Published in final edited form as:

Addiction. 2015 January ; 110(1): 51–58. doi:10.1111/add.12714.

A ‘Missing Not at Random’ (MNAR) and ‘Missing at Random’ (MAR) Growth Model Comparison with a Buprenorphine/Naloxone Clinical Trial

Sterling McPherson, Ph.D.^{a,c,d,e,f,b}, Celestina Barbosa-Leiker, Ph.D.^{a,c,d,e,f}, Mary Rose Mamey, M.A.^c, Michael McDonell, Ph.D.^g, Craig K. Enders, Ph.D.^h, and John Roll, Ph.D.^{a,c,d,e,f,g}

^aCollege of Nursing, Washington State University

^cWashington State University, Department of Psychology

^dPrograms of Excellence in the Addictions, Washington State University

^eProgram for Rural Mental Health and Substance Abuse Treatment, Washington State University

^fTranslational Addictions Research Center, Washington State University

^gDepartment of Psychiatry and Behavioral Sciences, University of Washington School of Medicine

^hDepartment of Psychology, Arizona State University

Abstract

Aims—To compare three missing data strategies: 1) Latent growth model that assumes the data are missing at random (MAR) model, 2) Diggle-Kenward missing not at random (MNAR) model where dropout is a function of previous/concurrent urinalysis (UA) submissions, and 3) Wu-Carroll MNAR model where dropout is a function of the growth factors.

Design—Secondary data analysis of a National Drug Abuse Treatment Clinical Trials Network trial that examined a 7-day versus 28-day taper (i.e., stepwise decrease in buprenorphine/naloxone) on the likelihood of submitting an opioid-positive UA during treatment.

Setting—11 outpatient treatment settings in 10 US cities.

Participants—516 opioid dependent participants.

Measurements—Opioid UAs provided across the 4-week treatment period.

Findings—The MAR model showed a significant effect ($B = -0.45$, $p < 0.05$) of trial arm on the opioid-positive UA slope (i.e., 28-day taper participants were less likely to submit a positive UA over time) with a small effect size ($d = 0.20$). The MNAR Diggle-Kenward model demonstrated a

^bCorresponding author: Please address all requests for reprints to Dr. Sterling McPherson, College of Nursing, PO 1495, SNRS 314C, Washington State University, Spokane, WA, 99210-1295, U.S.A.; Phone: (509) 324-7459; Fax: (509) 324-7341; smcpherson05@wsu.edu.

Declarations of Interest

None of the other authors have any financial, personal, or other type of relationship that would cause a conflict of interest that would inappropriately impact or influence the research and interpretation of the findings.

significant ($B=-0.64$, $p<0.01$) effect of trial arm on the slope with a large effect size ($d=0.82$). The MNAR Wu-Carroll model evidenced a significant ($B=-0.41$, $p<0.05$) effect of trial arm on the UA slope that was relatively small ($d=0.31$).

Conclusions—This performance comparison of three missing data strategies (latent growth model, Diggle-Kenward selection model, Wu-Carroll selection model) on sample data indicates a need for increased use of sensitivity analyses in clinical trial research. Given the potential sensitivity of the trial arm effect to missing data assumptions, it is critical for researchers to consider whether the assumptions associated with each model are defensible.

Keywords

substance use disorder treatment; randomized clinical trials; longitudinal missing data; missing not at random models; latent growth modeling; sensitivity analysis

Introduction

The missing data mechanism is often not thoroughly evaluated as part of the analytic strategy in many substance use disorder (SUD) randomized clinical trials (RCTs)(1–3). This is a problem because treatment efficacy can vary as a function of how the missing information is handled, which can jeopardize a research team's ability to make sound clinical inference (4–7). While many modern missing data methods exist, these methods have not been thoroughly explicated and compared with clinical trial data (2, 5, 8). Our approach is based on the technical publication by Enders (6) on missing data growth models and geared towards real-world modeling that would likely come before final model specification and interpretation (9). This paper investigates the impact of how different missing data procedures can lead to different inferences for clinical practice when the primary parameter of interest is the impact of trial arm on substance use (i.e., measured via urinalysis [UA]) over time.

