

# Chapter 4

## Clinical Trials: Handling the Data

Christina E. Bailey and George J. Chang

### 4.1 Introduction

Randomized control trials (RCTs) provide the foundation for evidence-based medicine, which is the cornerstone of medical practice. RCTs are prospective studies that compare the effect of an intervention between an intervention and control group. An understanding of statistical methods is fundamental to the interpretation of RCT methods and results. This chapter will not provide an in-depth description of the methods of statistical analysis (this information can be obtained from any introductory statistics textbook). Instead, this chapter will provide a brief review of common statistical methods used to analyze data and discuss some issues associated with data analysis.

### 4.2 Who Should Be Analyzed

The first question that should be answered before proceeding with data analysis is which study participants should be included in data analysis. Defining the study population has important implications for the feasibility of the study and generalizability of the results. Unfortunately, even some of the best-designed clinical trials often cannot be perfectly implemented. In retrospect, some participants may not have met the inclusion criteria, data for some participants may be missing, or the protocol may not have been completely followed. Some investigators prefer to eliminate participants who do not adhere to the inclusion criteria or the protocol, whereas other investigators believe that once a participant is randomized, he or she

---

C.E. Bailey, MD, MSCI • G.J. Chang, MD, MS (✉)  
Department of Surgical Oncology, University of Texas, MD Anderson Cancer Center,  
1400 Pressler Street, FCT 17.6060, 301402, Houston, TX 77230-1402, USA  
e-mail: gchang@mdanderson.org

should be included in the final analysis [1]. Both of these views will be discussed later.

Exclusions are potential participants who do not meet all of the entry requirements and are not randomized. Fortunately, exclusions do not bias the results, but it is important to document exclusion criteria in the trial protocol because exclusions can influence interpretation of the results. Participants may also be withdrawn from analysis. Multiple reasons exist for withdrawing participants from the analysis, including ineligibility, nonadherence, and poor quality and/or missing data. Participant withdrawal can bias the results, and it is important to develop a policy on the handling of withdrawals during the design of the trial. Investigators are responsible for convincing readers that the analysis was not biased secondary to participant withdrawal.

### 4.3 Expressing the Data

Before reviewing common statistical methods used to analyze data, we will first review hypothesis testing. Hypothesis testing allows investigators to make generalizations from a sample to the population from which the sample was obtained [2]. The first step in hypothesis testing is stating a null ( $H_0$ ) and alternative ( $H_A$ ) hypothesis. The  $H_0$  states that there is no difference between the hypothesized and population mean, whereas the  $H_A$  states that there is a difference between the hypothesized and population mean. The next step in hypothesis testing is to decide on the appropriate statistical test (reviewed in greater detail below). It is essential to account for random variation in order to conclude that the observed differences in the samples are not due to chance. The p-value estimates the probability of a true difference occurring by chance. If the observed results are highly unlikely (i.e.,  $p < 0.05$ ), we reject the  $H_0$  and accept the  $H_A$ . This means that 5 times out of 100, we will reject the  $H_0$  when it is true (i.e., state there is a difference between two populations when a difference does not exist). This is referred to as type 1 or alpha ( $\alpha$ ) error. Conversely, type 2 or beta ( $\beta$ ) error occurs when an investigator fails to reject the  $H_0$  when it is false (i.e., state there is no difference between populations when a difference does exist). Power ( $1-\beta$ ) is the ability of a study to detect a true difference, and is important in hypothesis testing. Whereas  $\alpha$  error of 0.05 is conventionally accepted,  $\beta$  error of 0.10 or 0.20 is most often used (i.e., power of 90 or 80 %).

An example of utilizing hypothesis testing is as follows. Researchers implemented a trial to determine whether work as a fire fighter affects pulmonary function tests. The study included 50 firefighters, and forced expiratory volume after 1 second ( $FEV_1$ ) was measured before and after a 2-year period on the job. The expected mean decline in  $FEV_1$  over 5 years in normal males is 0.10 L. In this study, the  $H_0$  is the mean decline in  $FEV_1$  will be equal to 0.10 L, and the  $H_A$  is the mean decline in  $FEV_1$  will not be equal to 0.10 L.

