Intricacy of missing data in clinical trials: Deterrence and management

Richa Singhal, Rakesh Rana
Department of AYUSH, Central Council for Research in Ayurvedic Sciences, Ministry of Health and Family Welfare, GOI, New Delhi, India

Abstract
Missing data is frequently encountered in clinical studies. Unfortunately, they are often neglected or not properly handled during data analysis and this may significantly bias the results of the study, reduce study power and lead to invalid conclusions. Substantial instances of missing data are a serious problem that undermines the scientific trustworthiness of causal conclusions from clinical trials. The assumption that statistical analysis methods can compensate for such missing data is not justified. Hence aspects of clinical trial design that limit the probability of missing data should be an important objective, while planning a clinical trial. In addition to specific aspects of trial design, many components of clinical trial conduct can also limit the extent of missing data. The topic of missing data is often not a major concern until it is time for data collection and data analysis. This article discusses some basic issues about missing data as well as prospective “watch outs” which could reduce the occurrence of missing data. It provides some possible design considerations that should be considered in order to alleviate patients from dropping out of a clinical trial. In addition to these the concept of the missing data mechanism has also been discussed. Three types of missing data mechanisms missing completely at random, missing at random and not missing at random have been discussed in detail.

Key words: Data monitoring and co-ordination, International Conference for Harmonization E9 Guidelines, missing data, missing data mechanisms, study conduct, study design

Submission: 06-09-2013 Accepted: 14-05-2014

Introduction
In a clinical trial major source of missing data are the participants who discontinue the treatment because of adverse events, intolerance of the investigated intervention, lack of efficacy or due to inconvenience. However, some of the less discussed issues for the occurrence of missing data in clinical trials are poorly designed assessment schedules, use of inefficient data collection tools, missed visits or omission of some questions. These problems can be reduced by some simple adjustments in the assessment schedules or minor modifications in the data collection tools, moreover, these measures can also reduce participants’ burden and inconvenience. An effective missing data monitoring system can retrospectively retrieve almost all types of missing data items except the one that arose due to patients dropping out from the study. It is a common assumption among the researchers that statistical analysis methods can compensate for such missing data, which is not really justified. It should always be kept in mind that there is no statistical fix for an inadequately designed and inefficiently executed study.[1,2]

Properly conducted clinical trial can limit the extent of missing data. This article aims to draw the attention of researchers on some of the issues in study design and its conduct which leads to missing values. Some measures for improving the study design and its conduct have also been suggested.
Reasons For The Occurrence of Missing Data

Missing data due to trial participant action or inaction
a. Drop out from the study due to ineffective treatment, occurrence of adverse events or poor drug palatability.
   In some situations, participants also drop out from the study due to early recovery
b. Sometimes participants miss some schedule assessment visits due to inadequate compensation, unreasonably excessive number of visits, inconvenient timings on which patients are called upon, long travel times, etc
c. Often, participants refuse to undergo some requisite study procedures
d. Many times participants refuses to provide some information that they consider sensitive or personal during the interaction with investigators or while filling up any self-administered questionnaires
e. At times, participants intentionally or unintentionally do not fill the complete information in the self-administered questionnaires which are too lengthy.

Missing data due to investigator’s action or inaction
a. At times, investigator fails, to explain in detail the requirements of the study to participant
b. Sometimes, investigators ask leading questions from the participants which they don’t respond fully
c. At times, investigators carelessly overlook that whether the participant has completely filled the self-administered questionnaires or not
d. Often, investigators fail to actively encourage the participants to participate sincerely in the study
e. In many studies, missing data also arise due to lack of tracking the patients follow-ups and failure to remind them for it, especially in studies with longer durations.

Missing data due to deficiencies in the study design
a. In many trials too many assessment schedules leading to the increased number of visits are planned than actually needed for testing the study hypothesis
b. The assessments are made too long by collecting the irrelevant information which at times leads to the intricacy for the participants in retaining attention, while interacting with the investigators
c. Self-administered questionnaires included in the study are not fully understood by the participants because of complicated and confusing language
d. In many clinical studies, too many questionnaires inheriting similar questions are administered, which reduces the participants interests in filling them completely.

Measures That Can Be Adopted to Curtail Missing Data

Most of the aforesaid described issues can be addressed to minimize the occurrence of missing data in clinical studies. Slight alterations in the study design along with the modifications in operational procedures implemented for execution of the study could result in minimizing the missing data problem in the clinical trials. Some of the measures that would be helpful in curtailing the trouble caused by missing data in the clinical trials are described below.

