FOCUS: RESEARCH AND CLINICAL ETHICS

Strategies for Dealing with Missing Data in Clinical Trials: From Design to Analysis

James D. Dziura[a,c*], Lori A. Post[a], Qing Zhao[b,c], Zhixuan Fu[b,c], and Peter Peduzzi[b,c]

[a]Department of Emergency Medicine, Yale School of Medicine, New Haven, Connecticut; [b]Yale School of Public Health, New Haven, Connecticut; [c]Yale Center for Analytical Sciences, New Haven, Connecticut

Randomized clinical trials are the gold standard for evaluating interventions as randomized assignment equalizes known and unknown characteristics between intervention groups. However, when participants miss visits, the ability to conduct an intent-to-treat analysis and draw conclusions about a causal link is compromised. As guidance to those performing clinical trials, this review is a non-technical overview of the consequences of missing data and a prescription for its treatment beyond the typical analytic approaches to the entire research process. Examples of bias from incorrect analysis with missing data and discussion of the advantages/disadvantages of analytic methods are given. As no single analysis is definitive when missing data occurs, strategies for its prevention throughout the course of a trial are presented. We aim to convey an appreciation for how missing data influences results and an understanding of the need for careful consideration of missing data during the design, planning, conduct, and analytic stages.

INTRODUCTION

Scientists require the best available clinical evidence, which can come from systematic reviews, experimental trials, and observational research [1-3]. Missing data are ubiquitous throughout various medical research designs, even randomized controlled trials (RCT†) that are considered the gold standard for evaluating a direct causal link between an intervention and outcome [4]. The randomized assignment protects against selection bias by equalizing known and unknown characteristics between intervention groups. Nevertheless, randomization alone is not sufficient to provide an unbiased intervention comparison [5]. Two additional requirements for an unbiased

*To whom all correspondence should be addressed: James Dziura, Suite 555, 300 George St., New Haven, CT 06519; Tele: 203-737-4468; Fax: 203-737-6983; Email: james.dziura@yale.edu.

†Abbreviations: RCT, randomized clinical trial; MCAR, missing completely at random; MAR, missing at random; MNAR, missing not at random; LOCF, last observation carried forward; NRC, National Research Council; ITT, intent-to-treat principle; MOP, manual of operations.

study are: 1) missing data from randomized patients do not bias the comparison of interventions [6,7] and 2) outcome assessments are obtained in a similar and unbiased manner for all patients [8,9]. For the latter, standardized participant assessments and blinding of the intervention are commonly employed to assure this balance [9] but are outside the scope of this review.

Despite efforts to minimize missing data through design, it is likely to occur in the majority of RCTs [10].

The intent-to-treat principle (ITT) requires the complete inclusion of all data from all randomized patients in the analysis and is considered the most appropriate criteria for assessment of the utility of a new therapy [8]. In an ITT analysis, all randomized participants have outcomes assessed and are analyzed in the group in which they were randomized (regardless of the actual intervention received). When participants drop out or miss visits, thus producing missing data, the ability to conduct an intent-to-treat analysis and draw conclusions about a causal link is compromised [11].

For nearly a century, scientists have been dealing with missing data by deleting or arbitrarily filling in missing cases post-hoc. These techniques are prone to bias to the extent the study results are meaningless [12,13], yet they continue to be utilized [14,15]. Over the past 2 1/2 decades, great strides have been made in the development of analytic techniques to estimate causal effects in the presence of missing data [16]. The increased utilization of methods such as inverse probability weighting [17], multiple imputation [18], and likelihood-based analysis [19] vastly improved rigor over the ad-hoc methods (e.g., last observation carried forward, complete case analysis) that previously dominated the RCT landscape. Still, it is important to understand that these methods are tools rather than solutions. When data are missing, the result of any statistical analysis relies on the unverifiable assumptions concerning the relation between the unobserved data and the reasons they are missing [20]. In other words, conclusions drawn from clinical trials with missing data

can vary depending on the assumptions made and the analytic method chosen.

Given there is no universal method to analyze missing data [21], the National Research Council (NRC) released guidelines on the Handling of Missing Data in Clinical Trials [22]. They advocate a more principled approach to design and analysis focusing on two critical elements: 1) careful design and conduct to limit the amount of missing data and 2) analysis that makes full use of information on all randomized participants and is based on careful attention to the assumptions about the nature of the missing data underlying estimates of treatment effects.

As guidance for researchers interested in performing clinical trials, this review is a non-technical overview of the consequences of missing data and a prescription for its treatment extending beyond the typically used analytic strategies to the entire research process including the stages of design, planning, and conduct. We provide an example-driven demonstration of the potential bias from incorrectly analyzing data with missing observations and explain the advantages/disadvantages of available analytic strategies. Most notably, we discuss the need for a paradigm shift in the way trials are managed, focusing on the prevention of missing data [23]. Many of the recommendations are applicable to research studies beyond the RCT.

TYPES OF MISSING DATA AND ANALYTIC EXAMPLES

When planning a study, conducting an analysis or critically reviewing trial results, it is necessary to contemplate how the missing data are generated. *Are certain groups more likely to having missing data? Are certain responses more likely to be missing?* To assist in our approach to designing and analyzing studies in the presence of missing data, Little and Rubin [24] have identified three categories for classifying how missing data are generated: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). Common examples of these miss-

Table 1. Common examples of the three missing data mechanisms in clinical trials.

Missing Mechanism	Examples
MCAR	Administrative censoring: follow-up is terminated because the study has ended. Migration-study participants move and are unable to complete visits. Random failure of the experimental instrument (e.g. test tube break, equipment failure)
MAR	Missing data caused by features of the study design such as participants being removed from the trial if their conditions are not controlled sufficiently well according to protocol criteria. Dropout based on recorded side-effects. Dropout based on known baseline characteristics.
MNAR	Dropout based on the unobserved response (e.g., a person not responding to treatment is more likely not to provide an observation). Participants miss a visit because they've had an outcome.

Table 2. MCAR Example: Results from a hypothetical clinical trial evaluating the effect of treatment on improvement in body fat (Outcome).

a. Results from the whole study population.

