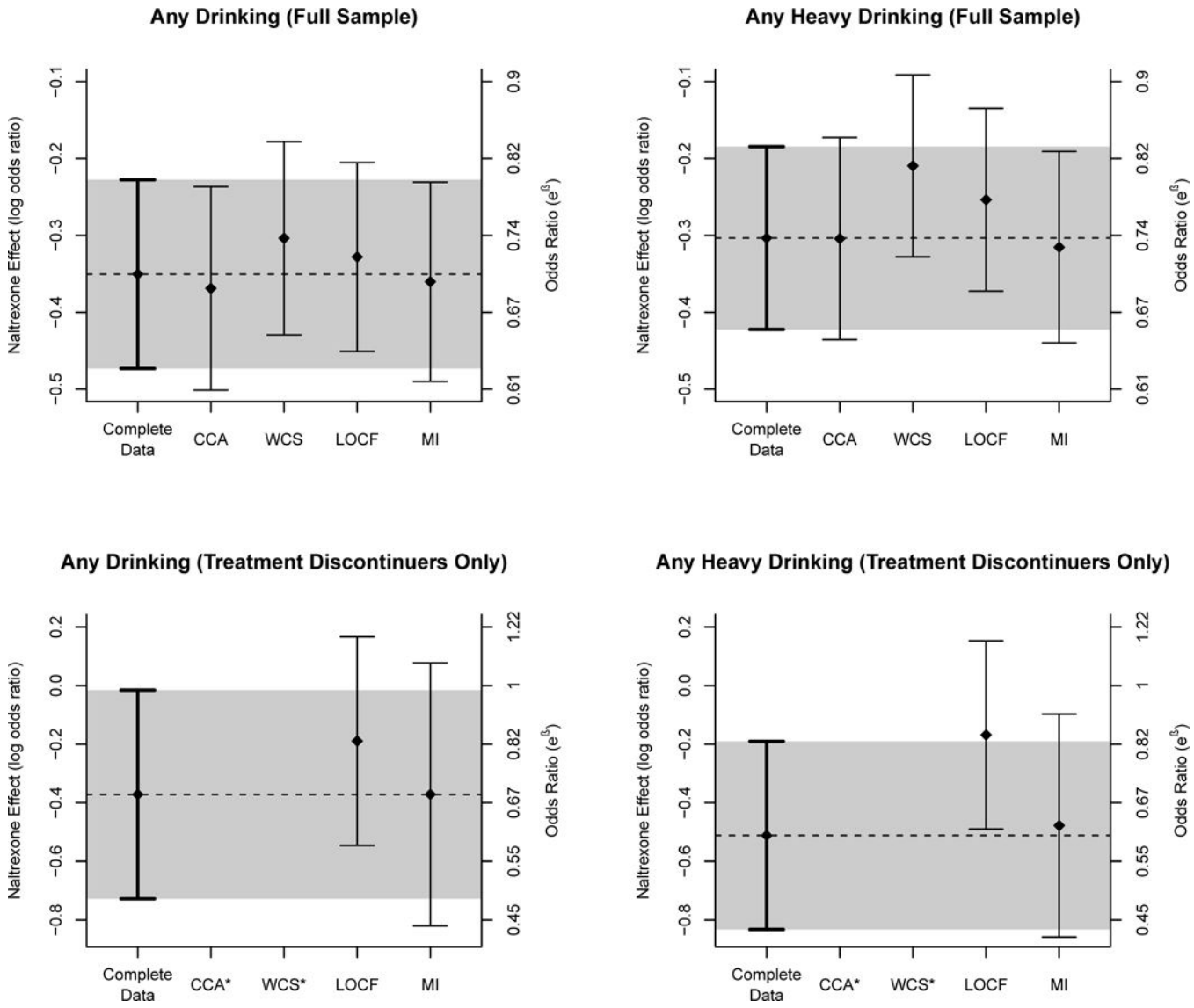

Figure 3.

Treatment effect estimates and ±1 SE intervals for the effects of naltrexone on any-drinking and heavy-drinking outcomes when data from participants who dropped out of treatment were set to missing. The dashed horizontal line represents "true" (i.e., "actual effect") before values were set to missing. The full sample results (top panels) represent analyses using the full sample ($N = 1146$); the withdrawers-only sample represents analyses using only the participants who discontinued treatment ($N = 185$). *In the withdrawers-only sample, CCA and WCS could not be computed due to all observations of the dependent variable being set to missing (CCA) or to drinking/heavy-drinking (WCS).