

RESEARCH ARTICLE

Open Access



Analysis of self-report and biochemically verified tobacco abstinence outcomes with missing data: a sensitivity analysis using two-stage imputation

Yiwen Zhang¹, Xianghua Luo^{2,3*} , Chap T. Le^{2,3}, Jasjit S. Ahluwalia⁴ and Janet L. Thomas⁵

Abstract

Background: Missing data are common in tobacco studies. It is well known that from the observed data alone, it is impossible to distinguish between missing mechanisms such as missing at random (MAR) and missing not at random (MNAR). In this paper, we propose a sensitivity analysis method to accommodate different missing mechanisms in cessation outcomes determined by self-report and urine validation results.

Methods: We propose a two-stage imputation procedure, allowing survey and urine data to be missing under different mechanisms. The motivating data were from a tobacco cessation trial examining the effects of the extended vs. standard Quit and Win contests and counseling vs. no counseling under a 2-by-2 factorial design. The primary outcome was 6-month biochemically verified tobacco abstinence.

Results: Our proposed method covers a wide spectrum of missing scenarios, including the widely adopted “missing = smoking” imputation by assuming a perfect smoking-missing correlation (an extreme case of MNAR), the MAR case by assuming a zero smoking-missing correlation, and many more in between. The analysis of the data example shows that the estimated effects of the studied interventions are sensitive to the different missing assumptions on the survey and urine data.

Conclusions: Sensitivity analysis has played a crucial role in assessing the robustness of the findings in clinical trials with missing data. The proposed method provides an effective tool for analyzing missing data introduced at two different stages of outcome assessment, the self-report and validation time. Our methods are applicable to trials studying biochemically verified abstinence from alcohol and other substances.

Keywords: Abstinence outcome, Imputation, Missing data, Sensitivity analysis

Background

Cigarette smoking is a risk factor for morbidity and mortality in the US and around the world [1–4]. Smoking cessation studies usually encourage cessation and provide either behavioral (e.g., counseling) or pharmaceutical (e.g., nicotine gum) interventions or both. In smoking cessation studies, missing binary abstinence outcomes (i.e., quit or not quit) are very common. These missing outcomes may

lead to bias or weakened statistical power in estimating the effect of the studied intervention. Choosing appropriate statistical methods to handle binary missing data has been a continuing source of controversy [5].

The choice of methods to deal with missing data would depend on assumptions about the missing mechanism [6]. Data are referred to as being missing at random (MAR) if the missing status (yes or no) is not related to the missing value itself, but can be dependent on some other observed variables. Data are referred to as being missing not at random (MNAR) or nonignorable missing if the probability of missing depends on the missing value. It is well known that from the observed data alone, it is impossible to

* Correspondence: luox0054@umn.edu

²School of Public Health, Division of Biostatistics, University of Minnesota, 420 Delaware St. SE, MMC 303, Minneapolis, MN 55455, USA

³University of Minnesota Masonic Cancer Center, Minneapolis, MN 55455, USA

Full list of author information is available at the end of the article



distinguish between MAR and MNAR. Therefore, statistical analyses based on one specific missing mechanism, such as the popular MAR assumption, could lead to misleading conclusions if it turns out that the missing is not at random. For example, consider a trial to incentivize smokers to quit smoking. Missing data due to a non-response in surveys or as a result of dropouts could be dependent on the smoking status of the participants, which renders the missing mechanism to be not at random. Sensitivity analysis can play a crucial role in assessing the robustness of the findings in clinical trials with missing data [7]. In this paper, we aim to study sensitivity analysis methods for analyzing smoking cessation outcome data under various missing data mechanisms including MAR and MNAR.

In the literature, the standard procedure used in smoking cessation trials is to assume that all non-respondents are smoking (referred to as the “missing = smoking” method hereinafter), which is a special case of single imputation under the MNAR assumption. Based on Jackson et al. [8], around 80% of reports of smoking cessation trials adopt this assumption. However, this simple imputation approach has been shown to lead to potentially biased results [8–12]. Other common single imputation methods frequently used in smoking cessation trials include the last observation carried forward (LOCF), the baseline observation carried forward (BOCF) and imputations based on predicted values from a regression model or the expectation-maximization (EM) algorithm [13]. In addition, Barnes and colleagues [14] used a multiple imputation procedure with the propensity score matching method to impute missing smoking status.

Hedeker and colleagues [12] demonstrated that the simple missing = smoking imputation is essentially based on the assumption that the missing status and the smoking status have a perfect correlation ($r = 1$) or, equivalently, the odds ratio (OR) between missing and smoking equals positive infinity ($OR = +\infty$). They developed both simple and multiple imputation approaches based on more relaxed assumptions which allow different levels of correlation between smoking and missing status. Although their imputation method provides a more flexible and useful alternative to the simple missing = smoking method and has been applied in various trials [8, 13], this method cannot be directly applied to data with missing values generated from multiple sources or stages. For example, when cessation outcome is determined by self-report survey data followed by a urine validation test, both a non-respondent survey and a missing urine sample can lead to missing cessation outcomes. In this case, an imputation procedure designed to account for missing data generated from two different stages, the survey collection stage and the urine collection stage, would be preferable. A naive imputation approach for dealing with this type of missing data would

treat only subjects who confirmed their abstinence by both self-report and a urine sample with a negative result as a treatment success (i.e., biochemically verified self-reported abstinence). All the other subjects including those who either failed to complete the survey or failed to provide the urine sample for confirmation of self-reported abstinence would be considered as treatment failures (i.e., not achieving abstinence). Note that this naive approach is an extreme case under the MNAR assumption, assuming a perfect correlation between survey missing and self-report failure and a perfect correlation between urine missing and urine-verified failure among people who self-reported abstinence. Hence, it does not have the flexibility to accommodate different levels of correlation between survey missing and self-report failure or between urine missing and urine-verified failure. In this paper, we extend Hedeker et al.’s method [12] to a two-stage imputation procedure to take into account missing data in either the self-report or the urine verification stages.