	Outcome			
Treatment	Y	N		
				• Risk Ratio=(30/40)/(20/40)=1.5
A	30	10	40	• SE(Ln(RR))=0.18
B	20	20	40	• 95% CI =1.05,2.15
	50	30	80	• Chi-Square=5.33 P=0.02

b. Results from completers following 30% missing data from a Missing Completely At Random Mechanism (MCAR)*.

	Outcome			
Treatment	Y	N		
				• Risk Ratio=(21/28)/(14/28)=1.5
A	21	7	28	• SE(Ln(RR))=0.22
B	14	14	28	• 95% CI =0.98,2.30
	35	21	56	• Chi-Square=3.73 P=0.053

*Each cell from Table 2a is multiplied by 70% to obtain 30% missing data from an MCAR mechanism in Table 2b.

ing data mechanisms are shown in Table 1 [25].

Missing Completely at Random

The definition of MCAR is that the likelihood of missing data is unrelated to any observed or unobserved variables. That is, the chance of missing data is the same for individuals in different treatment groups and those who have differential disease severity or treatment response. For example, a dropped test tube in a lab or an equipment failure may lead to missing data. As this is equally likely to occur in any subject in the study (i.e., regardless of treatment received, disease severity, etc.), it represents a completely random process. Subsequently, the average effect of the treatment will be the

Table 3. MAR example: Results from a hypothetical clinical trial evaluating the effect of treatment on improvement in body fat (Outcome).

a. Results from the whole study population.

Men	Outcome			
Treatment	**Y**	**N**		
A	225	75	300	• Outcome in A = 225/300 = 0.75
B	150	150	300	• Outcome in B = 150/300 = 0.50
	375	225	600	• Risk Ratio = 1.5

Women	Outcome			
Treatment	**Y**	**N**		
A	90	210	300	• Outcome in A = 90/300 = 0.30
B	60	240	300	• Outcome in B = 60/300 = 0.20
	150	450	600	• Risk Ratio = 1.5

Total	Outcome			
Treatment	**Y**	**N**		
A	315	285	600	• Outcome in A = 315/600 = 0.525
B	210	390	600	• Outcome in B = 210/600 = 0.35
	525	675	1200	• Risk Ratio =1.5
				• 95%CI=1.31, 1.71
				• Mantel Haenszel RR=1.5
				• 95% CI=1.33, 1.70

b. Results from completers following missing data from a Missing At Random Mechanism (MAR)*.

Men	Outcome			
Treatment	**Y**	**N**		
A	180	60	240	• Outcome in A = 180/240 = 0.75
B	120	120	240	• Outcome in B = 120/240 = 0.50
	300	180	480	• Risk Ratio = 1.5

Women	Outcome			
Treatment	**Y**	**N**		
A	81	189	270	• Outcome in A = 81/270 = 0.30
B	27	108	135	• Outcome in B = 27/135 = 0.20
	108	360	405	• Risk Ratio =1.5

Total	Outcome			
Treatment	**Y**	**N**		
A	261	249	510	• Outcome in A = 261/510 = 0.51
B	147	228	375	• Outcome in B = 147/375 = 0.39
	408	477	885	• Risk Ratio = 1.31
				• 95% CI=1.12, 1.52
				• Mantel Haenszel RR=1.5
				• 95% CI=1.30, 1.73

*Probabilities of missing are dependent on the combination of treatment and gender to mimic an MAR mechanism
Probability Missing for Men in Trt A = 0.20
Probability Missing for Men in Trt B = 0.20
Probability Missing for Women in Trt A = 0.10
Probability Missing for Women in Trt B = 0.55

same in those with and without missing data. Other examples include missing data due to administrative censoring or migration. To illustrate, Table 2a presents results for a hypothetical weight loss trial comparing the effect of two treatments (A and B) on the outcome, *improvement in body fat*, which is categorized as either yes or no. A common

measure of the relative effect of two treatments is the risk ratio (RR; the proportion of those in treatment A with an improvement divided by the proportion in treatment B with an improvement). In the study population, improvement in body fat was 1.5 times more likely for treatment A compared to B (i.e., RR = 1.5). The 95 percent confidence interval (1.05, 2.15) and the p-value (0.02) generated from the chi-square test show a statistically significant difference indicating that treatment A is superior to B.

To illustrate the MCAR mechanism, suppose that during the study the bioimpedence scale used to measure body fat was unreliable and failed in 30 percent of the participants. We also assume that this failure rate was not related to their treatment or whether or not they had an improvement in body fat (i.e., an MCAR mechanism). Table 2b shows the results of the study in those without missing data (i.e., completers). The risk ratio in the completers remains unchanged at 1.5. Thus, the estimate of the treatment effect was not biased by MCAR missing data. However, because of the reduced sample size due to missing data, the standard error is larger, resulting in a wider confidence interval (0.98, 2.30) that includes the null value of 1.0 and a chi-square statistic that is no longer significant at the 0.05 threshold (p = 0.053). This example illustrates that complete case analysis does not result in a biased estimate of the treatment difference in the absence of data arising from an MCAR mechanism; however, there is a loss of precision in the estimation of the treatment difference (i.e., larger standard errors and wider confidence limits), as well as a loss of power in the test of significance (i.e., more conservative test statistics yielding higher p-values).

Missing at Random

When the likelihood of missing data is related to observed variables but not to unobserved variables, the missing data mechanism is referred to as missing at random (MAR). If in a clinical trial dropout is more likely for men compared to women, but all men have the same chance of dropout and all women have the same chance of dropout, the missing data mechanism is MAR. Other examples of the MAR mechanism are missing data caused by features of the study design (e.g., providing rescue therapy when conditions are not sufficiently controlled according to protocol criteria), dropout based on recorded side effects or lack of efficacy, or dropouts based on known baseline characteristics.

