Part I

BACKGROUND AND SETTING

This section begins with a chapter illustrating how missing data can cloud inferences to be drawn from clinical trials – in other words, why missing data matter. Chapter 2 focuses on the mechanisms that give rise to missing data. Understanding these mechanisms is the essential background needed to understand the possible consequences of missing data. Chapter 3 discusses estimands – what is to be estimated from the trial. Together, these chapters form the basis for discussion on how to limit missing data and how to handle missing data that remains.

1

Why Missing Data Matter

The evidence to support new medicines, devices, or other medical interventions is based primarily on randomized clinical trials. Many of these trials involve assessments taken at the start of treatment (baseline), followed by assessments taken repeatedly during and in some scenarios after the treatment period. In some cases, such as cancer trials, the primary post-baseline assessments are whether or not some important event occurred during the assessment intervals. These outcomes can be summarized by expressing the multiple post-baseline outcomes as a time to an event, or as a percentage of patients experiencing the event at or before some landmark time point. Alternatively, the multiple post-baseline assessments can all be used in a longitudinal, repeated measures analysis, which can either focus on a landmark time point or consider outcomes across time points.

Regardless of the specific scenario, randomization facilitates fair comparisons between treatment and control groups by balancing known and unknown factors across the groups. The intent of randomization in particular, and the design of clinical trials in general, is that differences observed between the treatment and control groups are attributable to causal differences in the treatments and not to other factors.

Missing data is an ever-present problem in clinical trials and has been the subject of considerable debate and research. In fact, the U.S. Food and Drug Administration convened an expert panel to make recommendations for the prevention and treatment of missing data (NRC, 2010). The fundamental problem caused by missing data is that the balance provided by randomization is lost if, as is usually the case, the patients who discontinue the study differ in regards to the outcome of interest from those who complete the study. This imbalance can lead to biases

in the comparison of the treatment groups. As the proportion of missing data increases, the potential for greater bias increases. These biases cannot be overcome by larger sample sizes. In fact, biased results from larger studies can be even more problematic because the larger studies engender greater confidence – in the wrong result.

Missing data may arise in many ways. Intermittent missing data occurs when patients miss a scheduled assessment but attend a subsequent visit. Dropout (withdrawal, attrition) is when patients miss all subsequent assessments after a certain visit. In some trials, when patients are withdrawn from their randomly assigned treatment, no more assessments are taken. In other trials, follow-up assessments may continue. All these settings may lead to missing data, although the statistical approaches appropriate for each setting may vary.

The ICH E9 guideline (www.ich.org/cache/compo/276–254-1.html), which provides the fundamental principles that guide researchers and regulators in medical research, states that despite missing data, a trial may still be valid provided the statistical methods used are sensible. Carpenter and Kenward (2007) define a sensible analysis as one where:

1) The variation between the intervention effect estimated from the trial and that in the population is random. In other words, trial results are not systematically biased.
2) As the sample size increases, the variation between the intervention effect estimated from the trial and that in the population gets smaller and smaller. In other words, as the size of the trial increases, the estimated intervention effect hones in on the true value in the population. Such estimates are called consistent in statistical terminology.
3) The estimate of the variability between the trial intervention effect and the true effect in the population (i.e., the standard error) correctly reflects the uncertainty in the data.

If all these conditions hold, then valid inference can be drawn despite the missing data. However, the analyses required to meet these conditions may be different from the analyses that satisfy these conditions when no data are missing. Regardless, whenever data intended to be collected are missing, information is lost and estimates are less precise than if data were complete.

Table 1.1. *Hypothetical Trial Results*

	Treatment 1	Treatment 2
Success	56	42
Failure	84	98
Missing	60	60
Total	200	200

The following hypothetical data illustrates the ambiguity missing data can cause. Assume Treatment 1 is an investigational intervention or medicine and Treatment 2 is the standard of care. Results for each patient are categorized as success or failure and the outcomes are summarized in Table 1.1.

The success rates based on the observed data are 40% (56/140) for Treatment 1 and 30% (42/140) for Treatment 2. When basing results on only the patients with known outcomes, the success rates are not significantly different ($p = .103$). However, 30% of the outcomes are missing. Table 1.2 summarizes results that would be seen if:

1) It was assumed presence or absence of the observations was not related to the outcome. Hence, in each treatment group unknown outcomes were assumed to have the same proportion of successes as the known outcomes.
2) It was assumed all patients with unknown outcomes were failures.
3) It was assumed unknown outcomes for the investigational drug (Treatment 1) were failures, and unknown outcomes for the standard of care (Treatment 2) had an equal chance of success or failure (50% success).