This investigation considers two missing data assumptions. The first assumption of missing at random (MAR) indicates that observed scores can effectively predict the propensity or probability of missing data and after controlling for these observed scores, missing data is unrelated to UA results. Conversely, missing not at random (MNAR) assumes that the missing data is related to unobserved variables (e.g., attrition is related to one's UA result, such that those likely to test positive are more likely to dropout) even after controlling for the observed data. Empirically testing for the true cause of missing values is not currently possible because MAR and MNAR require propositions about the unobserved values. A systematic, comparative evaluation (i.e., sensitivity analysis) of different methods with different assumptions can prove vital in order to inform the research team's judgment regarding treatment effectiveness (1, 2, 5, 6, 8, 10–13). It is likely that many missing data situations fall somewhere on the continuum between MAR and MNAR (14, 15) making such sensitivity analyses all the more important in order to guide the research team in their treatment effect interpretation.

Typical maximum likelihood (ML) analyses invoke the MAR assumption and so does multiple imputation (MI); two state of the art methods (15). ML, or sometimes referred to as

full information or direct ML, is easily implemented when estimating a latent growth model (LGM). Under equivalent conditions (e.g., covariates, parameter constraints) multiple imputation and ML produce almost equivalent results (13, 15). However, given ML's ease of use within the LGM framework and to increase pedagogical clarity, this is the method we chose for the current investigation. ML and MI are related to the field of statistics, psychometrics, and biostatistics, and MI is a method based largely on Bayesian theory. Two MNAR approaches, both originating in biostatistics, with different assumptions regarding treatment dropout were compared to the MAR model.

The Diggle-Kenward approach models the MNAR assumption that the probability of attrition at a specific time point is *directly* related to the outcome score both before and after dropout occurred (6, 8, 16, 17). This model assumes that dropout is a function of time-specific outcomes (5, 6). The Wu-Carroll selection model makes the MNAR assumption that the probability of dropout is a function of one's developmental trajectory (5, 6, 8, 18). Thus, in latent growth modeling terminology, the intercept and slope growth factors are regressed onto each of the time-specific dropout indicators (8, 19, 20). Dropout is treated as a function of *indirect* change in the outcome scores over time. While these two MNAR modeling strategies offer only a sampling of the large and developing array of MNAR modeling strategies, they will introduce the plethora of models that can be fit when performing sensitivity analyses. These example MNAR models are a good starting point for beginning a series of sensitivity analyses because such models are estimable in existing software.

The Diggle-Kenward and Wu-Carroll selection models (6, 8, 16) are related to econometrics and biostatistics, and both have strong distributional assumptions that are easy to violate. Based on additional work in those fields, Follman and Wu (21) have presented a generalized model similar to the model we will estimate here but does not require specification of a distribution. Albert et al. (22) furthered this and other work (23–25) by proposing a model for longitudinal MNAR binary data where a Gaussian autoregressive process is shared between the outcome and the mechanism of missing data. Unfortunately, neither of these models is readily available in commercial software to the authors' knowledge. This limits their utility for the scope of the current investigation and one reason why these models were not included. Also, our models are the same as what Enders' (2010) paper (6) presents in a more technical example

This investigation compares three different missing data strategies intended to illustrate the importance of conducting sensitivity analyses. The first is the MAR growth model, which utilizes ML. This model is compared to two MNAR models; Diggle-Kenward (16) and Wu-Carroll (18) selection models. We hypothesized that the association of 1) trial arm, and 2) intercept UA with the linear UA slope is dependent on the missing data strategy used. While we have previously explored the impact of different missing data strategies on treatment effects both cross-sectionally using logistic regression (2) and longitudinally using generalized estimating equations (7), we have not compared or explicated MNAR strategies via sensitivity analyses.

Methods

Participants and Procedures

Below, we summarize more complete descriptions of the participants, procedures, and primary outcomes of this clinical trial, which can be found in the originally reported clinical trial (26) and in our previous missing data work (2, 7).