To illustrate MAR, the results of another hypothetical weight loss trial are shown in Table 3a. Although the overall likelihood of improvement was higher in men (62.5 percent = 375/600) than in women (25 percent = 150/600), risk ratios were identical (RR = 1.5). Pooling of the gender specific data also results in the same crude and gender adjusted risk ratio (i.e., Mantel Haenszel RR = 1.5, 95 percent CI = 1.33, 1.70). Suppose the likelihood of having missing data was dependent on the combination of the treatment and gender (i.e., an MAR mechanism). For example, at the end of the study, 20 percent of men in treatment A, 20 percent of men in treatment B, 10 percent of women in treatment A, and 55 percent of women in treatment B had missing data. Note that the missing rates are dependent on the observed data but not on the unobserved outcome. Table 3b demonstrates the cross tabulations with the missing data rates applied to the whole study population. Despite the differential missing data, gender specific risk ratios remain the same, 1.5. However, crude pooling of the gender specific completers data results in a risk ratio of 1.31 (95 percent CI = 1.12, 1.52) that is smaller than the true risk ratio of 1.5. After adjustment for gender using the Mantel-Haenszel method, the true risk ratio of 1.5 is recovered in the completers, albeit with reduced precision and slightly larger confidence limits (1.30, 1.73) compared to the whole study population.

Missing Not at Random

When the likelihood of missing data depends on the unobserved data, the missing data is termed missing not at random (MNAR). For example, in substance abuse

Table 4. MNAR example: Results from a hypothetical clinical trial evaluating the effect of treatment on improvement in body fat.

a. Results from the whole study population.

Treatment	Outcome Y	Outcome N		
A	315	285	600	• Risk Ratio=(315/600)/(210/600)=1.5
B	210	390	600	• 95% CI =1.31, 1.71
	525	675	1200	• Chi-Square=37.3 P<0.001

b. Results from completers following missing data from a Missing Not At Random Mechanism*.

Treatment	Outcome Y	Outcome N		
A	189	285	474	• Risk Ratio=(189/474)/(210/444)=0.84
B	210	234	444	• 95% CI =0.73, 0.98
	399	519	918	• Chi-Square=5.14 P=0.02

*Probabilities of missing are dependent on the treatment and outcome to mimic an MNAR mechanism
Probability Missing for Outcome "Y" in Trt A = 0.40
Probability Missing for Outcome "N" in Trt A = 0.00
Probability Missing for Outcome "Y" in Trt B = 0.00
Probability Missing for Outcome "N" in Trt B = 0.40

trials with abstinence as an outcome, it is usual that dropout is higher for those who have relapsed. The problem is in those who drop out, relapse status is not typically obtained. In this case, the probability of having missing data is dependent on their unobserved data — relapse status. In another example, consider a study evaluating treatments to reduce cocaine use in which the outcome is drug level from a urine drug test measured every Monday morning. Participants who use cocaine over the weekend and do not show up for their urine test would be expected to have higher levels of cocaine metabolites. Thus, the likelihood of the data being missing is directly related to the unobserved cocaine level.

Continuing with our hypothetical clinical trial evaluating the effect of treatment on improvement in body fat, Table 4 demonstrates the consequences of missing data arising from a MNAR mechanism. The proportion of missing data was set at 40 percent in those who received treatment A and had an improvement in body fat and 40 percent in those who received treatment B and *did not* have any improvement. Therefore, missing data were related to the unobserved outcome (i.e., MNAR). The net result was a complete reversal of the risk ratio when examining the whole study population (Table 4a, RR = 1.5) compared to the completers (Table 4b, RR = 0.84), both reaching statistical significance. If the complete case analysis was used, conclusions would have been the opposite from the true effect.

An alternative taxonomy refers to missing data arising from an MNAR mechanism as *non-ignorable* because ignoring the process that leads to the missing data will lead to biased results. In contrast, the probability of missing data for ignorable missingness (MCAR or MAR) on a particular variable does not depend on the values of that variable given other observed variables. Such data may still produce unbiased estimates without the need for a model to explain the missing mechanism. To obtain unbiased treatment differences when missing data are MNAR requires modeling the relation between the outcome of interest and the probability of non-response. Determining this relation is a difficult task that highlights the importance of obtaining outcome
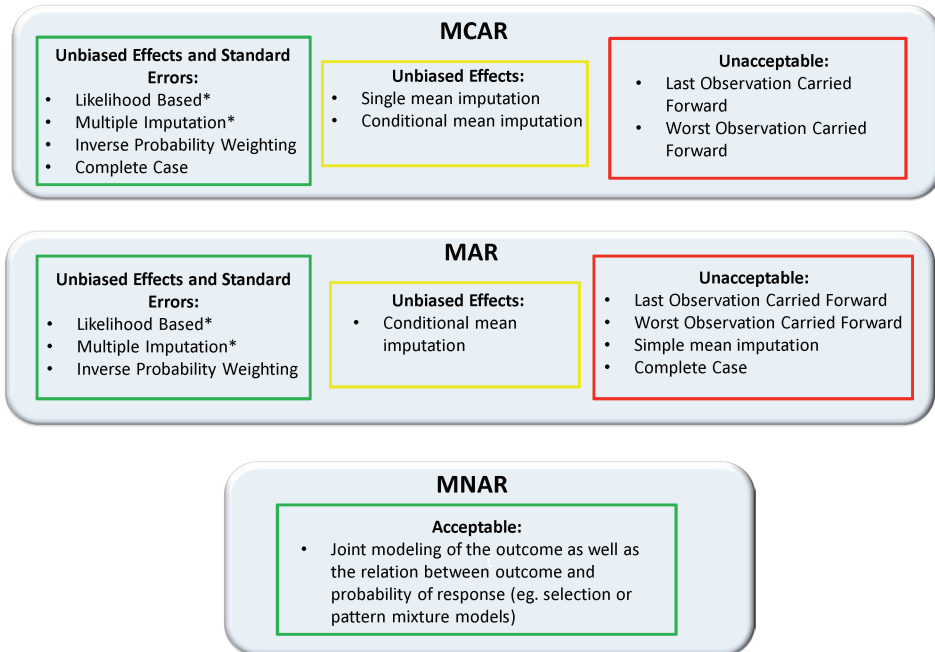
MCAR

Unbiased Effects and Standard Errors:	Unbiased Effects:	Unacceptable:
• Likelihood Based*	• Single mean imputation	• Last Observation Carried Forward
• Multiple Imputation*	• Conditional mean imputation	• Worst Observation Carried Forward
• Inverse Probability Weighting		
• Complete Case		

MAR

Unbiased Effects and Standard Errors:	Unbiased Effects:	Unacceptable:
• Likelihood Based*	• Conditional mean imputation	• Last Observation Carried Forward
• Multiple Imputation*		• Worst Observation Carried Forward
• Inverse Probability Weighting		• Simple mean imputation
		• Complete Case

MNAR

Acceptable:
• Joint modeling of the outcome as well as the relation between outcome and probability of response (eg. selection or pattern mixture models)

Figure 1. A summary of the acceptable and unacceptable analytic methods for types of missing data. Green boxes show methods giving unbiased estimates of treatment effects and correct estimates of standard errors and p-values, yellow boxes show methods giving only unbiased estimates of treatment effects, red boxes show unacceptable methods. *Preferred method as it uses all available data.

data on every randomized patient and collecting auxiliary data that may be predictive of dropout.