Table 1.2. *Hypothetical Trial Results Under Different Assumptions About the Missing Outcomes*

	Assumption 1		Assumption 2		Assumption 3	
	Treatment 1	Treatment 2	Treatment 1	Treatment 2	Treatment 1	Treatment 2
Success	80	60	56	42	56	72
Failure	120	140	144	158	144	128

Using assumption 1, results significantly ($p = .046$) favored Treatment 1 because of the increase in sample size – even though the percentages of success and failure did not change compared with the results in Table 1.1. Using assumption 2, results were trending to favor Treatment 1, but the difference was not significant ($p = .130$). Using assumption 3, results approached significance ($p = .108$) in favor of Treatment 2.

These results illustrate how missing outcomes limit the extent to which the trial can inform clinical practice. According to some assumptions the success rate for Treatment 1 was significantly better than Treatment 2, but under other assumptions Treatment 2 was favored. These results also provide the motivation for trial sponsors to limit the amount of missing data. The same rates of success led to non-significance with the missing observations, but would have been significant if no data were missing (assumption 1).

Given the wide range of conclusions that may be drawn based on differing assumptions about the missing data, it may seem that trials with nontrivial amounts of missing data are uninterpretable. With missing data, some information is irretrievably lost, but disregarding the 140 observed outcomes per treatment because 60 outcomes were missing is not an answer.

The extent to which useful information can be gleaned from trials with missing data depends on the amount of missing data, how well the reasons or mechanisms driving the missingness are understood, and how robust conclusions are across the plausible reasons (mechanisms).

Although it is impossible to know with certainty what mechanisms gave rise to the missing data, the extent to which it is understood why data are missing narrows the possibilities. Results can be compared across these various possibilities. Of course, all else equal, the more complete the data the more interpretable the findings.

The importance of reducing missing data and the bias from it is further illustrated in Figure 1.1. This graph depicts the power from the contrast between drug and control from 10,000 simulated clinical trials under three assumptions. In all scenarios the true treatment effect was equal to a standardized effect size (Cohen's D) (Cohen, 1992) of 0.50, and 200 patients were randomized to drug versus control in a 1:1 ratio. The ideal scenario had only 5% dropout and no bias from it. The medium

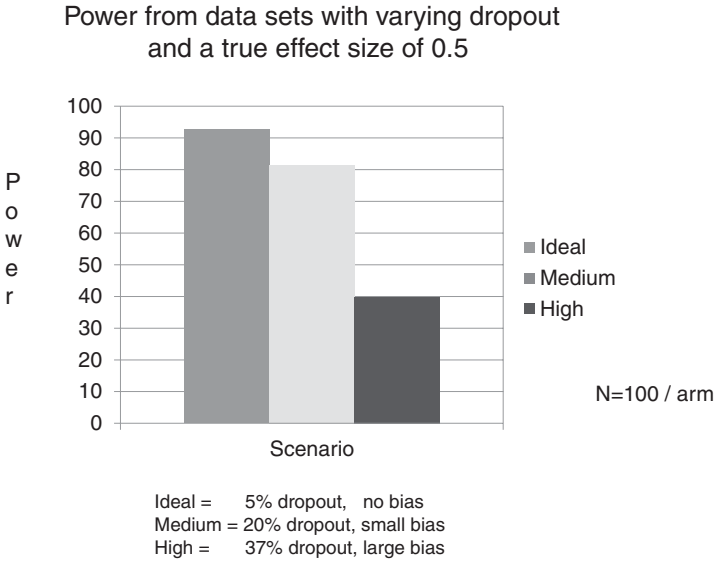

Figure 1.1. Power for the contrast between drug and control from 10,000 simulated clinical trials with low, medium, and high rates of dropout.

scenario had a moderate dropout rate of 20% and minimal bias as the average estimated treatment effect was 0.45, a bit smaller than the true value of 0.50. The high scenario had 40% dropout and appreciable bias as the average estimated treatment effect was only 0.30.

When using a p value of 0.05 as the cutoff for statistical significance, the ideal scenario had 92% power, the moderate scenario had 81% power, and the high dropout scenario had 40% power. In other words, moderate dropout more than doubled (8% vs. 19%) the rate of false negative findings. In the high-dropout scenario, power was reduced to only about 40%, more than doubling the false negative rate over that found with moderate dropout. Dropout can turn a study with a very high probability of success into something less sure than a coin flip.

Modern statistical analyses can reduce the potential for bias arising from missing data. However, principled means of handling missing data rely on untestable assumptions about the missing values and the mechanism giving rise to them (Verbeke and Molenberghs, 2000). The conundrum inherent to analyses of incomplete data is that data about which the missing data assumptions are made are missing. Hence, the

assumptions cannot be tested from data, and the appropriateness of analyses and inference cannot be assured. The greater the rate of missing data, the greater the potential for increased bias. Therefore, minimizing missing data is the best way of dealing with it (Fleming, 2011).