The publicly available clinical trial dataset used for this analysis was a National Drug Abuse Treatment Clinical Trials Network (0003) trial of two different buprenorphine/naloxone tapering schedules (7- versus 28-day)(26). The purpose of this trial was to compare the impact of an administration schedule of a 7-day versus 28-day taper of buprenorphine/naloxone on the likelihood of a positive UA for opioid use (UA+). The sample of 516 participants was scheduled for a weekly visit across the 4-week treatment period for a total of 5 possible UA submissions (including the pre-treatment, baseline UA). Data used for this set of comparative growth model analyses included the same participants (n = 516). As previously reported, 44% of the UAs were missing across the 4-week treatment period while approximately 72% of the sample dropped out by the end of the 4 weeks and 27% of the sample contained sporadic missing values but did not dropout (7).

Participants were stabilized on buprenorphine/naloxone prior to the 4-week treatment period. Patients were subsequently stratified across individualized maintenance doses of buprenorphine/naloxone and randomized to one of the two trial arms (i.e., 7-day or 28-day taper schedules) (2, 26). Just as other papers that have found when using this clinical trial dataset, both primary and secondary, we did not find any statistical differences between the two groups on series of relevant variables (2, 7, 27). We also did not find any significant differences in drug use characteristics or demographics between those who completed treatment versus those who dropped out. The main outcome in the primary paper by Ling et al. (26), and in the missing data analysis by McPherson et al. (2), was the percentage of opioid negative (UA-) urine specimens across the two groups at the completion of their respective taper. As the primary outcome paper describes, follow-up data was collected at 1-month and 3-months after the treatment period. Consistent with most previous SUD RCTs, we think it inappropriate to combine the follow-up analysis with the treatment period analysis. However, many of the analyses applied here could be used for follow-up analyses as well.

Analytic Strategy

We used latent growth modeling to compare three different analytic methods of modeling missing values: 1) LGM that uses ML to handle the missing data (i.e., MAR model) (8, 27, 28), 2) the Diggle-Kenward (16) MNAR model where dropout is a function of previous and concurrent time-specific UA submissions (see Figure 1), and 3) the Wu-Carroll MNAR model where dropout is a function of the intercept and linear slope growth factors (see Figure 2) (18). In standard latent growth modeling, the model parameters are estimated using ML, which uses the logit link for binary indicators, such as UA in our investigation. This strategy treats the outcome as longitudinal (i.e., similar to the commonly used method of generalized estimating equations model used in SUD RCTs), and the 'slope' of opioid use

change is what gets regressed onto trial arm and other covariates. Dropout was coded as follows for both of the MNAR models: 0 = observed or intermittent missing values, 1 = dropout. Coding dropout in such a way treats intermittent missing values as MAR and permanent dropout as MNAR. Importantly, these models do not provide direct estimates of missing values.

All models included several auxiliary variables (i.e., 4 opioid UAs obtained during the stabilization [pre-randomization] period). Auxiliary variables are covariates that are potentially correlated with the missingness or are correlated with one or more of the analysis variables. While auxiliary variables do not need to be predictors in the model, these variables should be correlates of the missing data mechanism or correlates of other analysis variables responsible for the missingness. A strong set of auxiliary variables is usually made up of a relatively small number of variables that will ideally serve both of the following functions: a) predict what the missing value would have been if it were observed; and b) predict the propensity for producing a missing value (29).

Each analysis regressed each of the binary opioid UA scores (0 = negative for opioids, 1 = positive for opioids) across the 4-week treatment period onto both an intercept and linear slope growth factor (i.e., latent variable, or random effect) using the logit link. When combined with fixed time scores, such as those presented in Figures 1 and 2, this forces all of the variance observed in the outcomes over time into either the growth factors or components of unaccounted for variability. Preliminary model fitting indicated a linear trend for the slope growth factor fit well, and that the estimation of a random effect (i.e., variance) for both the intercept and the slope was warranted. Both the intercept and linear slope growth factors were then regressed onto taper condition (7-days versus 28-days). We also reported effect size estimates that are similar to Cohen's d (30) to evaluate differences between the two taper conditions in the logit metric at the end of the trial, similar to analyses conducted on other SUD RCTs when reporting a single endpoint analysis (i.e., drug use at the end of treatment). We defined d as the model implied mean difference at the final time point (week 4) divided by the square root of the intercept variance. The mean difference between trial arms at the final time point was equal to the difference in the unit of logit scale standard deviations. We also regressed the linear slope growth factor onto the intercept growth factor given the growing body of evidence that baseline UA is often predictive of use during treatment (7, 31, 32). We also included gender and age to replicate analyses conducted on this dataset previously (2, 7, 26). Mplus 7.0 was used for all analyses (17).