ANALYTIC APPROACHES TO MISSING DATA: EFFECTIVE TREATMENT, BUT NOT THE CURE

As demonstrated above, failure to appropriately account for the missing data mechanism during analysis can lead to erroneous conclusions. The problem that exists at the end of a study is that there is no definitive way to demonstrate which mechanism has led to the missing data. Addressing missing data during the analysis is not an acceptable comprehensive approach to managing missing data [6,21,26].

Statistical tests are available to demonstrate that missing data are not MCAR [27]. However, exclusion of MCAR is insufficient to ensure validity of a particular analysis. For instance, in a study evaluating two interventions on smoking cessation, exploration might uncover that baseline level of smoking

was associated with dropout. While this observation precludes the validity of an analysis that excluded those with missing data (i.e., analysis that assumes MCAR) and suggest the need for an analytic method that is valid under an MAR mechanism, it would not rule out an MNAR mechanism. That is, unless the missing data occur as a function of the study design, such as administrative censoring, a single technique for analysis does not exist. Rather a series of missing data analytic techniques are necessary to examine missing data bias. The following is an overview of analytic approaches, their assumptions about the missing data mechanisms, and their advantages and disadvantages. The approaches fall under four general strategies for coping with missing data: 1) use only data from participants completing the trial with no missing data; 2) use all available data; 3) impute (either single or multiple) values for missing data and analyze with complete case methods; or 4) develop a model for the data that includes a model for the missing data process [28]. The utility of each method under the different

missing data mechanisms is described in Figure 1.

Predefining Approach: During the Design of the Study

It is important to note the need to specify the plan to address the inevitable missing data during the study design phase. Pre-specification of the statistical analysis plan is important to avoid data-driven selection of results.

Common Approaches: Computationally Simple, but Rarely Acceptable

Complete Case Analysis is performed on only those subjects with a complete set of outcome data observed. Subjects with any missing data are excluded from analysis and typical statistical methods are used (e.g., chi-square analysis, t-tests, ANOVA, regression) on the reduced set of observations. For complete case analysis to provide an unbiased assessment of the intervention effect, the assumption that completers are a random sample of the full study sample (i.e., MCAR) is required. In addition to its computational simplicity whereby common statistical tests are conducted, an advantage of this method is that estimates of treatment differences are unbiased when missing data are MCAR. Disadvantages include loss of power and precision in the estimation of the treatment effect from the reduced sample size and the reliance on the strong MCAR assumption. When missing data are not MCAR, the estimate of the intervention effect will be biased.

Single Imputation creates a complete set of data for all randomized subjects by using a rule to set missing responses to a value. These approaches are computationally simple. Once created, simple statistical methods are used on the full set of data. There are many forms of single imputation, including last observation carried forward (LOCF), worst observation carried forward, and simple and conditional mean (or regression) imputation.

Due to its simplicity and its perceived (but often not true) tendency to provide conservative estimates of the treatment effect, LOCF has been a popular imputation technique. In LOCF, missing data for a subject is replaced with the last observed value for that subject. This analysis is not valid under MCAR but rather under the very specific and unrealistic assumption that the missing outcomes are equal to the last observed response [29]. The argument that LOCF provides a conservative estimate of an intervention is not universally true. Take, for example, a trial comparing two interventions for maintaining cognitive function in patients with Alzheimer's. The typical downward trajectory for these patients may lead to an upwardly biased estimate of function in those who drop out. A higher proportion of dropouts in one of the intervention groups would favorably bias that intervention in an LOCF analysis. Imputing identical values for an individual can also lead to underestimation of variability and a low p-value [30-33].

Two other single imputation approaches are simple and conditional mean (or regression) imputation [20]. In simple mean imputation, missing values are replaced by the mean for that variable. This approach ignores information from other variables that may be relevant if the data are MAR. While it can lead to valid estimates of treatment differences under MCAR, it will result in an underestimation of variability from the unseen data because a constant is imputed for all missing data regardless of differing participant characteristics. In contrast, conditional mean imputation accommodates associations with other observed variables by regressing the outcome on other observed variables in the completers. The missing outcomes are imputed using the regression equation that includes a random error component. Given the use of the observed data in the imputation and provided all covariates that sufficiently explain the missing data are included, valid estimates of treatment differences can be achieved under MAR. Nevertheless, the use of a single value to replace missing data does not fully capture the uncertainty that this value is correct. As with other single imputation procedures, underestimates in the variability of treatment ef-

fects will lead to inappropriately low p-values.

The NRC guidelines [22] for missing data state that single imputation methods like "last observation carried forward" should not be used as the primary analytic approach unless the assumptions that underlie these methods are scientifically justified.

Acceptable Approaches Under MAR

Inverse Probability Weighting. This method has its origins in sample survey research in which responses by survey participants are weighted to accommodate unequal probabilities of selection [34]. The survey weight for each participant is the inverse of their probability of selection, and, thus, those with lower probabilities of selection have a greater weight in the analysis. In missing data, probabilities of being observed are estimated and analyses are weighted by their inverse. Consequently, those with a low probability of observation will be weighted more highly in the analysis. Weights may be obtained from a model such as a logistic regression that includes intervention group, previous values of the outcome of interest, and other covariates that may be predictive of being observed.