### 4.3.1 Comparison of Two Means

The student's  $t$ -test can be used to determine whether the means of two separate groups are equal. Student's  $t$ -test compares the means of two continuous variables and expresses the probability that any differences are due to chance or a "real" difference exists. Data can be obtained from paired or unpaired samples. Paired samples occur when observation are made in the same person (involves a before and after treatment measurement), and unpaired samples occur when observations in one group are independent from observations in another group [3]. There are three assumptions that must be met to utilize the student's  $t$ -test: the data in both groups must follow a normal distribution, the standard deviation (or variance) for both groups is equal, and both groups are independent. Violation of any of these three assumptions can lead to misleading conclusions. If the assumptions are violated, it is recommended that a nonparametric method (Mann–Whitney  $U$  test for unpaired data or Wilcoxon signed rank test for paired data) be used instead. The one-way analysis of variance (ANOVA) is used to compare the means of three or more groups.

An example of the student's  $t$ -test can be illustrated using the previous example comparing FEV<sub>1</sub> among firefighters before and after a 5-year period on the job. As mentioned earlier, the expected mean decline in FEV<sub>1</sub> over 5 years in normal males is 0.10 L. The mean decline in FEV<sub>1</sub> in the 50 firefighters included in the study is 0.2 L, and using student's  $t$ -test to compare means gives a  $p$ -value  $< 0.001$ . One therefore rejects the  $H_0$  and concludes that the observed decline in FEV<sub>1</sub> is significantly different from the expected decline.

### 4.3.2 Comparison of Two Proportions

The chi-square test and Fisher's exact test can be used to compare frequencies or proportions in two or more groups [4]. For example, consider a clinical trial comparing a new treatment (Drug A) to reduce mortality after pulmonary embolus. The primary end point is survival or death. A total of 1,000 patients with pulmonary embolus were randomized to receive Drug A ( $n=525$ ) or placebo ( $n=575$ ). In the treatment group 27 patients (5 %) died and in the placebo group 75 (13 %) died. The data can be displayed in a  $2 \times 2$  table.

	Died	Survived	
Drug A	27	498	525
Placebo	75	500	575
	102	998	1,100

The row totals are the total number of patients receiving Drug A and placebo, whereas the column totals are the total number of patients who died and survived. The chi-square test can be used to determine if there is a statistically significant association between death and treatment with Drug A. The  $H_0$  would be there is no association between death and treatment with Drug A, and the  $H_A$  would be there is an association between death and treatment with Drug A. Using the chi-square test,  $p < 0.001$  therefore rejecting the  $H_0$ , and there is an association between death (or improved survival) with Drug A. Of note, the Fisher’s exact test is used when the expected cell frequencies are  $< 5$ . The expected cell frequency is the probability of being in a given cell times the total sample size. For example, the expected cell frequency for the upper left cell is calculated as  $(525 \times 102) / 1,100 = 48.7$ .

### 4.3.3 Relative Risk and Odds Ratio

The relative risk (RR) is the ratio of the incidence in people with the risk factor (exposed persons) to the incidence in people without the risk factor (nonexposed persons). RR can only be calculated for cohort studies and clinical trials. In both instances, there is a group of subjects with the risk factor and a group of subjects without the risk factor. The subjects are then followed over time to determine which subjects develop the outcome of interest.

The odds ratio (OR) is the odds that a subject with an adverse event was at risk divided by the odds that a subject without an adverse event was at risk. OR can be calculated for cohort and case–control studies. The OR and RR can be easily calculated using a  $2 \times 2$  table.

	Disease	No disease
Treated/exposed	a	b
Control group	c	d
$OR = a * d / b * c$ $RR = \frac{a / (a + b)}{c / (c + d)}$		

For example, a trial was performed comparing thrombotic events in patients taking a nonsteroidal anti-inflammatory drug (NSAID) compared to placebo. In the NSAID group, 46 out of 1,000 patients had a thrombotic event compared to 26 out of 1,000 patients in the placebo group.

	Thrombotic event		Total
	Yes	No	
NSAID	46	954	1,000
Placebo	26	974	1,000

The calculated OR would be 1.81  $[(46 \times 974)/(954 \times 26)]$ . This can be interpreted as patients taking the assigned NSAID have 1.81 increased odds of having a thrombotic event compared to patients taking placebo. The calculated RR  $[(46/1,000)/(26/1,000)]$  is 1.76. This can be interpreted as patients taking the assigned NSAID have a 76 % increase in the rate of thrombotic events compared to patients taking placebo.

