

The design and conduct of clinical trials to limit missing data

R. J. Little,^{a,*†} M. L. Cohen,^b K. Dickersin,^c S. S. Emerson,^d
J. T. Farrar,^e J. D. Neaton,^f W. Shih,^g J. P. Siegel^h and H. Sternⁱ

This article summarizes recommendations on the design and conduct of clinical trials of a National Research Council study on missing data in clinical trials. Key findings of the study are that (a) substantial missing data is a serious problem that undermines the scientific credibility of causal conclusions from clinical trials; (b) the assumption that analysis methods can compensate for substantial missing data is not justified; hence (c) clinical trial *design*, including the choice of key causal estimands, the target population, and the length of the study, should include limiting missing data as one of its goals; (d) missing-data procedures should be discussed explicitly in the clinical trial protocol; (e) clinical trial *conduct* should take steps to limit the extent of missing data; (f) there is no universal method for handling missing data in the *analysis* of clinical trials – methods should be justified on the plausibility of the underlying scientific assumptions; and (g) when alternative assumptions are plausible, sensitivity analysis should be conducted to assess robustness of findings to these alternatives. This article focuses on the panel's recommendations on the design and conduct of clinical trials to limit missing data. A companion paper addresses the panel's findings on analysis methods. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: clinical trial ethics; dropouts; incomplete data; randomized withdrawal; run-in period

1. Introduction

Missing data is an important problem that potentially compromises the validity of treatment comparisons in clinical trials, because missingness may be related to the drug's effectiveness, safety, or patient prognosis [1]. Existing guidelines [2–5] for the design and conduct of clinical trials and the analysis of the resulting data provide only limited advice on how to handle missing data. At the request of the Food and Drug Administration, the National Research Council convened the Panel on the Handling of Missing Data in Clinical Trials to prepare 'a report with recommendations that would be useful for FDA's development of guidance for clinical trials on appropriate study designs and follow-up methods to reduce missing data and appropriate statistical methods to address missing data for analysis of results.' This article summarizes the main findings of that report [6] concerning the design and conduct of trials.

Despite best efforts, missing data are often inevitable, so the appropriate analysis of clinical trials when there are missing values remains an important issue. The panel advocates a principled approach to analysis, on the basis of careful attention to assumptions about the nature of missing data mechanisms. A companion article [7] addresses this aspect of the study.

^aDepartment of Biostatistics, University of Michigan, MI, U.S.A.

^bCommittee on National Statistics, National Academy of Sciences, Washington DC, U.S.A.

^cDepartment of Epidemiology, Johns Hopkins University, MD, U.S.A.

^dDepartment of Biostatistics, University of Washington, Seattle, WA, U.S.A.

^eDepartment of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, PA, U.S.A.

^fSchool of Public Health, University of Minnesota, MN, U.S.A.

^gDepartment of Biostatistics, University of Medicine and Dentistry of New Jersey School of Public Health, NJ, U.S.A.

^hJohnson & Johnson, Radnor, PA, U.S.A.

ⁱDepartment of Statistics, University of California, Irvine, CA, U.S.A.

*Correspondence to: Roderick J. Little, Department of Biostatistics, University of Michigan, 1415 Washington Hgts, Ann Arbor, MI 48109, U.S.A.

†E-mail: rlittle@umich.edu

Randomized clinical trials are the primary tool for evaluating new medical interventions. Randomization provides for a fair comparison between the treatment and control groups, balancing out, on average, distributions of known and unknown factors that have the potential to influence outcomes in the treatment arms. When a substantial fraction of the measurements of the outcome of interest are missing, the benefits of randomization are dissipated, treatment comparisons are potentially biased, and power is compromised. Increasing the sample size to compensate for the missing data reduces random error and increases power, but it does not reduce systematic errors that might bias treatment comparisons. We focus primarily on ‘phase III’ confirmatory clinical trials that are the basis for the approval of drugs and devices, where the bar of scientific rigor is set high. However, many of the recommendations are applicable to randomized trials and epidemiologic studies in general.

Missing data in clinical trials often arise when participants drop out of the study before its conclusion. It is important to distinguish between two kinds of dropout, ‘treatment dropout’ or discontinuation, where an individual goes off the study protocol, for example, by terminating an assigned treatment; and ‘analysis dropout’, where some of the measurements in a study are not recorded, for example, because of failure to attend study visits where measurements are taken [8]. Often these types of dropout are related, in that individuals who go off their assigned treatment are not followed-up, so that further measurements are not taken. However, this does not have to be the case, as discussed below.