Results

Maximum Likelihood (MAR) Model

The MAR model demonstrated an effect of trial arm ($B = -0.45, p < .05$) on the opioid UA linear slope growth factor such that those in the 28-day taper group were significantly less likely than the 7-day taper group to submit a UA+ over time and the mean difference at the end of the treatment period was $d = 0.202$ (see Table 1). Also, the intercept growth factor (i.e., baseline UA) was predictive of the UA linear slope ($B = -0.14, p < .05$), indicating that opioid use at baseline is associated with a decrease in UA+ submissions during the period of treatment controlling for all covariates in the model.

Diggle-Kenward (MNAR) Model

The Diggle-Kenward model demonstrated an effect of trial arm on the linear slope ($B = -0.64, p < .05$), and on the intercept growth factor ($B = 0.67, p < .05$; see Table 2, Figure 1). This indicates that those in the 28-day taper group were less likely to submit a UA+ for opioids, but they were also significantly more likely to submit a positive UA at baseline. The mean difference at the end of the trial was relatively large ($d = 0.820$). The UA intercept predicted the UA linear slope ($B = 0.15, p < .05$), but in the opposite direction compared to the MAR model. Please note that for Tables 2 and 3, the dropout regression coefficients presented in the bottom half of Table 2 are ORs, but the random effect regressions of the intercept and linear slope on covariates are presented in the top half of the table as unstandardized linear regression coefficients.

While time-specific UA did not predict either subsequent or concurrent dropout, trial arm was predictive of dropout after week 1 (odds ratio (OR) = 0.43, $p < .05$), 2 (OR = 0.41, $p < .05$), and 3 (OR = 82.01, $p < .05$). Those in the 28-day taper group were less likely to dropout at weeks 2 and 3, but they were more likely to dropout at week 4 compared to patients in the 7-day taper group. Importantly, this model did not include a week 1 dropout indicator because none of the patients dropped out at week 1.

Wu-Carroll (MNAR) Model

The Wu-Carroll MNAR model evidenced an effect of trial arm on the UA slope ($B = -0.41, p < .05$; see Table 3, Figure 2), but the effect size at the end of the treatment period was smaller ($d = 0.307$). Trial arm also predicted dropout at week 4 (OR = 71.31, $p < .05$). However, trial arm did not predict dropout at weeks 2 or 3. Baseline UA was predictive of the UA slope ($B = -0.12, p < .05$), similar in size and direction as the MAR model. The UA intercept and slope both were not significantly associated with dropout from weeks 2 through 4.

Discussion

Analyzing SUD RCTs with a substantial amount of missing values is a ubiquitous problem throughout the substance use treatment literature (1, 2). The association of trial arm, baseline UA, and other covariate effects changed in a meaningful way across the growth models, indicating that the missing data mechanism is critical to understand and explain when reporting on such trials. Sensitivity analyses are an active and much needed area of increased research (12, 13), in part because there are currently no objective criteria for comparing these models. However, we offer our tentative interpretation of this example set of sensitivity analyses below.

In our analysis, one of the primary parameters of the UA slope regressed onto trial arm was consistently significant, the effect size varied from $d = 0.202$ in the MAR model (i.e., small (30)) to $d = 0.307$ in the Wu-Carroll model, and $d = 0.820$ in the Diggle-Kenward model (i.e., large effect size (30)). This variance in effect size is important to note given that 1) RCTs often form the basis of evidence-based treatment modalities, and 2) effect sizes are commonly included in quantitative reviews. We also note the changing direction from model

to model when evaluating the regression path from baseline opioid UA to opioid UA specimens submitted during the treatment period (i.e., prediction of UA slope by the UA intercept, or baseline value). For the treatment effect, this provides an innovative ‘confidence interval’ for understanding the stability of the estimate, similar to how such sensitivity analyses have been reported elsewhere (10, 11).