OR or RR greater than 1 indicates that there is an increased risk of the measured event associated with the exposure. When the OR or RR equals 1, the measured event is no more likely to occur with or without the exposure. On the other hand, when the OR or RR is less than 1, the measured event is less likely to occur with the exposure [5]. Also of note, in the previous example, OR and RR approximate each other, 1.81 and 1.76, respectively. This is usually true when the event rates are low and/or the treatment effect is small.

Other terms to be familiar with include absolute risk reduction, number needed to treat, absolute risk increase, and relative risk reduction. The absolute risk reduction allows one to assess the reduction in risk compared with the baseline risk. Specifically, it is the reduction in risk of a new intervention compared to the risk without intervention, and it is the absolute value of the difference between the experimental and control event rates. The number needed to treat is the reciprocal of the absolute risk reduction and provides the number needed to treat in order to prevent one event. For example, if a new treatment decreases the relative risk of myocardial infarction and has an absolute risk reduction of 0.0086, then the number of people who need to be treated to prevent 1 myocardial infarction is approximately 116  $(1/0.0086=116.3)$ . Absolute risk increase is the opposite of the absolute risk reduction. It is the increase in risk with a new treatment compared with the risk without the treatment, and relative risk reduction is the reduction in risk with a new treatment relative to the risk without treatment [4].

### 4.3.4 Correlation and Linear Regression

Correlation is used to determine if a linear relationship exists between two quantitative variables. Linear correlation is a measure of the degree to which an increase or decrease in one continuous variable is associated with a proportional increase or decrease in a second continuous variable [6]. In other words, can the relationship between two variables be described by a straight line? For example, consider a

scatterplot depicting the hemoglobin A1c and serum glucose in ten patients with diabetes mellitus. If every point falls on a straight line, the two variables are perfectly correlated. The Pearson correlation coefficient ( $r$ ) can be used to calculate the strength of a relationship and ranges from  $-1$  to  $+1$ . A value of  $0$  represents no correlation,  $-1$  represents perfect negative correlation, and  $+1$  represents perfect positive correlation between two variables. The Pearson's correlation coefficient can be calculated for any dataset, but it is more meaningful if the two variables are normally distributed.

Linear regression allows investigators to analyze the relationship between two or more continuous variables when one variable depends on the others and allows investigators to predict one variable given the value of the other variables [3]. For example, investigators were interested in the relationship between height and forced expiratory volume (FEV) in children. Using linear regression, it was found that  $FEV = -6.07 + (0.14 \times \text{height})$ . Using this equation, the predicted FEV for a five foot ( $60''$ ) child would be  $2.34 \text{ l} (-6.06 + 0.14 \times 60)$ . Of note, when performing multivariate analysis (i.e., more than two variables are included in the model), the number of covariates used in the model depends on the sample size. Ideally the sample size should exceed ten times the number of independent variables. For example, if the sample size in a study is  $100$ , no more than ten independent variables should be included in the linear regression model. If too many independent variables are included in the model, investigators run the risk of overfitting the data. The same is also true for small sample size. Also, assumptions must be met in order to utilize linear regression models. They are as follows: the sample must be randomly selected,  $X$  and  $Y$  are normally distributed, and the  $Y$  values are independent of each other (i.e., not correlated).

### 4.3.5 *Survival Analysis*

Survival analysis is also referred to as time to event analysis. It allows for the analysis of binary categorical outcomes such as death, onset of disease, recurrence of disease, and onset of disability. Survival can be reported as a percentage (i.e., 1-year or 5-year survival), median survival, or survival curves. There are several different survival analysis methods: incidence density method, life table (actuarial method), Kaplan-Meier (product-limit method), and Cox Proportional Hazards model. Kaplan-Meier and Cox Proportional Hazards are the more commonly used survival methods for clinical trials. Kaplan-Meier survival analysis allows investigators to generate survival curves for each group which can be compared using the logrank statistic. Kaplan-Meier survival analysis is to be considered generally reliable up to two times the median follow-up time. Assumptions made when utilizing the Kaplan-Meier method include no change in the event rate over time and the outcome is the same for patients that are followed and those lost to follow-up. The Cox Proportional Hazard model provides a hazard ratio and allows for the comparison of two or more survival curves after adjusting for covariates. For example, a multi-institutional retrospective study identified  $3,500$  patients who underwent pancreatic resection for

pancreatic cancer. A multivariate-adjusted Cox Proportional Hazard model was used to evaluate the prognostic significance of adjuvant radiation therapy (AXRT). The hazard ratio for patients who received AXRT was 0.75. This can be interpreted as patient who received AXRT after surgical resection of pancreatic cancer had a 25 % decreased risk of death compared to patients who did not receive AXRT. During covariate adjustment, in general the ratio of the number of independent variables used in the Cox model to number of events should not exceed 1:10. For example, if a study has a sample size of 1,000 and 100 patients died, the maximum number of independent variables that should be included in the model is 10. Similar to linear regression, if too many independent variables are included in the model, one is at risk of overfitting the data. Assumptions for Cox regression are the same as Kaplan-Meier survival analysis, and the effect of the covariate does not change over time for any of the independent variables.