There is no ‘foolproof’ way to analyze trial data with substantial amounts of missing data. It is unlikely that participants who drop out of a trial are representative of all those in the trial, and no analysis method recovers the potential for robust treatment comparisons derived from complete follow-up of all randomized participants. Hence, we emphasize that the key is trial design and conduct to limit the amount and impact of missing data. Design options to limit missing data are focus of Section 2. The amount of missing data varies greatly depending on the extent to which efforts to prevent and limit missing data is a priority in the conduct of a trial; Section 3 describes some practical steps in the conduct of a trial that can reduce the incidence of missing data. Section 4 summarizes the Panel’s findings.

The study report uses three types of trials as case studies to illustrate different aspects of missing data: trials for chronic pain, trials for the treatment of HIV, and trials for mechanical circulatory devices for severe symptomatic heart failure. These will be referenced throughout the article.

Trials for chronic pain are used to assess the ability of an intervention to provide symptomatic pain relief from conditions such as osteoarthritis. A typical trial involves a 12-week maintenance period following dose titration, and the efficacy measures are typically pain scores assessed periodically by the patient as well as by the investigator [9, 10]. These types of clinical trials are subject to high rates of treatment discontinuation, and the reasons may differ between the treatment and control groups. For example, in placebo-controlled trials, discontinuation in the placebo group may stem from inadequate efficacy (i.e., lack of pain relief), whereas discontinuation in the treatment group may arise because the treatment is not well tolerated [10]. To further complicate matters, participants who stop study treatment may switch to a therapy considered effective. Trial sponsors typically stop collecting pain response data on participants who discontinue the assigned treatment.

Trials for the treatment of HIV have the goal of determining whether a new drug has comparable safety and efficacy with that of an approved drug used for initial antiretroviral treatment (ART). The studies typically involve participants not previously receiving ART [11]. Because combination ART is an existing effective treatment that is the norm for HIV, the typical design in this setting is an ‘add-on’ design, comparing ART plus a new drug A with same ART plus a current drug B, measured over a period of 24 or 48 weeks. A common primary efficacy outcome is treatment success at a fixed follow-up time, such as 24 or 48 weeks of follow-up. Treatment failure is generally a composite outcome defined to include study participants who (1) die; (2) discontinue the study drug for lack of efficacy or toxicity before the closeout visit; (3) are lost to follow-up (e.g., do not attend the prespecified visit time point; or (4) remain on the study drug but have a predefined unacceptably high viral load, as measured by plasma RNA measurement, at the end of the reference period. These trials can have moderate to large numbers of patients who either discontinue treatment before the end of follow-up or who do not attend the designated follow-up time point (e.g., 24 or 48 weeks) [12]. Participants are typically not followed-up after discontinuing treatment, making comparisons of different components of the composite failure outcome difficult. Moreover, information is not collected on effects of discontinuing the study drug on the performance of subsequent therapies.

Trials for mechanical circulatory devices for severe symptomatic heart failure. Heart implantable left ventricular assist devices (LVADs) have been shown to be effective for patients with advanced heart failure, as a bridge to heart transplantation or as a destination therapy for patients who are not eligible

for transplantation; its use is increasing with the development of smaller and more durable devices. We consider a superiority design trial in which the goal is to determine whether an LVAD substantially improves functional status compared with medical management, without negatively impacting survival. As their disease progresses, some patients assigned medical management may receive a heart transplant or an approved LVAD over the course of the study, complicating the interpretation of functional measures for treatments.

Survival status is ascertained for nearly all patients, but missing health status data are often major problems [13]. Functional status during follow-up may be missing because of early death (for example, as a consequence of the implantation procedure), failure to attend examinations, inability to perform functional tests such as a 6-min walk, or ‘questionnaire fatigue’ for self-administered quality-of-life instruments. Item nonresponse in quality of life instruments is also a problem. Furthermore, the inability to mask observers to the treatment hinders the objective assessment of health status measures.