The rest of the article is organized as follows. In Section 2, we first introduce a randomized controlled trial of college smokers [15] which motivated this research, and then, we introduce a sensitivity analysis method using a two-stage imputation procedure for missing abstinence data at the self-reporting and subsequent biochemical verification stages. In Section 3, we report the sensitivity analysis result of the college smokers study. Some discussions and concluding remarks can be found in Sections 4 and 5, respectively.

Methods

Aim, design, and setting

The data motivating this research were collected from 1217 subjects enrolled in a smoking cessation randomized clinical trial entitled “Enhanced quit and win contests to improve smoking cessation among college students” (henceforth referred to as the “Enhanced Quit & Win” study) during the academic years 2010–2013. This study utilized a two-by-two factorial design to examine the marginal effect of two distinct interventions: the impact of multiple vs. single Quit & Win contests and the effect of the Motivational and Problem Solving counseling (MAPS) counseling vs. no counseling on smoking cessation among college smokers. Specifically, participants were randomly assigned to one of four groups: (1) single contest (denoted by Tx1, $n = 306$), (2) single contest plus counseling (Tx2, $n = 296$), (3) multiple contests (Tx3, $n = 309$), and (4) multiple contests plus counseling (Tx4, $n = 306$). The primary cessation outcome was measured at 6 months post-randomization when all participants were encouraged to complete an online survey to report their smoking status and other tobacco use in the past 30 days. Only people who reported no tobacco use in the past 30 days

were invited to provide urine to biochemically (cotinine assay) confirm their self-reported abstinence. Both self-reported abstinence and biochemically verified abstinence were of interest. The study design and the characteristics of participants are described in greater detail in the parent study manuscript [15]. This trial was registered at [ClinicalTrials.gov](https://clinicaltrials.gov) as number NCT01096108.

Sensitivity analysis using two-stage imputation

As we described earlier, the missing data in the Enhanced Quit & Win study occurred at two different stages: the survey collection stage and the urine verification stage. A common and conservative imputation approach for dealing with such two-stage missing data would treat only subjects who self-reported abstinence and provided a urine sample which confirmed the abstinence as a treatment success (i.e., biochemically verified abstinence). All the other subjects including those who either failed to complete the survey or failed to provide urine would be considered as a treatment failure. This is analogous to the missing = smoking method for one-stage missing data. Note that this approach is an extreme case of the single imputation approach under the not missing at random (MNAR) assumption, assuming a perfect correlation between the survey missing and self-report failure, or equivalently an infinite odds ratio between the two (denoted by $OR_1 = \infty$), and at the same time a perfect correlation between the urine missing and urine-verified failure (denoted by $OR_2 = \infty$).

In this paper, we propose a two-stage imputation approach under the MNAR assumption, which takes into account the two-stage missing process and allows (1) different levels of correlation between the survey missing and self-report failure (i.e., varying OR_1) and (2) different levels of correlation between the urine missing and urine-verified failure among those who self-reported abstinence (i.e., varying OR_2). This can be considered as an extension of the imputation method in Hedeker et al. [12] for one-stage missing data to a two-stage missing data situation. In this section, we present a two-stage imputation approach conducted on a summary or aggregated data basis.

The one-stage imputation method by Hedeker et al. [12] for the self-report data

We first introduce some general notation. We code “tobacco use status”, the binary dependent variable as 1 = used tobacco/failure and 0 = did not use tobacco/abstinence and “missing status”, the binary indicator of whether the data is missing or not as 1 = missing and 0 = observed. Let $j = 1, 2, 3, 4$ index the four treatment groups, Tx1 to Tx4, respectively. Let subjects be indexed by $i = 1, 2, \dots, n_j$ where n_j denotes the total number of subjects in treatment group Tx_j. Since we propose to

perform imputations within each treatment, in the sequel we omit j from all symbols to simplify notation. Moreover, we use superscripts 11, 12, 21, and 22 to denote the four entries of the two-by-two table between the tobacco use status and missing status, as illustrated in Table 1. Note that, in the second row of Table 1, only the total number of individuals with missing data, n^2 , can be observed; the abstinence statuses of these people, n^{21} and n^{22} (the second row in Table 1) are unknown and need to be estimated. Furthermore, in the summation row of Table 1, the total number of abstinence (denoted by n^1) and the total number of failure (denoted by n^2) are also unknown. Note that the ‘dot’ in the superscripts indicates summation over a row or column.

Following Hedeker et al. [12], in order to impute the numbers for abstinence and failure for participants with missing survey data (n^{21} and n^{22}), we will assume an odds ratio for the missing survey status and self-report tobacco use status (OR_1) to reflect the strength of correlation between them (denoted by r_1). Note that the widely adopted missing = smoking method corresponds to the situation of $r_1 = 1$ or $OR_1 = \infty$. In that case, n^{21} is imputed with 0 and n^{22} is imputed with n^2 . More generally, we have

$$OR_1 = \frac{(n^{22}/n^{21})}{(n^{12}/n^{11})}, \text{ or equivalently } \frac{n^{22}}{n^{21}} = OR_1 \frac{n^{12}}{n^{11}}, \quad (1)$$

and then it can be shown that the unobserved values, n^{21} and n^{22} can be imputed with the assumed OR_1 by:

$$n^{22} = n^2 \frac{OR_1 * Odds}{1 + OR_1 * Odds} = \pi n^2 \text{ and } n^{21} = n^2 - n^{22},$$

where *Odds* is the odds of tobacco using among survey respondents and can be calculated from the observed survey data by $\frac{n^{12}}{n^{11}}$; and $\pi = \frac{OR_1 * Odds}{1 + OR_1 * Odds}$ is a multiplicative factor relating n^{22} to n^2 .