## 4.4 Analyzing the Data

Careful analysis of data obtained from clinical trials requires a major investment of time and effort. Inappropriate statistical analysis can result in misleading conclusions and impairs the credibility of the trial and investigators. Two important issues that should be considered in the analysis of clinical trial results are intention-to-treat analysis and the role for subgroup analysis.

### 4.4.1 *Intention-to-Treat Analysis*

Intention-to-treat (ITT) analysis is a technique commonly used in randomized control trials. The definition is as follows: “All patients randomly allocated to one of the treatments in a trial should be analyzed together as representing that treatment, whether or not they completed, or indeed received that treatment” [7]. In other words, ITT compares outcomes between study groups with each participant analyzed according to their randomized group assignment regardless of receiving the assigned treatment, withdrawal from the study, or deviation from the protocol.

An alternative to ITT is “per protocol” analysis, which only evaluates those participants who complied with the assigned treatment. This appears to be an appropriate approach to analysis because participants can only be affected by an intervention they actually received. However, the problem arises when participants who adhere to the study treatment differ from those that are noncompliant or drop out, thus introducing bias [8]. For example, in the Postmenopausal Estrogen/Progestin Intervention (PEPI) Trial, 875 healthy postmenopausal women aged 45–64 years of age who had no known contraindication to hormone therapy were randomly assigned to four different estrogen or estrogen plus progesterone regimens and placebo. Of the 175 women assigned to the unopposed estrogen arm, 41 (23 %) discontinued treatment because of

endometrial hyperplasia, which is a precursor of endometrial cancer [9]. If “per protocol” analysis was performed, these women would have been eliminated from analysis, and the association of estrogen therapy and endometrial cancer may have been missed.

ITT analysis not only minimizes bias but it also maintains the similarities between treatment groups in regard to prognosis. This is the reason for randomization, and this feature may be lost if analysis is not performed on the groups produced by the randomization process. This can be illustrated by the European Coronary Surgery Study Group Trial comparing medical and surgical treatment for stable angina. A total of 768 men under the age of 65 with angina were included in the study (373 men were randomized to medical treatment, and 395 men were randomized to surgical treatment). A total of 26 men assigned to the surgical arm did not undergo surgery, and 50 men assigned to the medical arm underwent surgery. Using ITT analysis, there was no significant difference in mortality between the two groups at 2 years [10]. Alternatively, using “per protocol” analysis, the mortality rate would be 8.4 % for the medical treatment and 4.1 % for the surgical treatment ( $P=0.018$ ) [7]. In “per protocol” analysis, surgery appears to have a falsely low mortality rate.

Despite the advantages of ITT analysis, the major disadvantage is that participants who choose not to take the assigned intervention will be included in the estimate of the effects of that intervention. There is potential for the magnitude of the effect of the treatment to be underestimated if there are a significant number of participants who “cross over” between treatments. For this reason, results of trials are often evaluated using both ITT and “per protocol” analysis, and if both analyses have similar results, the confidence in the trial conclusions is increased. However, if the results from the two analyses are different, the results of ITT analyses dominate because randomization is preserved and bias minimized.

The utilization of ITT analysis has increased over the years. In 1999, Hollis et al. surveyed all reports of randomized controlled trials published in 1997 in the *BMJ*, *Lancet*, *JAMA* and *New England Journal of Medicine*. A total of 119 (48 %) trials mentioned ITT analysis. Of these, 12 trials excluded any patients who did not start the allocated intervention, and three trials did not analyze all randomized subjects as allocated. The authors concluded that the ITT approach is often inadequately described and inadequately applied, and readers should critically assess the validity of reported ITT analysis [11]. More recently, Gravel et al. conducted a cross-sectional literature review of randomized control trials reported in ten medical journals in 2002. Of the 403 articles, 249 (62 %) reported the use of ITT. Among these, 192 (77 %) clearly analyzed patients according to the groups to which they were randomized. Authors used a modified ITT approach in 23 (9 %) and clearly violated a major component of ITT in 17 (7 %). The approach used in 17 (7 %) was unclear [12].