Reducing frequency of missing data through improvement in study design
The study should be designed with limited number of assessment schedules that are really needed for achieving the study objectives. Measures should be taken to consolidate the questionnaires by merging the overlapping questions to make them more effective and less time consuming for the participants. Efforts should be made to minimize the number of hospital visits for the participants; if possible try to collect the desired information through telephonic discussion with the participants. In studies with long term follow-ups provide more incentives to the participants in the last visits, so as to ensure participants retention in the study till the end. However, adequate care should be taken in distinguishing these measures from undue inducements, and hence the compensation that will be paid to the participants at each visit should be approved by the Institutional Ethics Committee. Too frequent assessment visits should be avoided and a time window should be framed for allowing the participants to turn up as per their availability, if applicable as per the nature and sensitivity of the study. Care should be taken while selecting the study site and investigators. Emphasis should be made on selecting the investigators and study sites who have a good track record in the previous studies of enrolling the participants, respecting them and extracting the complete information from them.

Reducing the frequency of missing data by investigators and study staff conduct
Investigators, while obtaining the consent from the participants should highlight that the study aims at contributing some scientific knowledge and validation of the drugs in the particular disease area. Investigators should explain the participants that their contribution is important in moving the scientific knowledge ahead in the disease area. These efforts make the participants feel that they are an important part of the study and it helps them in retaining interest in the study and its procedures till end. Sometimes, in the studies with long follow-up participants loose interest in the mid of the study, therefore, the investigators and staff should keep
on encouraging the participants and make them realize that their contribution is important at every visit. The study staff should be trained to treat the participants with utmost care and to create a welcoming and friendly atmosphere for the participants and their family members. Participants should realize that the investigator and study staff keep their health on priority than the study itself. Study staff should be flexible in allowing the participants to visit the hospital within the allowable time window and should arrange the transport and child care facilities, etc., if the participants need them, this helps in making the participants realize that study staff is aware of the difficulties faced by them during the visits and also appreciate their efforts.

Managing Missing Data

Missing data are often unavoidable, despite the best efforts to reduce their occurrence through trial design and conduct. The “International Conference for Harmonization Guidelines E9 Statistical Principles for Clinical Trials” states that missing data is a potential source of bias in clinical trials. Hence every effort should be undertaken to fulfill all the requirements of the protocol concerning the collection and management of data. In reality, however, there will almost always be some missing data. Unfortunately, no universally applicable methods of handling missing values can be recommended. Hence, the best way of handling missing data is to avoid it. Some measures that can be taken to manage missing data are described below:

a. A data coordinating and monitoring team should be constituted for all the clinical trials. It is a common practice done by statisticians to monitor the data collection and quality of the data. The statisticians should thoroughly check the data for missing values also along with other data quality indicators.

b. The data coordinating team should respond quickly and intimate to the study investigators if any missing value is encountered during the course of the study by generating the missing data reports at regular intervals, so that the majority of the missing data items can be salvaged in time (i.e. before the investigator loses the participant due to completion of the study or due to dropping out in between the study).

c. Missing data can also be avoided by the use of properly designed questionnaires and data collection tools. All instruments used in the study should be user-friendly and should be translated into simple language for the participants to easily understand. Language in the self-administered questionnaires given to the participants should be simple and unambiguous.

d. Electronic formats should be designed by providing the pop up messages for completing the requisite fields and hence that while entering the data at the site itself investigators could be made aware of the missing data items and necessary steps could be taken in time.

Understanding The Nature Of Missing Data

The most appropriate way to handle missing data will depend upon how the data points are missing. Recent progress in dealing with missing data has come from an understanding of the reasons why data may be missing. Most statistical methods dealing with missing data are strongly dependent on the nature of these mechanisms. Hence, violation of these assumptions can result in biased parameter estimation. Little and Rubin (1987) developed the very useful taxonomy for describing the assumptions regarding the type of missing data. They have classified missing data mechanisms into three broad categories.

Missing completely at random

Missing completely at random (MCAR) means that the missing data mechanism is unrelated to the values of any variables, whether missing or observed. Under MCAR the probability of the missing value is independent of both observed and unobserved measurements. The missing mechanism is unrelated to any inference the researcher wish to draw about the intervention effect. For example, some observations may be missing because of equipment failure in the clinic, or because a patient was unable to attend for some reason not related to his/her illness or its intervention (e.g. his/her child was unwell). Other examples of MCAR occur when a participant misses a survey administration due to scheduling difficulties or other unrelated reasons (such as a doctor’s appointment), or an administrative blunder causes several test results to be misplaced prior to data entry. In such type of cases the average effect of intervention under study will be the same among those who do and do not have missing data.