Participants who do not respond or are lost to follow-up in a smoking cessation study may differ from those who are retained in the study with regard to their smoking status. We often expect that the odds of tobacco use among non-respondents is equal to or higher than that of respondents (i.e., $OR_1 \geq 1$), especially in studies where people are incentivized to quit as in the Enhanced Quit & Win

Table 1 Two-by-two table of tobacco use status by missing for self-report data

Missing status of self-report data	Self-report tobacco use status		
	Abstinence	Failure	Total
Observed	n^{11}	n^{12}	n^1
Missing	n^{21}	n^{22}	n^2
Total	n^1	n^2	n

Bolded entries indicate values that are not observable

study. Note that a larger OR_1 would imply a stronger relationship between missing and tobacco use.

The two-stage imputation method for the urine-verified data

When estimating biochemically verified abstinence, more complex conditions should be considered since missing data can be present at both the survey and the urine verification stage. Without specification, the notation for the survey data are the same as those defined previously (see Table 1 and the top half of Fig. 1). Some additional notations, shown in the lower half of Fig. 1, specific to the urine data are defined as follows. Let $u^{(obs)}$ and $u^{(imp)}$ denote the number of urine samples provided by people who self-reported abstinence (n^{11}) and the estimated number of urine samples that could be collected from people who would report abstinence

if they did not fail to respond to the survey (n^{21}), respectively; similarly, $v^{(obs)}$ and $v^{(imp)}$ are used for the number of missing urine samples of n^{11} and n^{21} , respectively. For the urine-verified abstinence outcome, similar notation is defined as for the self-report abstinence outcome except for using f , instead of n . The superscript 11, 12, 21, and 22 have the same meaning as those for n . In addition, we use $f^{11(obs)}$ to denote the number of urine-verified abstinence and $f^{12(obs)}$ the number of urine-verified failures obtained from people who actually provided urine samples, and we have $u^{(obs)} = f^{11(obs)} + f^{12(obs)}$. Similarly, we use $f^{11(imp)}$ to denote the number of urine-verified abstinence and $f^{12(imp)}$ the number of urine-verified failures obtained from the estimated available urine samples, and we have $u^{(imp)} = f^{11(imp)} + f^{12(imp)}$. Then we combine $f^{11(obs)}$ and $f^{11(imp)}$ to obtain the total number of