#### **4.4.2 Subgroup Analysis**

Clinical trials are labor intensive and costly, and investigators often use subgroup analysis to extract as much information as possible regarding the effect of a particular

treatment. Subgroup analysis compares subsets of randomized participants. Specifically, investigators compare the treatment effect between one and more subgroups rather than the entire cohort of participants. Subgroups are usually defined based on baseline characteristics. Using subgroup analysis, investigators have the potential to determine in which participants a specific treatment is more (or less) effective (or harmful). For example, a double-blind, placebo-controlled trial was conducted in which the reduction in the incidence of death or hospitalization for cardiovascular reasons with the use of the beta-blocker carvedilol was compared to placebo in patients with heart failure. In subgroup analysis, the investigators further examined whether carvedilol decreased the incidence of cardiovascular events according to the patients' severity of disease, age, sex, left ventricular ejection fraction, 6-min walk, cause of congestive heart failure, systolic blood pressure, and heart rate. Patients treated with carvedilol had a 65 % lower risk of death than those given placebo, and the beneficial effect of carvedilol on survival was consistent in all evaluated subgroups [13].

Even though subgroup analysis allows investigators to identify who, if anyone, benefits from an intervention, care must be utilized in the interpretation of subgroup findings. There are several issues that arise during subgroup analysis [14]:

1. Most trials are not sufficiently powered to detect a difference between treatment groups. Subgroups are by definition smaller than the entire trial cohort. Therefore, if a difference does exist between subgroups, it may not be detected because the size of the trial is not large enough. Investigators may also examine results in a large number of subgroups, thus increasing the likelihood that a difference in treatment effect in a subgroup may be due to chance (type I error).
2. A number of subgroups can be identified based on baseline characteristics, and subgroups can be specified either before or after examination of the data. These two methods are referred to as prespecified subgroup analysis and post hoc analysis, respectively. Prespecified subgroup analysis is planned and documented before any data analysis is performed. Post hoc analysis is often referred to as "data dredging" or "fishing" and can be of particular concern because it can be unclear how many subgroups were analyzed and whether some subgroups were identified secondary to inspection of the data [15].
3. Statistical tests for interaction examine the strength of treatment differences between varying subgroups. This is the best method for making inferences from subgroup analysis. Tests for interaction take into consideration that data available for subgroup analysis is limited. Even though tests for interactions protect investigators from making false or premature claims from subgroup analysis, the test is not routinely used. In a survey of 50 trial reports in 4 major journals conducted by Pocock et al. in 2002, only 15 (43 %) of the 35 reports with subgroup analysis used tests for interaction [14]. A common mistake made by investigators is presenting separate  $p$ -values for treatment differences within each subgroup. For example, testing the hypothesis that there is no treatment effect in patients younger than 50 years of age and then testing the hypothesis separately in patients older than 50 years of age does not address whether treatment differences vary according to age. Separate subgroup  $p$ -values can be misleading.

4. The results of subgroup analysis are often overinterpreted by authors and readers, and caution has to be exercised when drawing conclusions. In the same survey conducted by Pocock et al., 21 trials (42 %) claimed to find differences in subgroups that were not compatible with the overall treatment comparison, and 13 of these featured these claims in the summary and/or conclusion [14]. As readers analyze trials that utilize subgroup analysis, biological plausibility, the number of subgroup analyses performed, prespecification of the subgroups, and the trial size have to be considered when drawing conclusions.

Treatment decisions in multiple fields of medicine are directed by the results from randomized clinical trials (RCTs). One field in which there have been hundreds of RCTs is cardiology. Hernandez et al. reviewed 63 cardiovascular RCTs published from 2002 to 2004 in major medical journals. Of the selected RCTs, 39 reported subgroup analysis, and 26 had more than 5 subgroups. Only 14 (35.8 %) prespecified the subgroups, and only 11 (28 %) reported interaction tests. The authors concluded that the reporting of subgroup analysis in cardiovascular RCTs had several shortcomings, including lack of prespecification and testing of a large number of subgroups without the use of tests for interactions. Based on these results, the authors made several recommendations to appropriately perform and interpret subgroup analysis [16]:

1. Specify subgroups in advance with a clear rationale.
2. Use statistical tests for interaction in the full RCT population.
3. Be skeptical if subgroups were not prespecified, not biologically plausible, or no interaction test was performed.
4. Utilize subgroup analysis as a hypothesis-generating tool for future studies.
5. Emphasis should be placed on the overall results, which for the most part are better estimates of treatment effects compared to subgroup effects.