It is critical to consider whether or not the additional assumptions associated with each MNAR and MAR model are reasonable (1, 8, 19). For example, the Diggle-Kenward model assumes that the logistic regression portion of the model is correctly specified (e.g., no omitted regression paths, no extra paths), and that the repeated outcome adheres to a strictly normal distribution. Given that the outcome in this example is binary, the normality assumption is associated with the underlying distribution of the continuous variable and the binary indicator represents a threshold value that separates that ‘latent’ continuous variable into binary observed values. The Wu-Carroll model also makes the assumption of correct specification for the dropout portion of the model, assumes that the random effects are normally distributed, and assumes conditional independence (i.e., no association between outcomes and missing data indicators after controlling for the random effects).

Consider the trial presented here, where substance use status is provided on a weekly basis during a treatment trial. In this instance, the research team might observe a developmental trajectory of scores wherein the subjects show an increasing pattern of positive UA results over time, and these assessments “snowball” to the point where a participant feels increasingly discouraged and eventually drops out. It is the trajectory, rather than any one point in time that leads to dropout (i.e., Wu-Carroll). This makes sense for the current example because the data for this trial is during a taper period that commenced after receiving treatment for at least 5 weeks prior and a combination of treatment fatigue and discomfort experienced during the taper. The Diggle-Kenward model may be an inappropriate model because the research team must be confident that time-specific missing values are associated with the concurrent outcome assessment. Another important aspect of our example is that key parameters in the MAR model and the Wu-Carroll (i.e., effect of trial arm and UA intercept on the UA slope; see Tables 1 and 3) were very similar in terms of significance and magnitude. The results from the Diggle-Kenward model are less consistent with both of these. Consistency across multiple sensitivity analyses will often lend weight to results because the results appear more ‘stable’.

Overall, a prudent approach would be to describe and present as many models as is reasonable given space limitations, or as supplemental online materials, while noting their similarity on key parameters in order to help make one’s case that the key findings are in fact what the research team says they are in the presence of a high percentage of missing values. Please note that this is a tentative, example interpretation. Ideally, a clinical interpretation offered after completion of a series of sensitivity analyses would be done by the original research team (e.g., McDonnell and colleagues(10); Roll and colleagues(11)). This investigation also highlights the potential for these modern approaches to missing data to shed new light on outcomes of interest (e.g., time-specific dropout) other than the primary outcome of UA or reported days of use. We have found such sensitivity analyses extremely

helpful when examining treatment effects in other, similar substance use disorder clinical trials (10, 11).

This investigation has some notable limitations. One limitation is the strict distributional assumptions of the Diggle-Kenward and Wu-Carroll selection model (6, 8, 16) noted above. This potentially limits the utility of these models, but it is nevertheless an important model to consider given its intuitive appeal for modeling time-specific dropout. Another limitation of this manuscript is our exclusion of pattern-mixture modeling (PMM) strategies for comparison in our sensitivity analyses. PMM methods can control for the missing data and use MAR methods such as ML or MI to stratify the analysis across multiple missing data patterns (or sub-groups, ‘classes’) in an effort to more accurately model the data and control for the data being MNAR (5, 6, 8, 34). Our approaches presented above focus on dropout and assume that intermittent missing data is MAR, while PMM can model different types of missing data patterns, including intermittent missing data and dropout. Given the rapidly expanding number and types of pattern-mixture models being reviewed and discussed in the literature (e.g., (5)), we felt that this investigation would be better served by choosing a select few to demonstrate how key parameters can vary across the missing data strategy employed. Lastly, we would normally control for potential ‘site effects’ by including site as a fixed or random effect, but those data were not available for this clinical trial due to the Clinical Trial Network’s policy of remaining consistent with HIPAA regulations governing the release of data to the public.