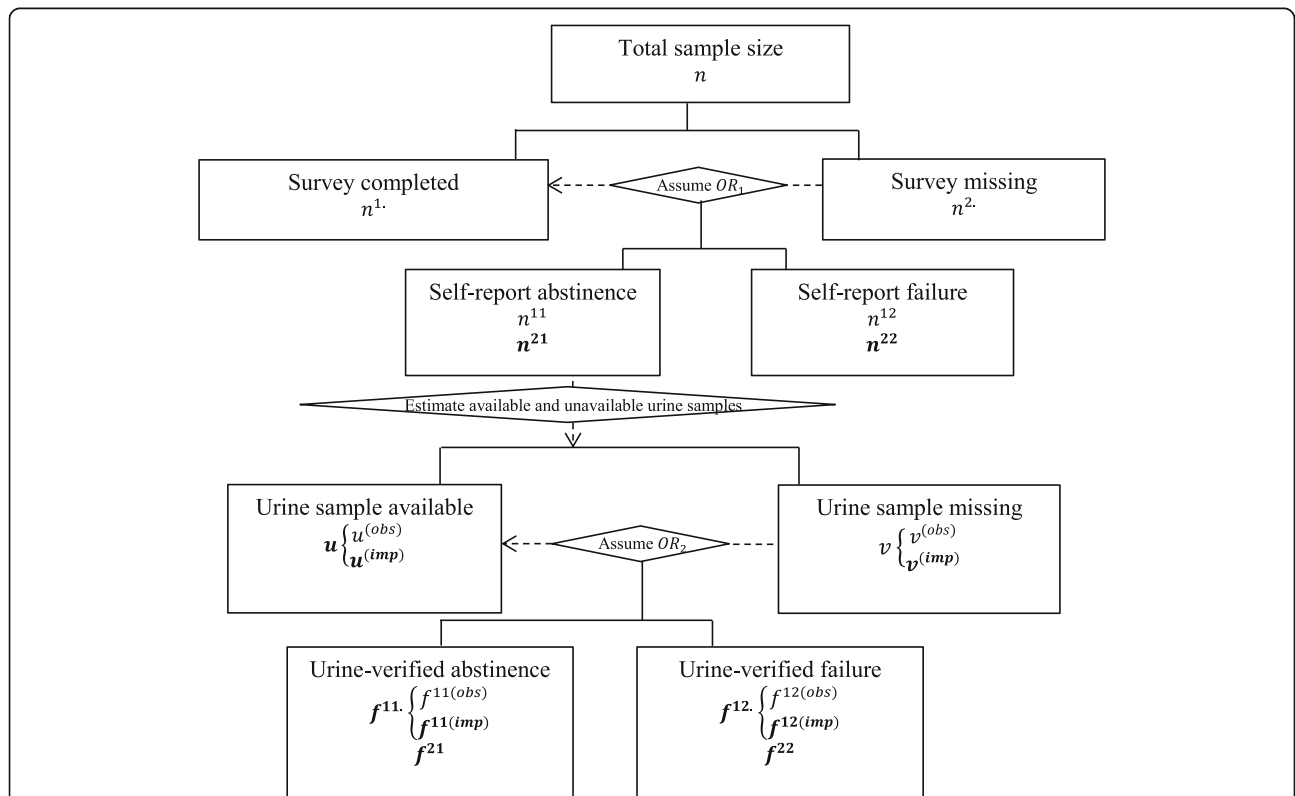


Fig. 1 Data structure and notation for a single treatment group. Note n is the total sample size, n^1 is the number of survey respondents, and n^2 is the number of survey non-respondents. Then among survey respondents, denote n^{11} as the number of observed self-report abstinence and n^{21} as the number of imputed self-report abstinence. Similarly, n^{12} and n^{22} represent the number of observed failures and imputed failures based on the self-report data, respectively. For the urine samples, $u^{(obs)}$ and $u^{(imp)}$ represent the number of observed and estimated (based on the imputed survey data) urine samples being provided; similar notations, $v^{(obs)}$ and $v^{(imp)}$ are used for the number of unavailable urine samples. For the urine data, analogous notations are defined as for the survey data except for using f , to denote the numbers of subjects under different conditions (with the superscript 11, 12, 21, and 22 having the same meaning). In addition, we used $f^{11(imp)}$ to denote the abstinence and $f^{12(imp)}$ to denote the failure obtained from the estimated available urine samples $u^{(imp)}$. Then we combined the $f^{11(obs)}$ and $f^{11(imp)}$ to obtain the number of urine-verified abstinence f^{11} ; among the urine samples what were actually provided or could have been provided if all surveys were completed, whereas combined $f^{12(obs)}$ and $f^{12(imp)}$ to obtain the urine-verified failure f^{12} . Denote OR_1 as the assumed odds ratio between missing and smoking for self-report data and OR_2 for urine data. Dashed lines indicate where missing data are reallocated based on certain assumptions or estimations. Bolded notation denotes values that are not observed

participants with urine-verified abstinence f^{A1} , among the urine samples what were actually provided or could have been provided if there were no surveys missing; similarly, we combine $f^{A2(obs)}$ and $f^{A2(imp)}$ to obtain the total number of urine verified failures f^{A2} .

Based on the previous imputation results for missing data at the survey stage, self-report abstinence (n^{11} , n^{21}) and failed abstinence (n^{12} , n^{22}) have been generated based on the imputed survey data under the assumed OR_1 within each treatment. Next, we proceed to estimate urine-verified abstinence or failure under the assumed OR_2 for the imputed, “complete” self-report data. Prior to imputing missing urine sample data, the numbers of subjects who would have provided urine samples ($u^{(imp)}$) or would not provide urine samples ($v^{(imp)}$) among survey non-respondents need to be estimated. One can assume that the urine missing rate among survey non-respondents, compared with respondents, varies by a known factor λ ($\lambda > 0$), that is

$$\frac{u^{(imp)}}{v^{(imp)}} = \lambda \frac{u^{(obs)}}{v^{(obs)}} \tag{2}$$

Consequently, the number of available ($u^{(imp)}$) and unavailable ($v^{(imp)}$) urine samples among imputed self-report abstinence cases can be calculated based on Equation (2) and the fact that $u^{(imp)} + v^{(imp)} = n^{21}$.

Similarly, one can assume that, compared to the actually provided urine samples, the urine-verified abstinence rate among the urine samples that could have been provided if the survey were completed, varies by a known factor η ($\eta > 0$):

$$\frac{f^{11(imp)}}{f^{12(imp)}} = \eta \frac{f^{11(obs)}}{f^{12(obs)}} \tag{3}$$

Therefore, the number of urine-verified abstinence ($f^{A1(imp)}$) and failure ($f^{A2(imp)}$) among imputed self-report abstinence cases people can be estimated based on Equation (3) and the fact that $f^{A1(imp)} + f^{A2(imp)} = u^{(imp)}$.

We then can calculate the total number of urine-verified abstinence cases by $f^{A1} = f^{A1(obs)} + f^{A1(imp)}$ and the urine-verified failure by $f^{A2} = f^{A2(obs)} + f^{A2(imp)}$ among all the “available” urine samples (including actually observed or imputed). Up to this point, the urine-verified abstinence (f^{21}) and urine-verified failure (f^{22}), among people whose urine was not actually provided ($v^{(obs)}$) or would not be provided even if their survey data were completed ($v^{(imp)}$), have not yet been imputed. Next, we use the fact that $v = f^{22} + f^{21} = v^{(obs)} + v^{(imp)}$ and propose a similar imputation procedure for the urine missing data as for the survey missing data described in the previous subsection as follows:

$$\frac{f^{22}}{f^{21}} = OR_2 \frac{f^{12}}{f^{11}} \text{ or equivalently,}$$

$$f^{22} = v \frac{OR_2 * Odds'}{1 + (OR_2 * Odds')} = v\pi'$$

where the second equality follows from Equation (3), and $Odds'$ and π' are the odds of tobacco use and probability of tobacco use among people who provided urine sample, respectively. The overall number of participants with urine-verified abstinence can then be obtained by simply adding f^{11} and f^{21} , and similarly, the overall number of urine-verified failure is $f^{12} + f^{22}$. After all the above steps are completed for each treatment arm, we can estimate the various treatment effects based on the imputed data.