Research on the merits of various analytic approaches for incomplete data includes literally hundreds of scenarios and millions of data sets. Comparing analytic methods can be done using simulated data with known values of the parameters being estimated. Various analytic approaches can be compared within each simulated data set or within each actual clinical trial data set. Hence, much is known about the comparative merits of analytic approaches. Parts III and IV discuss and illustrate some of the common analytic approaches.

A useful by-product of the debates on the appropriateness of various analytic approaches has been consideration of the primary estimand – that is, the primary research question in the clinical trial (see Chapter 3). Obviously, what is being estimated can influence the best way to estimate it. Determining the primary estimand is more complicated than simply stating the primary analysis. Further consideration must be given to a variety of issues. For example, should data after initiation of rescue medications and/or discontinuation of initially randomized study medication be included in the primary analysis?

Sensitivity analyses are a series of analyses with differing assumptions. The aim is that by comparing results across sensitivity analyses it becomes apparent how much inference about the treatment effect relies on the assumptions. In fact, many of the newer statistical approaches are finding their best application as sensitivity analyses rather than as a primary means of analysis (Molenberghs and Kenward, 2007; Mallinckrodt et al., 2008; NRC, 2010; also see section 12.4).

Reasonable measures to reduce missing data combined with appropriate analytic plans that include sensitivity analyses can markedly reduce the uncertainty in results and increase the information gained from medical research. Recent research has provided useful guidance on these various approaches, and the intent of this book is to provide researchers with a practical guide to make use of them.

2

Missing Data Mechanisms

2.1 Introduction

One of the keys to understanding the potential impact of missing data is to understand the mechanism(s) that gave rise to the missingness. However, before considering missing data mechanisms, two important points are relevant. First, there is no single definition of a missing value. Even if restricting focus to dropout (withdrawal), several possibilities exist. For example, values may be missing as the result of a patient being lost to follow-up, with nothing known about treatment or measurements past the point of dropout. Alternatively, a patient may withdraw from the initially randomized study medication and be given an alternative (rescue) treatment, but with no further measurements taken. Or, follow-up measurements may continue after initiation of the rescue treatment. All these and other scenarios may happen within a single trial, with differing implications for appropriate handling of the data (Mallinckrodt and Kenward, 2009).

Moreover, the consequences of missing values are situation dependent. For example, in a clinical trial for diabetes, if a patient is lost to follow-up halfway through the trial, information needed to understand how well the drug worked for that patient is indeed missing. On the other hand, in a trial for a treatment to prevent breast cancer, if a patient dies from breast cancer midway through the trial, follow-up data are again incomplete; however, information about how well the treatment worked for that patient is not missing because it is known that the treatment did not work.

Knowing that missingness (dropout) may or may not be associated with changes in treatment raises the second important point: how the

handling of treatment changes influences outcomes and inferences to be drawn from them. This is, of course, an issue in its own right, but it is also relevant when considering appropriate analyses for incomplete data, especially in the Intention to Treat (ITT) framework that is often the primary basis on which results of confirmatory trials are judged (Mallinckrodt and Kenward, 2009).

While ITT and its alternatives have a direct bearing on the formulation of analyses with missing data, ITT is not a method for handling missing data. Rather, ITT defines the data to be analyzed, and to some extent the inferences drawn from them (Mallinckrodt and Kenward, 2009). An ITT analysis is one in which each patient is assigned for analysis to the treatment group to which he or she was randomized, irrespective of actual subsequent behavior or compliance, and which includes all randomized patients. This definition is the same regardless of whether data are missing or not. The problem, which is discussed in Chapter 3, is how to conform to ITT when some data are missing (Mallinckrodt and Kenward, 2009).

2.2 Missing Data Taxonomy

In order to understand the potential impact of missing data and to choose an appropriate analytic approach for a particular situation, the process (i.e., mechanisms) leading to the missingness must be considered. The following taxonomy of missing data mechanisms is now well established in the statistical literature (Little and Rubin, 2002).

Data are *missing completely at random* (MCAR) if, conditional upon the independent variables in the analysis, the probability of missingness does not depend on either the observed or unobserved outcomes of the variable being analyzed (dependent variable).

Data are *missing at random* (MAR) if, conditional upon the independent variables in the analysis and the observed outcomes of the dependent variable, the probability of missingness does not depend on the unobserved outcomes of the dependent variable.

Data are *missing not at random* (MNAR) if, conditional upon the independent variables in the analysis model and the observed outcomes of the dependent variable, the probability of missingness *does* depend on