In conclusion, our hope is that the above study demonstrates for clinical research teams the importance of performing sensitivity analyses and thoroughly evaluating what the most defensible model is when reporting their unique missing data situation for a clinical trial. While there is still going to be a level of uncertainty with any model reported, the goal should be to report the most likely model from multiple models with varying assumptions, and defend that choice using a logical argument.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This project was supported by grants from the Department of Justice, the Life Science Discovery Fund (Roll, PI) and a grant to the Clinical Trials network Pacific Northwest Node (award number 5 U10 DA013714-10) from the National Institute on Drug Abuse (NIDA; Donovan and Roll, Co-PIs). In addition, this project was supported by funds from the Pilot Study Support Program as part of the Center for Advancing Longitudinal Drug Abuse Research (CALDAR, award number P30DA016383; McPherson, PI) from the National Institute on Drug Abuse (NIDA), and the Washington State University Spokane Seed Grant Program (McPherson, PI).

Drs. McPherson, Barbosa-Leiker, and Roll have received research funding from the Bristol-Myers Squibb Foundation. This funding is in no way related to the investigation reported here.

References

1. 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]

2. McPherson S, Barbosa-Leiker C, Burns GL, Howell D, Roll J. Missing data in substance abuse treatment research: current methods and modern approaches. *Experimental and clinical psychopharmacology*. 2012; 20:243–250. [PubMed: 22329556]
3. Hedden SL, Woolson RF, Malcolm RJ. A comparison of missing data methods for hypothesis tests of the treatment effect in substance abuse clinical trials: a Monte-Carlo simulation study. *Substance Abuse Treatment Prevention and Policy*. 2008; 3
4. Yang X, Shoptaw S. Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. *Drug and Alcohol Dependence*. 2005; 77:213–225. [PubMed: 15734221]
5. Muthen B, Asparouhov T, Hunter AM, Leuchter AF. Growth modeling with nonignorable dropout: alternative analyses of the STAR*D antidepressant trial. *Psychological methods*. 2011; 16:17–33. [PubMed: 21381817]
6. Enders CK. Missing not at random models for latent growth curve analyses. *Psychological methods*. 2011; 16:1–16. [PubMed: 21381816]
7. McPherson S, Barbosa-Leiker C, McDonell MG, Howell D, Roll J. Longitudinal missing data strategies for substance use clinical trials using generalized estimating equations: An example with a buprenorphine trial. *Human Psychopharmacology: Clinical and Experimental*. (In press.).
8. Enders, CK. *Applied Missing Data Analysis*. New York: Guilford Press; 2010.
9. Muthén BO. Latent variable modeling in heterogeneous populations. *Psychometrika*. 1989; 54:557–585.
10. McDonell MG, Srebnik D, Angelo F. Randomized controlled trial of contingency management for stimulant use in community mental health patients with serious mental illness. *Am J Psychiatry*. 2013; 170:94–101. [PubMed: 23138961]
11. Roll JM, Chudzynski J, Cameron JM, Howell DN, McPherson S. Duration effects in contingency management treatment of methamphetamine disorders. *Addict Behav*. 2013; 38:2455–2462. [PubMed: 23708468]
12. Ware JH, Harrington D, Hunter DJ, D'Agostino RB. Missing data. *N Engl J Med*. 2012; 367:1353–1354.
13. Little RJ, D'Agostino R, Cohen ML. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012; 367:1355–1360. [PubMed: 23034025]
14. Graham JW. Missing data analysis: making it work in the real world. *Annual review of psychology*. 2009; 60:549–576.
15. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological methods*. 2002; 7:147–177. [PubMed: 12090408]
16. Diggle P, Kenward MG. Informative Drop-out in Longitudinal Data-Analysis. *Applied Statistics-Journal of the Royal Statistical Society Series C*. 1994; 43:49–93.
17. Muthen, LK.; Muthen, B. *Mplus User's Guide*. Los Angeles, CA: Muthen & Muthen; 1998–2010.
18. 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.
19. Enders CK. Missing not at random models for latent growth curve analyses. *Psychol Methods*. 2011; 16:1–16. [PubMed: 21381816]
20. Muthen B, Asparouhov T, Hunter AM, Leuchter AF. Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychol Methods*. 2011; 16:17–33. [PubMed: 21381817]
21. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics*. 1995; 51:151–168. [PubMed: 7766771]
22. Albert PS, Follmann DA, Wang SA, Suh EB. A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics*. 2002; 58:631–642. [PubMed: 12229998]
23. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*. 1989; 45:939–955. [PubMed: 2486189]
24. Ten Have TR, Kunselman AR, Pulkstenis EP, Landis JR. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*. 1998; 54:367–383. [PubMed: 9544529]