Inverse probability weighting is a natural extension of a familiar technique that provides unbiased estimates of the treatment difference under the MAR assumption. Unlike the methods described below, it does not directly make use of all available data as only subjects with complete data are included in the final weighted model. Thus, power may be attenuated. Furthermore, to obtain valid standard errors and confidence intervals, specialized software is required to appropriately accommodate the weighted analysis. Packages such as SUDAAN, SAS, and STATA are equipped to do this.

Multiple Imputation (MI). While conditional mean imputation produces unbiased estimates of the treatment effect, it consistently underestimates the variability of this effect (i.e., artificially lowering p-values). This underestimation results from treating the observed and imputed values the same despite uncertainty in the imputed values.

The method of multiple imputation corrects for this by generating m completed data sets (typically m ranges from 5-20), each with the observed values and different plausible values imputed for the missing observations. After creating the m complete data sets, each is analyzed using usual methods (e.g., ANCOVA, regression), and the results are then combined across the analyses. It is important to note that imputations are not meant to be actual observations for that individual, but rather a statistically plausible set of values based on other information for that individual [35]. Thus, analysis from "filled-in" data sets provide statistically plausible results that would have been obtained if there were no missing data. Refer to [36,37] for examples of trials using MI.

Under the MAR assumption, MI produces unbiased estimates of the intervention effect and correct p-values. Other advantages of MI include the ability to handle not only missing outcome but missing covariate information, its relatively easy implementation, and the flexibility it provides by separating the imputation from the analytic model. The latter allows for increased complexity of the imputation model to make the MAR assumption more plausible. It also provides a simple and attractive framework for exploring sensitivity to non-random missing data. Drawbacks to MI include the inability to produce a unique estimate of the treatment effect (provides a different result each time you use MI) and the requirement for compatibility between the imputation and analysis models (e.g., the analysis model cannot contain variables, non-linearities or interactions that are not in the imputation model) [38,39]. The imputation model can be more complex than the analysis model, but the latter cannot contain variables that are not in the imputation model.

Likelihood-based Analysis. Maximum likelihood estimation (MLE) is a common estimation method in statistics contingent upon finding estimates of the treatment differences that maximize the probability of the observed data. To illustrate the MLE approach, suppose we do an experiment in N people where the probability of a success for

an individual is p and the probability of a failure is 1-p. If n people succeed and N-n people fail, the likelihood is proportional to the product of the probabilities of successes and failures or $p^n (1-p)^{N-n}$. The value of p that maximizes the likelihood is n/N or the overall proportion of success. In the weight loss trial with no missing data, maximum likelihood will produce the best estimate of the difference in body fat between intervention groups that maximizes the probability of observing the data. The problem when missing data occur is that we only observe a subset of the data yet we would like to draw conclusions based on the full data. Under MAR, the likelihood-based analysis allows us to accomplish this by averaging out the missing data from the joint likelihood of the observed and missing data. This is possible under MAR, because the future statistical behavior for a subject, conditional upon observed data, is the same whether or not that subject drops out. See [40,41] for examples of trials using MLE.

Under the MAR assumption, MLE produces unbiased estimates of the intervention effect and correct p-values and software is readily available in statistical packages (SAS, STATA, SPSS) for both continuous and discrete outcomes. Unlike MI, MLE provides a unique estimate of the treatment difference. MLE also requires fewer decisions than MI and is not reliant on compatibility of the imputation (as there is no imputation in MLE) and analysis model. Disadvantages of MLE include its reliance on parametric assumptions (e.g., normality) and that it is only appropriate for missing outcome data (i.e., it is unable to accommodate missing covariate data).

ITT or Per Protocol. ITT analyses determine the effect of intervention assignment (i.e., intention to give treatment) on the outcome regardless of the actual intervention received. It evaluates the random assignment. Per-protocol analyses evaluate the effect of the intervention for participants who adhered to the protocol. IPW, MI, and MLE methods are consistent with the ITT principle in that they do not exclude data from participants with incomplete responses and participants are analyzed within the group to which they were randomized [42]. However, in that they assume people with missing data would have had the same outcome experience if they had completed the study as similar people without missing data (i.e., those who adhered to the protocol) the MAR methods deviate from the ITT toward the per-protocol analysis [20]. Nevertheless, when data are missing, there is no unequivocal ITT analysis. As such, the MAR methods often provide a sensible approach for the primary analysis, but sensitivity analyses are recommended to understand the robustness to departures from the MAR assumption.

Acceptable Approaches Under MNAR

The basic approach to handling ignorable missing data (i.e., from an MCAR or MAR mechanism) is to adjust for all observable differences between missing and non-missing cases and assume that all remaining differences are unsystematic [38], thereby ignoring the process by which missing data happen. When missing data occur from an MNAR process, appropriate analysis requires the joint modeling of the outcome along with the missing data mechanism. This can be very complicated, given that under MNAR the model thereby the missing data process is rarely known for creating unverifiable assumptions for the analysis. For instance, we suspect relapse to lead to missing data in a study of treatment for substance abuse, but it is unlikely that all missing data is the result of relapse. To conduct an MNAR analysis, it is necessary to specify the strength of this relation, i.e., *what's the probability of having missing data given relapse* or similarly (but not equal) *what's the probability of relapse in those with missing data?*

Paralleling these related questions are two approaches to analysis under MNAR, selection [24], and pattern-mixture models [43]. Selection models require the specification of the relation between the outcome and the probability of being missing. For instance, in our weight loss study, the probability of a non-missing observation might be lower in those who have a recent unob-

served increase in body fat. On the other hand, pattern-mixture models specify the outcome distribution across the observed missing data patterns. For the weight loss example, this could correspond to indicating the probability of various weight loss profiles for those who drop out after their first visit compared to those who drop out after their second visit or those who don't ever drop out. While arbitrary, the chosen weight loss profiles could be further informed by recorded data such as the reason for dropout, where different profiles could be adopted for those dropping out because of migration compared to lack of efficacy.

While we can conceive of an endless number of MNAR analyses, in practice, a few reasonable scenarios are chosen in sensitivity analyses. Mallinckrodt et al. discuss a structure for conducting MNAR sensitivity analyses [44]. Ranges of likely values for the non-ignorable missing data can be solicited from experts (see Jackson et al. for examples of this process) [45]. The robustness of the conclusion with regard to the treatment difference is evaluated across these scenarios. For example, in the pattern mixture model described above, a range of weight loss profiles could be examined for their impact on study conclusions. See Hedeker et al. for an applied example [46].