When the data is MCAR there is no impact on bias however; a loss of statistical power can occur. MCAR data exists when missing values are randomly distributed across all observations. In this case, observations with complete data are indistinguishable from those with incomplete data. That is, whether the data point on Y is missing is not at all related to the values of Y or to the values of any Xs in that dataset. E.g., if we are asking people their weight in a trial, some people might fail to respond for no good reason – that is, their nonresponse is in no way related to what their actual weight is, and is also not related to anything else we are measuring. The missing values are MCAR if the reason for the move is unrelated to other variables in the data set (e.g. socioeconomic status, disciplinary problems, or other study-related variables). Hence, when the data is MCAR than estimating the effect of intervention by analyzing
the data of only those participants who do not have missing data will give the sensible results.

Missing at random
In practice trial, data are rarely MCAR. Usually there is an association between the chance of patient withdrawal and observations – typically intervention, baseline and (in longitudinal follow-up) measurements prior to withdrawal. Missing at random (MAR) data exists when the observations with incomplete data differ from those with complete data, but the pattern of data missing on Y can be predicted from other variables in the dataset (Xs) and beyond that bears no relationship to Y itself – that is, whatever nonrandom processes existed in generating the missing data on Y can be explained by the rest of the variables in the dataset. MAR assumes that the actual variables where data are missing are not cause of the incomplete data instead; the cause of the missing data is due to some factors that we also measured. E.g., only female sex may be less likely to disclose its weight. In this case, it is not sensible to include in the analysis only those with complete data. For example, suppose that worse health at baseline is associated both with increased risk of withdrawal and poor response to intervention. Analyzing data from the patients who remain to the end of the trial will thus give an over optimistic view of the intervention effect. Another example of MAR data is, a participant may be removed from a trial if his/her condition is not controlled sufficiently well (according to pre-defined criteria on the response).

Not missing at random or missing not at random or nonignorable
The pattern of data missingness is nonrandom and it is not predictable from other variables in the dataset. Not missing at random (NMAR) data arise due to the data missingness pattern being explainable only by the very variable (s) on which the data are missing. E.g., heavy or light people may be less likely to disclose their weight. NMAR data commonly occurs (unless “discomfort” is measured and observed for all patients). NMAR data commonly occurs when people do not want to reveal something very personal or unpopular about them.

Discussion
Missing data issues have been discussed and debated for many years. Handling of missing data in clinical trials has been recognized as an important issue not only for statisticians who analyze the data, but also for the clinical study team who conduct the study. Decision to handle such study participants data must be taken prior to the initiation of the study. Decision to do intention-to-treat analysis must be taken prior to the start of the study. Data cleaning should be done at frequent intervals in long term follow-up studies. Moreover, sample size should be calculated taking into account the attrition rate. Investigator’s performance must be gauged based on completed cases. For reducing the lost to follow-up missing values complicated patients must be identified and regular follow-up reminders should be given to them. A nonformal relation between investigator and patient should be maintained. All these must be approved by Ethics Committee.

References
1. Chong NG, Yusoff M. Missing values in data analysis, ignore or impute? Educ Med J 2011;391:e6-11.
2. Fitzmaurice G. Missing data: Implications for analysis. Nutrition 2008;24:200-2.
3. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: National Academies Press; 2010.
4. Ibrahim JG, Chu H, Chen MH. Missing data in clinical studies: Issues and methods. J Clin Oncol 2012;30:3297-303.
5. Ethical Guidelines for Biomedical Research on Human Participants. Delhi: Indian Council of Medical Research; 2006.
6. Statistical Principles for Clinical Trials; Step 5; Note for Guidance on Statistical Principles for Clinical trials; International Conference on Harmonization, Topic E9. London: European Medicines Agency; 1998.
7. Little RJ. Methods for handling missing values in clinical trials. J Rheumatol 1999;26:1654-6.
8. Rubin DB. Inference and missing data. Biometrika 1976;63:581-92.
9. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: John Wiley and Sons; 1987.
10. Little RJ. Modeling the drop out mechanism in repeated-measures studies. J Am Stat Assoc 1995;90:1112-21.