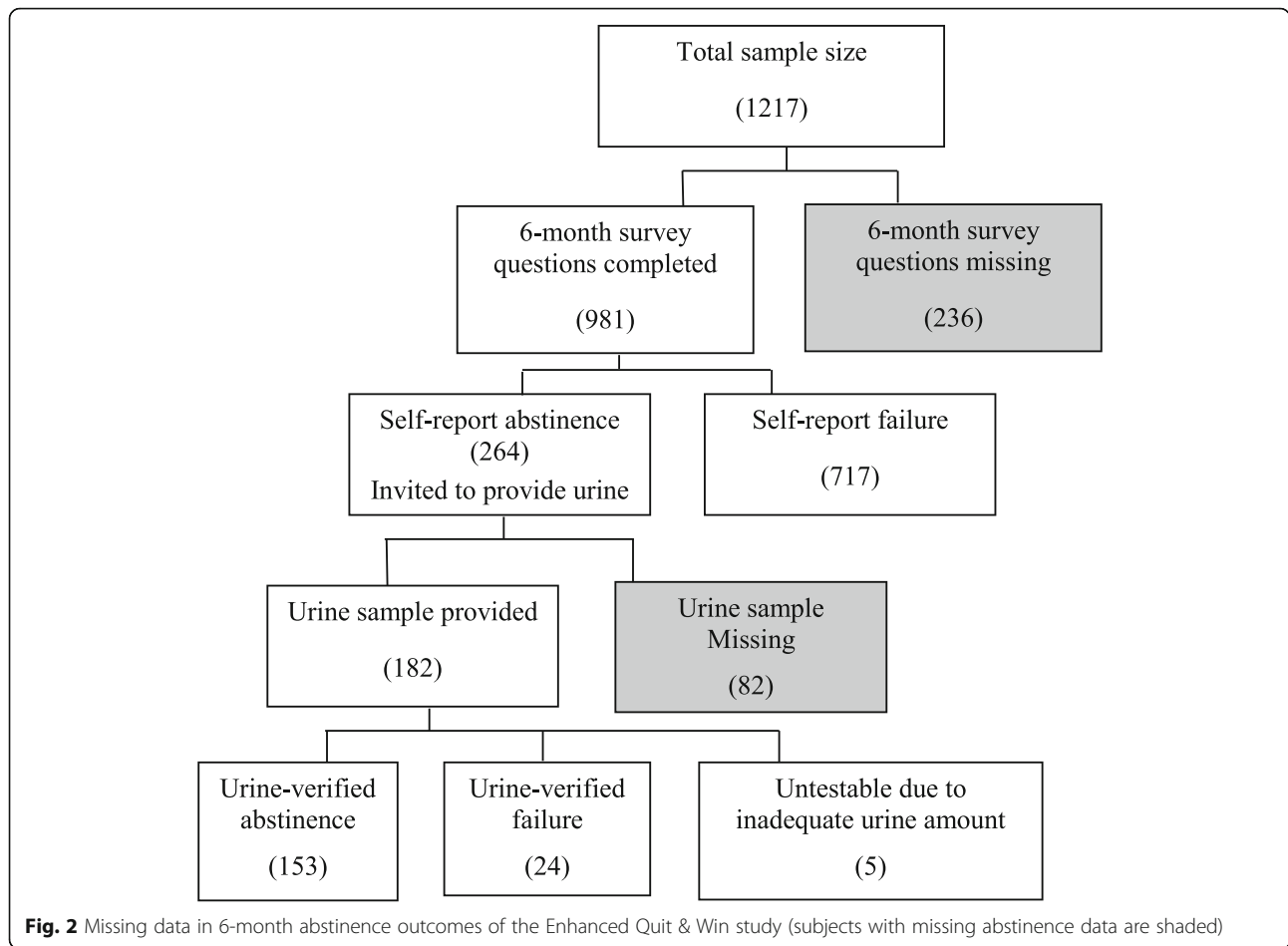
For the Enhanced Quit & Win data, we assumed a series of ≥ 1 values for OR_1 and OR_2 (1, 2, 3, 4, 5, and positive infinity) and that $\lambda = \eta = 1$ for the ease of presentation, but certainly more values can be examined for these parameters in the sensitivity analysis. SAS Version 9.4 (SAS Institute Inc., Cary, NC, USA) was used for all analyses and the SAS computing code for the proposed two-stage imputation method is provided in the Additional file 1: Supplementary Material.

Results

Summary of missing data

Figure 2 shows the summary of the 6-month abstinence outcomes and missing data. Of the 1217 randomized participants, 981 (81%) completed the 6-month survey and 236 (19%) did not. Among the 981 survey completers, 264 (27%) self-reported tobacco abstinence. Among the 264 participants who self-reported abstinence, 182 (69%) provided urine. Among the 182 participants who provided urine samples, 5 were not of adequate amount for testing and 153 (84%) were biochemically confirmed as abstinent.

Table 2 presents the differential missing data patterns across treatment arms and intervention conditions by both survey missing and urine missing. Note that the five missing urine test results due to inadequate urine amount were assumed to have the same distribution as the other 177 urine samples (86% verified abstinence and 14% verified failure) and added to the corresponding columns in Table 2. We found that the no counseling groups, Tx1 and Tx3, had significantly ($p = 0.003$) lower survey missing rates (15.4 and 16.5%, respectively and 15.9% for the combined group) than the two counseling arms, Tx2 and Tx4 (22.6 and 23.2%, respectively and 22.9% for the combined group), whereas the single- and multiple-conditions groups were found to have similar survey missing rates ($p = 0.798$). The urine missing rate was similar between the single- and multiple-conditions groups



and between the counseling and no counseling groups (both $ps > 0.05$).

Self-report abstinence outcome

The imputation results of the self-report abstinence outcome are summarized in Table 3. As a comparison, we also present the results from a complete case only analysis, where only subjects with no missing survey or urine were included. We can prove that the abstinence rate decreases as OR_1 increases. As expected, the estimated abstinence rates and treatment effect based on the imputed data under the MAR assumption (i.e., $OR_1 = 1$) are the same as those based on the complete case only analysis. However, the statistical significance is stronger (smaller p) in the former as more data are utilized. Under the MAR assumption, the estimated treatment effect of counseling vs. no counseling is significant (OR for abstinence = 1.31, $p = 0.034$); however, as OR_1 increases, the estimated treatment effect becomes less significant (all $ps > 0.05$ for $OR_1 \geq 2$), indicating that this treatment effect is sensitive to different assumed values of OR_1 . On the contrary, the estimated treatment effects of multiple vs. single contest are all close to 1.16 (all $ps > 0.05$), indicating that this

treatment effect estimation was robust to different assumed values of OR_1 . This phenomenon can be explained by the different survey missing rate between the counseling and no counseling groups, but not between the multiple and single contest groups (see the left panel in Table 2).

Urine-verified abstinence outcome

The results obtained from the imputed urine-verified abstinence data were summarized in Table 4. By considering all the combinations of OR_1 and OR_2 , each ranging from 1 to 5 and positive infinity, we found that the abstinence rate decreases as the assumed level of dependence between missing and tobacco use, OR_1 or OR_2 increases, as expected. Notice that the abstinence rates for the two studied conditions were found consistently higher than their corresponding control groups in all scenarios (i.e., the estimated treatment effect as indicated as odds ratios of abstinence are all > 1).