EXPLORING MISSING DATA PATTERNS

In the presence of missing data, it is important to understand how much is missing and in what patterns. Exploratory analyses include examining proportions and time to drop out, differential reasons for dropout, and characteristics of those who do and do not drop out. Kaplan Meier curves are helpful to compare time to dropout between intervention groups or other important study variables. Similar rates and reasons for dropouts between intervention groups increase the confidence in validity of an ignorable analysis. Plots of outcomes at each time point for those who do and do not drop out at the next visit determine whether dropout is conditional on observed out-

comes. Logistic regression can identify factors most strongly associated with dropout. While a hypothesis test is available for MCAR, its use is limited as it only provides evidence that data are not MCAR and cannot be used as confirmation that data are MCAR [27]. No test can provide definitive proof of the mechanism that produced missing data.

HOW MUCH MISSING DATA IS TOO MUCH?

The proportion of missing data alone is not sufficient to indicate study validity, but studies with minimal missing data are more likely to produce valid conclusions. Schulz and Grimes [47] suggest that losses to follow-up less than 5 percent usually have little impact whereas losses greater than 20 percent raise serious flags about study validity. In-between levels lead to problems somewhere in the middle. To support this, Kristman et al. [48] demonstrated through simulation that substantial bias in the estimation of odds ratios under MNAR conditions may arise in cohort studies with loss to follow-up of 20 percent. However, the 5-20 general rule of thumb has no statistical justification and oversimplifies the problem as the bias resulting from missing data also depends on the missing data mechanism and the analytic method. It is certainly possible that the use of a complete case analysis when 5-20 percent of the data are missing from an MAR mechanism can lead to a biased treatment effect.

PRACTICAL APPROACHES TO MINIMIZE MISSING DATA

Given the uncertainties involved with the identification of the missing data mechanism and thus conclusions from analyses that rely on this identification, the best method for dealing with missing data is to prevent it. O'Neill described the need for a cultural shift, focusing on strategies to prevent missing data during the conduct and management of clinical trials rather than relying on imperfect analytic methods [23].

Table 5. Approaches to handling and preventing missing data during trial design, planning, conduct and analysis.

Design Stage	
Anticipate Expected Missing Data	1. Estimate the expected amount of missing data and likely reasons for it. 2. Account for missing data in the sample size calculations and develop a suitable pre-specified analytic plan.
Methods to Encourage Participant Retention	3. Limit burden to participant by reducing required visits and amount of data collected. 4. Adopt data collection methods that don't require face to face visits. 5. Utilize run-in periods, ascertainable treatment outcomes, shorter follow-up periods, randomized withdrawal designs where appropriate. 6. Budget for monetary incentives for participants that are weighted toward study completion.

Planning Stage	
Study Documentation	7. Develop detailed study documentation in the form of manual of operations addressing all aspects of the study including screening procedures, training requirements, methods of communication, delivery of treatment, schedule and windows for assessments, and data collection/entry/editing procedures.
Informed Consent	8. Develop an informed consent that distinguishes the difference between withdrawing from the treatment and withdrawing from the study.
Study Sites	9. Select study sites with strong track records for enrolling, following, and completing participants. 10. Adopt a reimbursement mechanism that encourages study completion.
Training Study Personnel	11. Train/certify study personnel for participant enrollment, data collection, data entry, delivery of treatment, etc. prior to enrollment with re-certification throughout trial if necessary. 12. Highlight the continued collection of data in participants that are not adherent to treatment but remain in the study.
Pilot Study	13. Test operational aspects of the trial (e.g., enrollment, retention, clarity of study manuals and data collection instruments, study burden on participants, randomization, treatment delivery).

Conduct Stage	
Create Monitoring Reports	14. Develop monitoring reports to regularly track amounts of missing data at the levels of the study site and study personnel. 15. Keep track of reasons for withdrawal from the study or intervention.
Enhance Participant Contact	16. Utilize approaches to keep the study participants engaged in the study including incentives, visit reminders, newsletters, and intermittent phone calls to monitor status. 17. Outline procedures for contacting individuals with missed visits in manual of operations. Identify and intervene in participants that are likely to drop out.
Data Entry and Management	18. Timely data entry allows earlier detection of problems with missing data. 19. Implement a verification process requiring fields to be checked for accuracy and all discrepancies resolved before data entry.
Communication	20. Devise an efficient method of communication with study personnel for identifying and resolving unanticipated issues that arise during the study.

Analytic Stage	
Explore Missing Data	21. The amount of missing data, missing data patterns and variables associated with missingness will help to inform the primary and sensitivity analyses.
Use All Available Data	22. For primary analysis, use methods that make use of all available data such as multiple imputation or likelihood-based approaches. These methods make weaker assumptions about the missing data compared to complete case analysis. 23. For primary analysis, avoid the use of ad-hoc solutions (e.g., last observation carried forward) as they make unreasonable assumptions about the mechanism that produced the missing data.
Perform sensitivity analysis	24. Use methods such as pattern mixture or selection models to examine robustness of conclusions to reasonable MNAR mechanisms.

The National Research Council (NRC) has provided recent guidance, listing several recommendations to prevent missing data during the design and conduct of clinical trials. Interpretations of these guidelines and additional approaches (see Table 5) to prevent missing data have been described [6,21,23,26] as well as the potential dynamic between statisticians and clinical investigators to achieve them [49]. The overarching

goal is the creation of procedures and a climate and culture that will maximize the collection of complete data [6]. Collaboration with a data coordinating center with a history of conducting clinical trials is strongly encouraged to improve implementation.

Approaches in Study Design

While it won't necessarily attenuate missing data, *predicting the expected proportions of missing data* is recommended during the design phase as it 1) can impact the variability and estimate of the effect size thus influencing the required sample size and 2) be helpful in directing the range of sensitivity analyses required [50]. Of note, *inflation of the sample size estimate for dropouts* helps to preserve power, but it is not a comprehensive strategy for dealing with missing data as it does not preclude the opportunity for missing data-related bias.