In summary, subgroup analysis is important in clinical trials. The results of subgroup analysis can be used to generate a hypothesis for future studies, but the results must be interpreted with caution, and broad, general conclusion statements should not be made based on subgroup analysis.

## 4.5 Handling Missing Data

Missing data is a serious problem and has the potential to compromise conclusions drawn from clinical trials. Missing data is defined as “values that are not available and that would be meaningful for analysis if they were observed” [17]. It occurs when participants drop out of a study before its conclusion. Dropout can be secondary to treatment or analysis dropout. Treatment dropout occurs when the assigned treatment is terminated, and analysis dropout occurs when some study measurements are not recorded [18]. If dropout is secondary to the intervention, whether it is treatment dropout or analysis dropout, bias can be introduced into the analysis.

Unfortunately, limited information is available on how to handle missing data. Wood et al. reviewed all randomized trials published between July and December 2001 in the *British Medical Journal*, *Journal of the American Medical Association*, *Lancet*, and *New England Journal of Medicine*. They focused on trial design and how missing outcome data was described and how statistical methods were used to deal with missing data. The conclusion of their review was that missing outcome data is a common problem in randomized controlled trials and it is often inadequately handled in the statistical analysis [19]. To help address this problem, the National Research Council convened the Panel on the Handling of Missing Data in Clinical Trials at the request of the Food and Drug Administration (FDA). The objective of the panel was to prepare “a report with recommendations that would be useful for FDA’s development of a guidance for clinical trials on appropriate statistical methods to address missing data for analysis of results” [20]. The recommendations of the panel are summarized below.

The first step in minimizing missing data occurs during the design of the clinical trial. Every effort should be made to clearly define the target population and outcome measures prior to the initiation of the trial. The trial should be designed to maximize adherence to the protocol and ensure participants adhere to follow-up visits and measurements. Little et al. published several suggestions (adopted from the Panel on the Handling of Missing Data in Clinical Trials) for limiting missing data in the design of clinical trials [17, 18]:

1. Target a population that is not adequately served by available treatments and thus have incentive to remain on the study.
2. Include a run-in period in which all participants are initially placed on active treatment. After a specified time, the participants who were adherent to the therapy are randomized to continue active treatment or begin placebo.
3. Allow flexibility in the treatment regimen in order to reduce the dropout rate because of a lack of efficacy or treatment intolerance.
4. Consider add-on designs (study treatment is added to an existing treatment).
5. Shorter follow-up periods for the primary outcome.
6. Allow the use of rescue medications.
7. Consider a randomized withdrawal design to assess long-term efficacy (participants who have received study treatment without dropping out are randomized to continue to receive the treatment or switch to placebo).
8. Try to avoid using outcome measures that are likely to lead to substantial missing data.

Another important factor to take into consideration during the design phase is how missing data will affect the power of the trial. Most investigators “inflate” the initial sample size to account for anticipated missing data.

Even when investigators take every step to minimize missing data during the design of the trial, every participant will not follow their assigned intervention to the completion of the trial. The question then presents itself in regard to which data, if any, should be collected for participants who do not complete the assigned treatment. Investigators have two opposing views. Some believe that participants who

do not complete the assigned treatment are no longer relevant to the study. The opposing view is that continued data collection may be informative and can potentially allow for the ability to analyze end points for all participants and explore whether assigned therapy effects the use and efficacy of subsequent therapies [18].

After taking steps to minimize missing data during the design of the clinical trial, attention must then be turned to minimizing missing data during the conduct of the trial. Little et al. once again have several suggestions (also adopted from the Panel on the Handling of Missing Data in Clinical Trials) for limiting missing data during conduct of the trial [17, 18]:

1. Select investigators who have good track records enrolling participants, following participants, and collecting complete data.
2. Set acceptable rates for missing data in the study protocol.
3. Provide incentives (as long as they comply with ethical requirements) to investigators and participants for completeness of data collection.
4. Minimize the participant inconvenience and burden associated with data collection.
5. Provide effective treatment to participants after the trial.
6. Train investigators and their research staff on the negative impact of missing data.
7. Train investigators and their research staff on the informed consent process as a tool for encouraging complete data.
8. Monitor missing data during the trial.