As shown in the upper-left corner of Table 4, significant treatment effects were estimated for the counseling group when both OR_1 and OR_2 were small. Otherwise, there seemed to be no significant treatment effects for

Table 2 Summary of 6-month self-reported and urine verified abstinence and missing data by treatment arms and by type of intervention

Treatment group	Total N	Self-report abstinence and survey missing at 6 months			Urine-verified abstinence and urine missing at 6 months		
		Self-report abstinence n (%)	Self-report failure n (%)	Survey question missing n (%)	Urine-verified abstinence n (%) ^a	Urine-verified failure n (%) ^a	Urine missing n (%) ^a
By treatment arm							
Tx1	306	65 (21.2%)	194 (63.4%)	47 (15.4%)	38 (58.5%)	6 (9.2%)	21 (32.3%)
Tx2	296	59 (19.9%)	170 (57.4%)	67 (22.6%)	34 (57.6%)	6 (10.2%)	19 (32.2%)
Tx3	309	61 (19.7%)	197 (63.8%)	51 (16.5%)	35 (57.4%)	6 (9.8%)	20 (32.8%)
Tx4	306	79 (25.8%)	156 (51.0%)	71 (23.2%)	50 (63.3%)	7 (8.9%)	22 (27.8%)
By intervention							
No counseling (Tx1 + Tx3)	615	126 (20.5%)	391 (63.6%)	98 (15.9%)	73 (57.9%)	12 (9.5%)	41 (32.5%)
Counseling (Tx2 + Tx4)	602	138 (22.9%)	326 (54.2%)	138 (22.9%)	84 (60.9%)	13 (9.4%)	41 (29.7%)
Single contest (Tx1 + Tx2)	602	124 (20.6%)	364 (60.5%)	114 (18.9%)	72 (58.1%)	12 (9.7%)	40 (32.2%)
Multiple contests (Tx3 + Tx4)	615	140 (22.8%)	353 (57.4%)	122 (19.8%)	85 (60.7%)	13 (9.3%)	42 (30.0%)
Overall	1217	264 (21.7%)	717 (58.9%)	236 (19.4%)	157 (59.5%) ^b	25 (9.5%) ^b	82 (31.1%)

Tx1: single contest + no counseling; Tx2: single contest + counseling; Tx3: multiple contests + no counseling; Tx4: multiple contest + counseling

^aPercentage out of those who self-reported abstinence

^bFive subjects whose urine samples were not of adequate amount for testing. These 5 missing urine test results were assumed to have the same distribution as the rest 177 urine samples (86% verified abstinence and 14% verified failure) and added to the two columns accordingly

Table 3 Summary of imputation results for self-report abstinence assuming different levels of association between the survey missing status and self-report abstinence

OR_1	Counseling vs. no counseling				Multiple vs. single contests			
	Counseling Tx2 + Tx4	No counseling Tx1 + Tx3	Estimated treatment effect (odds ratio for abstinence)	<i>P</i> -value	Multiple contests Tx3 + Tx4	Single contest Tx1 + Tx2	Estimated treatment effect (odds ratio for abstinence)	<i>P</i> -value
Complete case only	29.7%	24.4%	1.31	.058	28.4%	25.4%	1.16	.291
1	29.8%	24.4%	1.31	.034	28.6%	25.4%	1.18	.212
2	27.0%	22.7%	1.26	.086	26.2%	23.4%	1.16	.251
3	25.8%	22.0%	1.23	.125	25.2%	22.5%	1.16	.275
4	25.1%	21.7%	1.21	.154	24.7%	22.1%	1.15	.290
5	24.7%	21.5%	1.20	.175	24.3%	21.8%	1.15	.301
$+\infty$	22.9%	20.5%	1.15	.303	22.8%	20.6%	1.14	.359

OR_1 : odds ratio between missing and tobacco use status for self-report data, where $OR_1 = 1$ corresponds to the situation when missing is independent of tobacco use and $OR_1 =$ positive infinity ($+\infty$) corresponds to the situation when missing = smoking; Tx1: single contest + no counseling; Tx2: single contest + counseling; Tx3: multiple contests + no counseling; Tx4: multiple contest + counseling. *P*-values are based on the Chi-square test

the counseling or the multiple contests groups under different combinations of OR_1 and OR_2 . We also found that the estimated treatment effect of counseling vs. no counseling is more sensitive to the assumed level of dependence between the survey missing and self-report abstinence, but less sensitive to the assumed level of dependence between the urine missing and urine-verified abstinence. For the estimated treatment effect of the multiple- vs. single-contest, we observed no obvious pattern, no matter what values were assumed for OR_1 or OR_2 . This can be explained by the comparable survey and urine missing rates between the two contest groups as shown in Table 2. We performed additional sensitivity analysis by assuming that survey non-respondents would be less likely to provide urine than survey respondents ($\lambda = 0.5$). Results (shown in Additional file 1: Table S1) are consistent with the results reported above which are based on the equal urine missing rate assumption ($\lambda = 1$).

Discussion

In many smoking cessation studies, researchers are interested in biochemically verified abstinence (e.g., urine cotinine verified abstinence). To conserve resources, it is common to only invite people who self-report abstinence to provide biochemical samples to validate self-reported abstinence. Hence, missing data can be present at either the survey completion stage or the biochemical sampling stage. The imputation approaches presented in this paper take into account this two-stage missing data challenge and describes a two-step imputation approach allowing the survey missing and biochemical sample missing to have different missing mechanisms. Our proposed imputation approach includes both the missing = smoking imputation (an extreme case of MNAR) and the MAR imputation as special cases, hence providing a more thorough sensitivity analysis result than any simple imputation

method alone. The estimated effect of the treatments tested in the Enhanced Quit & Win study were sensitive to the different missing mechanisms depending on the differential missing data patterns across treatment arms. Although the overall results were not universally impacted, these findings demonstrate that the use of one simple imputation method alone could result in misleading conclusions regarding a treatment effect estimate.