Other design techniques are aimed at reducing the number of participants with a missing primary outcome, usually from dropout. Although collecting an abundance of data to answer secondary, tertiary, and exploratory questions is tempting, the focus of a trial is to answer a primary question. Therefore, the benefits of collection of data beyond what is required to answer the primary question and crucial supportive evidence for it must be carefully weighed against the threat of missing data. *Limiting participant burden* by reducing the number of visits and the amount of data collected at each visit are universally applicable study design approaches.

The applicability of *other design strategies* is study-specific. Run-in periods are beneficial to minimize missing data even though the original intent was to maximize adherence and exclude those intolerant to treatment. Addressing target populations with incentive to remain in the study, using flexible treatment regimens that increase adherence, using outcomes that can be ascertained in a high proportion of participants, and shortening the follow-up period for the primary outcome also are design approaches to minimize missing data [26,51].

The randomized withdrawal design was initially proposed to evaluate long-term effectiveness. Participants deemed "responders" to a treatment are randomized to either maintain their current treatment group or receive placebo [52]. For example, to minimize the long-term impact of a weight loss therapy, a short-term endpoint (e.g., 3 months), where the propensity for missing data is low, could be used to define those responding to the treatment. Responders could then be randomized to continue or receive standard care and then be evaluated at 12 months to examine maintenance of effectiveness.

Approaches in Study Planning and Conduct

Several common-sense opportunities should be considered during the planning and conduct of a study to minimize missing data [26,51]. Approaches are directed at the participant, the data collection process, and the study team.

In addition to limiting participant burden, *enhancing participant engagement* is an effective approach to promote study completion. Updating contact information regularly and providing incentives for participation decrease the likelihood of dropout. Monetary incentives can be increased for subsequent visits beyond the index visit to encourage study completion [51]. *Informed consent procedures should clearly distinguish between discontinuation of study treatment and discontinuation of follow-up* so participants are clearly aware of their ability to complete the study despite non-adherence [21]. Feelings of investment and appreciation for participation may be reinforced through newsletters, study websites and social networking, access to study findings, and frequent expressions of gratitude from the study team. It's particularly important to *engage participants at the greatest risk of dropout*. Those at highest risk may be identified by asking at each visit about their likelihood for attendance at the next visit. If the chances of attending are low, reasons or barriers along with potential solutions can be explored.

When a participant drops out, it's helpful to *determine the reason for withdrawal*

as well as ancillary factors that may have been associated with their decision. This information can be utilized to meet the MAR assumption (e.g., by including them as adjustment covariates in the analysis) or to inform sensitivity analyses [20].

In addition to participant cooperation, complete data collection is contingent upon adequate study documentation, appropriate training of study personnel, a robust clinical management procedure, and a structure for efficient communication among the study team. A *manual of operations (MOP)* provides documentation to guide the conduct of the study including timing windows for assessments, instructions for data collection, and training requirements. Many NIH institutes such as the National Institute on Aging offer online guidelines and templates for the MOP [53]. Prior to enrollment, *training and certification of study personnel* is essential for proper execution of the protocol. Highlighting the need for complete data collection regardless of participant treatment adherence is a key component of this training. Pilot testing the protocol can identify possible sources of missing data. Development of participant calendars for visit reminders and automated participant contact can be made possible by an *electronic clinical data management system*. Visual data completeness checks may be performed prior to the participant leaving or electronic data capture systems programmed to require data completeness.

Constant exchange of information among the study team is essential so a *structure for efficient communication* is required. The process for dissemination of revised operations must be delineated. Regular study team meetings (phone or face to face) or web-based discussion boards permit the identification and an opportunity to resolve potential missing data issues [26].

Finally, data completeness should be evaluated through *regular monitoring reports* made available to the entire study team [26]. These reports require the *automization of data collection or entry of data in a timely manner* (i.e., as close to real-time as possible) in order to be utilized to improve study

conduct. The reports should include site-specific summaries of subject retention and data completeness, particularly for the essential study outcomes. *A priori* targets for unacceptable rates of missing data will aid in interpretation and possible remediation [21,51]. Underperforming sites can undergo additional training or be closed. To avoid this, the *track record of the site with regard to enrolling, following and collecting complete data should be an important criterion for study-site selection*.

CONCLUSION

Missing data are common in clinical trials. Given that no single analysis will be definitive in the presence of missing data, limiting missing data through the creation of a climate and culture that maximizes the collection of complete data is necessary. Our proposed recommendations operationalize this by providing specific guidance for each stage of the trial. In the design stage, researchers should anticipate missing data patterns and causes and consider methods/designs that encourage participant retention. Developing detailed study documentation, training study personnel and testing operational aspects of the trial are important during the planning stage. Regular monitoring of missing data and enhanced participant contact is recommended for the conduct stage. While easy to implement, ad hoc methods such as complete case analysis and last observation carried forward are not advocated as primary analytic strategies. As a primary strategy, the use of all available data is recommended through methods such as multiple imputation and likelihood-based analysis. Sensitivity analyses under MNAR are appropriate to evaluate robustness of conclusions to a range of sensible conditions.

REFERENCES

1. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why observational studies should be among the tools used in comparative effectiveness research. Health Aff (Millwood). 2010;29(10):1818-25.
2. Kaplan BJ, Giesbrecht G, Shannon S, McLeod K. Evaluating treatments in health care: The instability of a one-legged stool. BMC Med Res Methodol. 2011;11:65.

3. Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. Per Med. 2010;7(1):33-47.

4. Hennekens CH, DeMets D. Statistical Association and Causation Contributions of Different Types of Evidence. JAMA. 2011;305(11):1134-5.

5. Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]. March 2012. Available from: http://www.effectivehealthcare.ahrq.gov/ehc/products/322/998/MethodsGuideforCERs_Viswanathan_IndividualStudies.pdf.

6. Scharfstein S, Neaton J, Hogan J, Little R, Daniels M, Roy J, et al. Minimal Standards in the Prevention and Handling of Missing Data in Observational and Experimental Patient Centered Outcomes Research [Internet]. 2012 Mar 15. Available from: http://www.pcori.org/assets/Minimal-Standards-in-the-Prevention-and-Handling-of-Missing-Data-in-Observational-and-Experimental-Patient-Centered-Outcomes-Research.pdf.

7. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407-15.

8. Lachin JM. Statistical considerations in the intent-to-treat principle. Control Clin Trials. 2000;21(3):167-89.

9. Evans SR. Fundamentals of clinical trial design. J Exp Stroke Transl Med. 2010;3(1):19-27.

10. Trials PoHMDiC. The Prevention and Treatment of Missing Data in Clinical Trials. National Academy Press; 2010.

11. Can O, Yilmaz A, Hasdogan M, Alkaya F, Turhan S, Can M, et al. Has the quality of abstracts for randomised controlled trials improved since the release of Consolidated Standards of Reporting Trial guideline for abstract reporting? A survey of four high-profile anaesthesia journals. Eur J Anaesthesiol. 2011;28(7):485-92.

12. Rubin DB, Little RJ. Statistical analysis with missing data. Hoboken, NJ: J Wiley & Sons; 2002.

13. Wilkinson L. Task Force on Statistical Inference, APA Board of Scientific Affairs. Statistical methods in psychology journals: Guidelines and explanations. American Psychologist. 1999;54:594-604.

14. Bodner TE. Missing data: prevalence and reporting practices. Psychol Rep. 2006;99(3):675-80.

15. Peugh JL, Enders CK. Missing data in educational research: A review of reporting practices and suggestions for improvement. Review of Educational Research. 2004;74(4):525-56.

16. Enders CK. Applied missing data analysis. New York: The Guilford Press; 2010.

17. Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2006;169(3):571-84.

18. Beale EM, Little RJ. Missing values in multivariate analysis. Journal of the Royal Statistical Society Series B (Methodological). 1975;129-45.

19. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-92.

20. Carpenter JR, Kenward MG. Missing data in randomised controlled trials — a practical guide [Internet]. 2007. Available from: http://www.hta.nhs.uk/nihrmethodology/reports/1589.pdf.

21. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012;367(14):1355-60.

22. NAS. The Prevention and Treatment of Missing Data in Clinical Trials. The National Academies Press; 2010.

23. O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol Ther. 2012;91(3):550-4.

24. Little RJA, Rubin DB. Statistical analysis with missing data. New York: Wiley; 1987. 278 p.

25. Little RJ, Rubin DB. The analysis of social science data with missing values. Sociological Methods & Research. 1989;18(2-3):292-326.

26. Wisniewski SR, Leon AC, Otto MW, Trivedi MH. Prevention of missing data in clinical research studies. Biol Psychiatry. 2006;59(11):997-1000.

27. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. J Am Stat Assoc. 1988;83(404):1198-202.

28. Carpenter J, Pocock S, Johan Lamm C. Coping with missing data in clinical trials: A model-based approach applied to asthma trials. Stat Med. 2002;21(8):1043-66.

29. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer; 2009. 568 p.

30. Liu G, Gould AL. Comparison of alternative strategies for analysis of longitudinal trials with dropouts. J Biopharm Stat. 2002;12(2):207-26.

31. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. J Biopharm Stat. 2001;11(1-2):9-21.

32. Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. Stat Med. 1995;14(17):1913-25.

33. Barnes SA, Mallinckrodt CH, Lindborg SR, Carter MK. The impact of missing data and

how it is handled on the rate of false-positive results in drug development. Pharm Sat. 2008;7(3):215-25.

34. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res. 2013;22(3):278-95.

35. Raghunathan TE. What do we do with missing data? Some options for analysis of incomplete data. Annu Rev Public Health. 2004;25:99-117.

36. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet. 2011;378(9785):49-55.

37. Savoye M, Shaw M, Dziura J, Tamborlane WV, Rose P, Guandalini C, et al. Effects of a weight management program on body composition and metabolic parameters in overweight children: a randomized controlled trial. JAMA. 2007;297(24):2697-704.

38. Allison PD. Multiple imputation for missing data. Sociological Methods & Research. 2000;28(3):301-9.

39. Allison PD. Missing data. Thousand Oaks, CA: Sage Publications; 2002. 93 p.

40. Piacentini J, Woods DW, Scahill L, Wilhelm S, Peterson AL, Chang S, et al. Behavior therapy for children with Tourette disorder: a randomized controlled trial. JAMA. 2010;303(19):1929-37.

41. Weintraub WS, Spertus JA, Kolm P, Maron DJ, Zhang Z, Jurkovitz C, et al. Effect of PCI on quality of life in patients with stable coronary disease. N Engl J Med. 2008;359(7):677-87.

42. Mallinckrodt CH, Sanger TM, Dubé S, DeBrota DJ, Molenberghs G, Carroll RJ, et al. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. Biol Psychiatry. 2003;53(8):754-60.

43. Little RJ, Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. Biometrics. 1996;52(1):98-111.

44. Mallinckrodt CH, Lin Q, Molenberghs M. A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. Pharm Stat. 2013;12(1):1-6.

45. Jackson D, White IR, Leese M. How much can we learn about missing data?: an exploration of a clinical trial in psychiatry. J R Stat Soc Ser A Stat Soc. 2010;173(3):593-612.

46. Niaura R, Spring B, Borrelli B, Hedeker D, Goldstein MG, Keuthen N, et al. Multicenter trial of fluoxetine as an adjunct to behavioral smoking cessation treatment. J Consult Clin Psychol. 2002;70(4):887-96.

47. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet. 2002;359(9308):781-5.

48. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751-60.

49. Hughes S, Harris J, Flack N, Cuffe RL. The statistician's role in the prevention of missing data. Pharm Stat. 2012;11(5):410-6.

50. Guideline on missing data in confirmatory clinical trials. European Medicines Agency [Internet]. 2010. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.

51. Scharfstein DO, Hogan J, Herman A. On the prevention and analysis of missing data in randomized clinical trials: the state of the art. J Bone Joint Surg Am. 2012;94(Suppl 1):80-4.

52. Amery W, Dony J. A clinical trial design avoiding undue placebo treatment. J Clin Pharmacol. 1975;15(10):674-9.

53. National Institute on Aging. Guidelines for Developing a Manual of Operations and Procedures (MOP) [Internet]. Available from: http://www.nia.nih.gov/research/dgcg/clinical-research-study-investigators-toolbox/startup.