25. Pulkstenis EP, Ten Have TR, Landis JR. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*. 1998; 93:438–450.
26. Ling W, Hillhouse M, Domier C. Buprenorphine tapering schedule and illicit opioid use. *Addiction*. 2009; 104:256–265. [PubMed: 19149822]
27. Enders CK, Peugh JL. Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling-a Multidisciplinary Journal*. 2004; 11:1–19.
28. Muthen B, Kaplan D, Hollis M. On Structural Equation Modeling with Data That Are Not Missing Completely at Random. *Psychometrika*. 1987; 52:431–462.
29. Acock AC. Working with missing values. *Journal of Marriage and Family*. 2005; 67:1012–1028.
30. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum; 1988.
31. McDonell MG, Srebnik D, Angelo F. Randomized Controlled Trial of Contingency Management for Stimulant Use in Community Mental Health Patients With Serious Mental Illness. *The American journal of psychiatry*. 2012
32. Stitzer ML, Peirce J, Petry NM. Abstinence-based incentives in methadone maintenance: interaction with intake stimulant test results. *Exp Clin Psychopharmacol*. 2007; 15:344–350. [PubMed: 17696681]
33. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials*. 2004; 1:368–376. [PubMed: 16279275]
34. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*. 1997; 2:64–78.

Figure 1. Estimated Diggle and Kenward (1994) random linear growth model of time-specific dropout predicting both subsequent and concurrent dropout with additional covariates effects. (1) and (2) indicates regression paths that were held equal. Note: The short arrows near the ovals represent residual (unaccounted) variance in the growth factors.

Figure 2.

Estimated Wu and Carroll (1988) random linear growth model of intercept and slope growth factors predicting dropout with additional covariates effects. (1) and (2) indicates regression paths that we held equal. Note: The short arrows near the ovals represent residual (unaccounted) variance in the growth factors.

Table 1

Random Linear Growth Model of UAs Across Baseline and 4 Subsequent Weekly Visits During Treatment: Missing at Random (using Maximum Likelihood)

Covariates	Outcomes	Unstandardized Estimate	(95% CI)
Trial Arm →	UA Intercept	0.944 ^{ns}	(0.289 to 2.178)
Age (years) →	UA Intercept	−0.046 ^{ns}	(−0.127 to 0.035)
Sex →	UA Intercept	−0.234 ^{ns}	(−1.248 to 0.781)
Trial Arm →	UA Slope	−0.454 [*]	(−0.678 to −0.229)
Age (years) →	UA Slope	−0.001 ^{ns}	(−0.012 to 0.010)
Sex →	UA Slope	0.330 [*]	(0.083 to 0.576)
UA Intercept →	UA Slope	−0.140 [*]	(−0.214 to −0.066)

UA = opioid urine analysis. The dependent measure was a negative or positive across the 4-week treatment period (0 = negative UA; 1 = positive UA). Trial Arm represents the 7- and 28-day taper groups (0 = 7-day; 1 = 28-day). Sex represents male and female groups (0 = males; 1 = female). ns = non-significant.

*
p < .05.

Table 2

Growth Model of UAs Across Baseline and 4 Subsequent Weekly Visits: Missing not at Random, Diggle and Kenward (1994) Selection Model