There has been a debate regarding whether treatment should be adjusted or stratified in the imputation models. Jackson et al. [16] adjusted for treatment in their imputation model since treatment was found to be associated with the missing status and predicted missing outcomes. Alternatively, in this paper, we performed imputations stratified by treatment rather than adjusting for treatment in the model [17, 18]. Although some researchers may argue that this may overestimate the treatment effect [16], it has not been demonstrated by the preponderance of evidence. Research with more data examples to investigate the difference between these two strategies is certainly warranted.

In this paper, all the imputations were performed on aggregated data. In other words, no individual-level variation has been considered. Currently, we are working on extending the proposed imputation approach for aggregated data to take into account the uncertainty in the individual probability of tobacco use as in multiple imputations. One advantage of the imputations based on aggregated data is the ease of computing, while the multiple imputations approach is expected to give more conservative results as individual level variability is taken into account in the estimation of treatment effect. Also in this paper, we focus on the analysis of cessation outcome at a single time point. However, with repeatedly measured outcomes, longitudinal data analysis methods for dealing with missing data could be considered. [12, 19–21]. Note that our proposed

Table 4 Summary of imputation results for urine-verified abstinence assuming different levels of association between missing and abstinence

OR ₁	OR ₂	Counseling vs. no counseling				Multiple vs. single contests			
		Counseling Tx2 + Tx4	No counseling Tx1 + Tx3	Estimated treatment effect (odds ratio for abstinence)	P-value	Multiple contests Tx3 + Tx4	Single contest Tx1 + Tx2	Estimated treatment effect (odds ratio for abstinence)	P-value
1	1	25.8%	20.9%	1.31	.046	24.8%	21.8%	1.18	.212
	2	24.8%	20.1%	1.32	.046	23.9%	20.9%	1.19	.204
	3	24.1%	19.4%	1.32	.047	23.3%	20.2%	1.20	.200
	4	23.6%	18.9%	1.32	.047	22.7%	19.7%	1.20	.197
	5	23.1%	18.5%	1.32	.047	22.2%	19.2%	1.20	.195
	+∞	18.1%	14.1%	1.35	.058	17.4%	14.8%	1.22	.209
	2	1	23.3%	19.5%	1.26	.102	22.7%	20.0%	1.18
2		22.5%	18.7%	1.26	.101	21.9%	19.2%	1.18	.241
3		21.9%	18.1%	1.27	.100	21.3%	18.6%	1.19	.236
4		21.3%	17.6%	1.27	.100	20.8%	18.1%	1.19	.233
5		20.9%	17.2%	1.27	.100	20.4%	17.7%	1.19	.231
+∞		16.4%	13.1%	1.30	.109	15.9%	13.6%	1.21	.244
3		1	22.3%	18.9%	1.23	.143	21.9%	19.3%	1.17
	2	21.5%	18.2%	1.24	.140	21.1%	18.5%	1.17	.264
	3	20.9%	17.6%	1.24	.139	20.5%	17.9%	1.18	.258
	4	20.4%	17.1%	1.24	.137	20.0%	17.4%	1.18	.255
	5	20.0%	16.7%	1.25	.137	19.6%	17.0%	1.19	.253
	+∞	15.7%	12.8%	1.27	.143	15.3%	13.1%	1.20	.264
	4	1	21.8%	18.6%	1.22	.171	21.4%	18.9%	1.16
2		21.0%	17.9%	1.22	.168	20.6%	18.2%	1.17	.278
3		20.4%	17.3%	1.23	.165	20.0%	17.6%	1.17	.272
4		19.9%	16.8%	1.23	.164	19.6%	17.1%	1.18	.269
5		19.5%	16.4%	1.23	.163	19.2%	16.7%	1.18	.266
+∞		15.3%	12.6%	1.26	.166	15.0%	12.8%	1.20	.277
5		1	21.4%	18.4%	1.21	.192	21.1%	18.7%	1.16
	2	20.7%	17.7%	1.21	.188	20.3%	17.9%	1.17	.288
	3	20.1%	17.1%	1.22	.185	19.8%	17.4%	1.17	.282
	4	19.6%	16.6%	1.22	.183	19.3%	16.9%	1.18	.279
	5	19.2%	16.3%	1.22	.181	18.9%	16.5%	1.18	.276
	+∞	15.1%	12.4%	1.25	.183	14.8%	12.7%	1.20	.285
	+∞	1	19.8%	17.6%	1.16	.315	19.7%	17.7%	1.15
2		19.1%	16.9%	1.17	.306	19.0%	16.9%	1.15	.342
3		18.6%	16.3%	1.17	.300	18.5%	16.4%	1.16	.335
4		18.1%	15.9%	1.17	.295	18.0%	15.9%	1.16	.331
5		17.8%	15.5%	1.18	.292	17.7%	15.6%	1.16	.328
+∞		14.0%	11.9%	1.20	.278	13.8%	12.0%	1.18	.333

OR₁: odds ratio between missing and tobacco use status for self-report data; OR₂: odds ratio between urine missing and urine-verified failure among those who self-reported abstinence; Tx1: single contest + no counseling; Tx2: single contest + counseling; Tx3: multiple contests + no counseling; Tx4: multiple contests + counseling. P-values are based on the Chi-square test

methods are applicable to various tobacco or other substance use trials where the treatment goal is biochemically verified self-reported abstinence.

Conclusions

The proposed two-stage imputation method provides an effective sensitivity analysis tool for analyzing missing

data introduced at two different stages of outcome assessment, the self-report and validation time, frequently encountered in tobacco cessation studies. Our methods are also applicable to trials studying biochemically verified abstinence from other substance use such as alcohol and recreational drugs.