Unfortunately, there is no universal method for handling missing data during data analysis. The Panel on the Handling of Missing Data in Clinical Trials identified four different methods to adjust for missing data: complete-case analysis, single imputation methods, estimating equation methods, and methods based on a statistical method. Complete-case analysis excludes participants with missing data from the analysis. Single imputation methods fill in a value for each missing value using methods such as the last observation or baseline observation carried forward. Estimating equation methods weigh complete cases by the inverse of an estimate of the probability of being observed, and methods based on statistical methods include maximum likelihood, Bayesian methods, and multiple imputations. In general, the panel favored estimating equation methods and methods based on a statistical model for the data [17].

In summary, missing data is a major issue that has to be addressed in the design and analysis of clinical trials. Missing data can lead to bias and affect the interpretation of trial results. Therefore, it is important to try to minimize missing data during the design and conduct of clinical trials.

## 4.6 CONSORT Statement

In an effort to facilitate the interpretation of data from randomized trial and to facilitate their complete and transparent reporting such that some of the issues described above can be deliberated during the interpretation of the results, scientists and

editors have developed the Consolidated Standards of Reporting Trials (CONSORT) statement [21]. It is comprised of a 25-item checklist and a flow diagram focusing on reporting how the trial was designed, analyzed, and interpreted. While use of the CONSORT statement improves the communication of the study design and its findings, it is important to understand these key issues for the interpretation of data.

## 4.7 Conclusion

RCTs provide the foundation for evidence-based medicine. The design and implementation of RCTs are labor extensive and expensive; therefore, it is important that investigators have a clear understanding of the design and implementation of clinical trials and the analysis of the results obtained during a clinical trial. This chapter provided a brief review of common statistical methods utilized to express the results of clinical trials and common issues associated with data analysis.

## References

1. Friedman LM, Furberg C, DeMets DL. *Fundamentals of clinical trials*. 3rd ed. New York: Springer; 1998. xviii, 361 p.
2. Dawson B, Trapp RG. *Basic & clinical biostatistics*. 4th ed. New York: Lange Medical Books-McGraw-Hill, Medical Pub. Division; 2004. x, 438 p.
3. Wang D, Bakhai A, Maffulli N. A primer for statistical analysis of clinical trials. *Arthroscopy*. 2003;19(8):874–81.
4. Dawson B, Trapp RG. *Basic & clinical biostatistics*. 4th ed. New York: Lange Medical Books/McGraw-Hill; 2004.
5. Straus SE. *Evidence-based medicine : how to practice and teach it*. 4th ed. Edinburgh: Elsevier Churchill Livingstone; 2011. xvii, 293 p.
6. Troidl H. *Principles and practice of research: strategies for surgical investigators*. Berlin/New York: Springer; 1986. xvii, 380 p.
7. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol*. 1992;21(5):837–41.
8. Hulley SB. *Designing clinical research: an epidemiologic approach*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2001. xv, 336 p.
9. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. The Writing Group for the PEPI Trial. *JAMA*. 1995;273(3):199–208.
10. Coronary-artery bypass surgery in stable angina pectoris: survival at two years. European Coronary Surgery Study Group. *Lancet*. 1979;1(8122):889–93.
11. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319(7211):670–4.
12. Gravel J, Opatny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials*. 2007;4(4):350–6.
13. Packer M, et al. The effect of carvedilol on morbidity and mortality in patients with chronic heart failure. U.S. Carvedilol Heart Failure Study Group. *N Engl J Med*. 1996;334(21):1349–55.
14. Pocock SJ, et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21(19):2917–30.

15. Wang R, et al. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med.* 2007;357(21):2189–94.
16. Hernandez AV, et al. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J.* 2006;151(2):257–64.
17. Little RJ, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med.* 2012;367(14):1355–60.
18. Little RJ, et al. The design and conduct of clinical trials to limit missing data. *Stat Med.* 2012;31(28):3433–43.
19. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials.* 2004;1(4):368–76.
20. Council NR. The prevention and treatment of missing data in clinical trials. Washington, DC: National Academic Press; 2010.
21. Moher D, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol.* 2010;63(8):e1–37.