<i>Covariates</i>	<i>Outcomes</i>	<i>Unstandardized Estimate</i>	<i>(95% CI)</i>
Trial Arm →	UA Intercept	0.677*	(0.167 to 1.188)
Age (years) →	UA Intercept	−0.022*	(−0.044 to 0.000)
Sex →	UA Intercept	−0.072 ^{ns}	(−0.683 to 0.540)
Trial Arm →	UA Slope	−0.637*	(−0.921 to −0.352)
Age (years) →	UA Slope	−0.001 ^{ns}	(−0.014 to 0.011)
Sex →	UA Slope	0.339*	(0.018 to 0.660)
UA Intercept →	UA Slope	0.149*	(0.064 to 0.235)
<i>Covariates</i>	<i>Outcomes</i>	<i>Odds Ratio</i>	<i>(95% CI)</i>
Trial Arm →	Dropout Week 2	0.432*	(0.231 to 0.806)
Age (years) →	Dropout Week 2	0.982 ^{ns}	(0.954 to 1.012)
Sex →	Dropout Week 2	0.481 ^{ns}	(0.230 to 1.005)
Trial Arm →	Dropout Week 3	0.410*	(0.238 to 0.677)
Age (years) →	Dropout Week 3	1.006 ^{ns}	(0.985 to 1.027)
Sex →	Dropout Week 3	0.909 ^{ns}	(0.538 to 1.536)
Trial Arm →	Dropout Week 4	82.013*	(39.915 to 168.510)
Age (years) →	Dropout Week 4	1.015 ^{ns}	(0.985 to 1.046)
Sex →	Dropout Week 4	1.015 ^{ns}	(0.518 to 1.989)
UA Weeks 1–3 →	Dropout Weeks 2–4	1.367 ^{ns}	(0.683 to 2.734)
UA Weeks 2–4 →	Dropout Weeks 2–4	0.850 ^{ns}	(0.279 to 2.586)

UA = opioid urine analysis. The dependent measure was a negative or positive across the 4-week treatment period (0 = negative UA; 1 = positive UA). Trial Arm represents the 7- and 28-day taper groups (0 = 7-day; 1 = 28-day). Sex represents male and female groups (0 = males; 1 = female). ns = non-significant.

* p < .05.

Table 3

Growth Model of UAs Across Baseline and 4 Subsequent Weekly Visits: Missing not at Random, Wu and Carroll (1988) Selection Model

<i>Covariates</i>	<i>Outcomes</i>	<i>Unstandardized Estimate</i>	<i>(95% CI)</i>
Trial Arm →	UA Intercept	0.502 ^{ns}	(−0.346 to 1.351)
Age (years) →	UA Intercept	−0.059 [*]	(−0.115 to −0.002)
Sex →	UA Intercept	−0.287 ^{ns}	(−1.208 to 0.633)
Trial Arm →	UA Slope	−0.413 [*]	(−0.721 to −0.105)
Age (years) →	UA Slope	−0.001 ^{ns}	(−0.011 to 0.009)
Sex →	UA Slope	0.318 [*]	(0.096 to 0.540)
UA Intercept →	UA Slope	−0.121 [*]	(−0.210 to −0.032)
<i>Covariates</i>	<i>Outcomes</i>	<i>Odds Ratio</i>	<i>(95% CI)</i>
Trial Arm →	Dropout Week 2	0.382 ^{ns}	(0.118 to 1.013)
Age (years) →	Dropout Week 2	0.984 ^{ns}	(0.955 to 1.013)
Sex →	Dropout Week 2	0.531 ^{ns}	(0.172 to 1.645)
Trial Arm →	Dropout Week 3	0.355 ^{ns}	(0.105 to 1.198)
Age (years) →	Dropout Week 3	1.006 ^{ns}	(0.985 to 1.028)
Sex →	Dropout Week 3	1.006 ^{ns}	(0.384 to 2.635)
Trial Arm →	Dropout Week 4	71.307 [*]	(25.374 to 200.394)
Age (years) →	Dropout Week 4	1.017 ^{ns}	(0.986 to 1.048)
Sex →	Dropout Week 4	1.135 ^{ns}	(0.373 to 3.452)
UA Intercept →	Dropout Weeks 2–4	1.002 ^{ns}	(0.734 to 1.367)
UA Slope →	Dropout Weeks 2–4	0.759 ^{ns}	(0.055 to 10.483)

UA = opioid urine analysis. The dependent measure was a negative or positive across the 4-week treatment period (0 = negative UA; 1 = positive UA). Trial Arm represents the 7- and 28-day taper groups (0 = 7-day; 1 = 28-day). Sex represents male and female groups (0 = males; 1 = female). ns = non-significant.

*
p < .05.