Additional file

Additional file 1: SAS Computing Code for Analyzing Enhanced Quit & Win Data. **Table S1.** Summary of imputation results for urine-verified abstinence assuming different levels of association between missing and abstinence when $\lambda = 0.5$. (DOCX 58 kb)

Abbreviations

MAR: Missing at random; MCAR: Missing completely at random; MNAR: Missing not at random; LOCF: Last observation carried forward; BOCF: Baseline observation carried forward; EM: Expectation-maximization; OR: Odds ratio; MAPS: Motivational and problem solving counseling

Acknowledgements

The authors thank the Enhanced Quit and Win study team for collecting the data and the two referees whose comments have helped to improve the manuscript substantially.

Funding

This study was supported by the Biostatistics Core of the University of Minnesota Masonic Cancer Center (funded by the National Cancer Institute 5P30CA077598) to CTL and XL, by the National Heart, Lung, and Blood Institute (5R01HL094183) to JLT, JSA, and XL, and by the Clinical and Translational Science Institute of University of Minnesota (National Center for Advancing Translational Sciences UL1TR002494). The funding body played no role in the design of the study and collection, analysis, and interpretation of data and/or in writing the manuscript.

Availability of data and materials

This is a manuscript demonstrating a novel application of a statistical method on data collected from a previous study [15]. Data requests should be addressed to JLT, the principle investigator of the Enhanced Quit & Win study.

Authors' contributions

YZ performed all the analyses and SAS programming and co-wrote the manuscript, which was part of her dissertation when she was a master of science (MS) student at the University of Minnesota; XL developed the original idea, supervise YZ's dissertation research, and co-wrote the manuscript; CTL and JSA were YZ's dissertation committee members and participated discussions; TLJ was the principle investigator of the Enhanced Quit & Win study and supervised the conduct of the trial and the interpretation of the analysis results; All authors contributed to the writing and revisions of the manuscript and have read and approved the manuscript.

Ethics approval and consent to participate

The Enhanced Quit & Win study was approved by the University of Minnesota's human subjects committee. Written informed consent was obtained from all participants in the "Quit and Win Study".

Consent for publication

Not applicable.

Competing interests

X L is a member of the editorial board (Associate Editor) of this journal.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Joseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, 1240 N 10th St, Milwaukee, WI 53205, USA. ²School of Public Health, Division of Biostatistics, University of Minnesota, 420 Delaware St. SE, MMC 303, Minneapolis, MN 55455, USA. ³University of Minnesota Masonic Cancer Center, Minneapolis, MN 55455, USA. ⁴Brown University School of Public Health, Box G-S121-5, Providence, RI 02912, USA. ⁵Division of General Internal Medicine, Department of Medicine, University of Minnesota, 717 Delaware St. SE, Minneapolis, MN 55414, USA.

Received: 23 August 2018 Accepted: 3 December 2018

Published online: 18 December 2018

References

- Lopez AD, Collishaw NE, Piha T. A descriptive model of the cigarette epidemic in developed countries. *Tob Control*. 1994;3:242–7.
- Peto R, Lopez AD, Boreham J, Thun M, Heath C Jr. Mortality from tobacco in developed countries: indirect estimation from national vital statistics. *Lancet*. 1992;339:1268–78.
- Peto R, Lopez AD, Boreham J, Thun M, Heath C Jr, Doll R. Mortality from smoking worldwide. *Br Med Bull*. 1996;52:12–21.
- Pirie K, Peto R, Reeves GK, Green J, Beral V, Collaborators MWS. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. *Lancet*. 2013;381:133–41.
- Delucchi KL. Methods for the analysis of binary outcome results in the presence of missing data. *J Consult Clin Psychol*. 1994;62:569–75.
- Little RJA, Rubin DB. *Statistical analysis with missing data*. 2nd ed. New York NY: Wiley; 2002.
- Thabane L, Mbuagbaw L, Zhang S, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol*. 2013; 13:92.
- 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.
- Borland R, Balmford J, Hutn D. The effectiveness of personally tailored computer-generated letters for tobacco cessation. *Addiction*. 2004;99:369–77.
- Nelson DB, Parlin MR, Fu SS, Joseph AM, An LC. Why assigning ongoing tobacco use is not necessarily a conservative approach to handling missing tobacco cessation outcomes. *Nicotine Tob Res*. 2009;11:77–83.
- Blankers M, Smit ES, van der Pol P, de Vres H, Hoving C, van Laar M. The missing=smoking assumption: a fallacy in internet-based smoking cessation trials? *Nicotine Tob Res*. 2016;18:25–33.
- 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;102:1564–73.
- Smolkowski K, Danaher BG, Seeley JR, Kosty DB, Severson HH. Modeling missing binary outcome data in a successful web-based smokeless tobacco cessation program. *Addiction*. 2010;105:1005–15.
- Barnes SA, Larsen MD, Schroeder D, Hanson A, Decker PA. Missing data assumption and methods in a smoking cessation study. *Addiction*. 2010; 105:431–7.
- Thomas JL, Luo X, Bengtson J, et al. Enhancing Quit & win contests to improve cessation among college smokers: a randomized clinical trial. *Addiction*. 2016;111:331–9.
- Jackson D, Mason D, White IR, Sutton S. An exploration of the missing data mechanism in an internet based smoking cessation trial. *BMC Med Res Methodol*. 2012;12:157.
- White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statist Med*. 2011;30:377–99.
- Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res*. 2018;27:2610–26.
- Daniels MJ, Hogan JW. *Missing data in longitudinal studies*. Taylor & Francis Group; 2008.
- Demirtas H. Multiple imputation under Bayesian smoothed pattern-mixture models for non-ignorable drop-out. *Statist Med*. 2005;24:2345–63.
- Yang X, Shoptaw S. Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. *Drug Alcohol Depend*. 2005;77:213–25.