2. Trial design and missing data

2.1. Consider the potential for missing data when defining the primary estimand

Good clinical trial design should define clearly the target population and outcomes that will form the basis for decisions about efficacy and safety. The treatment of missing data depends on how outcomes are defined, and lack of clarity in their definition translates into a lack of clarity in how to deal with missing data issues. Given the difficulties of adequately addressing missing data at the analysis stage, trials should be designed to limit missing values, in a manner consistent with the trial objectives. In the words of Recommendation 1 of the panel report [6, p. 26]:

The trial protocol should explicitly define (a) the objective(s) of the trial; (b) the associated primary outcome or outcomes; (c) how, when, and on whom the outcome or outcomes will be measured; and (d) the measures of intervention effects, that is, the causal estimands of primary interest. These measures should be meaningful for all study participants, and estimable with minimal assumptions. Concerning the latter, the protocol should address the potential impact and treatment of missing data.

For illustration, we describe five possible estimands. For simplicity, we consider a trial comparing a new and control treatment. Outcome refers to the primary outcome of the trial (e.g., symptom or pain relief, a surrogate marker such as HIV RNA level). The ‘duration of protocol adherence’ is the time after randomization for which a participant received the study intervention according to protocol. For the purposes of this example, this time is assumed to be no longer than the planned study duration.

- (1) *(Difference in) mean outcome improvement for all randomized participants.* This so-called ‘intention-to-treat’ estimand assesses the overall benefits of a treatment policy or strategy relative to a control in the population of all potentially randomized participants. Other comparison measures, such as the relative risk or odds ratio, are also possible. Because the estimand relates to a treatment policy, observed differences reflect the effect of the initially assigned treatment as well as subsequent treatments adopted as a result of intolerance or lack of efficacy. A trial design that supports the use of this estimand is a parallel-group randomized trial in which outcome data are collected on all participants, regardless of whether the assigned treatment is received. A trial design that does not support the use of this estimand is a parallel-group randomized trial in which outcome data are not collected on participants after they stop taking the assigned treatment or switch to a different treatment.
- (2) *(Difference in) outcome improvement in those who adhere to treatment.* This estimand quantifies the degree of outcome improvement in the participants who tolerate and adhere to a particular treatment. A challenge with this estimand is that it is difficult to identify in advance the members of this subpopulation, and assessed performance in a trial may therefore be an overestimate of performance in practice. An example of a design for this estimand is a parallel-group randomized trial with an active treatment run-in period followed by placebo washout prior to randomization, limited to individuals who adhered to the active treatment during the run-in period. Outcome data are then collected on all randomized participants regardless of treatment received and subsequent adherence.
- (3) *(Difference in) outcome improvement if all participants had adhered.* This estimand quantifies the degree of outcome improvement in all participants in the trial, assuming they all received treatment according to the protocol for the planned study duration. This estimand requires imputation

of what would have been the outcome in individuals who did not adhere with the protocol, had they adhered. The question of whether this estimand is reasonable depends on the degree to which the lack of adherence is avoidable, because otherwise this estimand measures the effects of an infeasible treatment policy. A trial design for this estimand is a parallel-group randomized design in which participants are provided adjunctive or supportive therapies along with the study treatments to ensure that they are tolerated. This design assumes that such adjunctive therapies to enhance tolerability are available and do not have a direct effect on the outcome of interest. Outcome data are collected on all participants.

- (4) (*Difference in*) *areas under the outcome curve during adherence to treatment.* This estimand compares the arm-specific means of the area under the outcome curve over the duration of protocol adherence. This estimand simultaneously quantifies the effect of treatment on both the outcome measure and the duration of tolerability or adherence in all participants. A trial design that supports the use of this estimand is a parallel-group randomized trial with or without a run-in period. In such a trial with this estimand, there is no need to collect outcome data after an assigned treatment is discontinued or the participant switches to an alternative, other than to address secondary analysis issues such as delayed side effects.
- (5) (*Difference in*) *outcome improvement during adherence to treatment.* This estimand is the difference in mean outcomes from the beginning of the trial to the end of the trial or the end of adherence to the protocol, whichever occurs earlier. This estimand reflects both the duration of tolerability or adherence and outcome improvement in all participants. Estimating the treatment effect on the primary outcome does not require collection of outcome data after the assigned treatment is discontinued.

Estimands that combine features of pharmacological efficacy with tolerance and adherence, such as (4) area under the outcome curve during adherence to treatment or (5) outcome improvement during adherence to treatment, have the potential for misinterpretation. In particular, such estimands may not distinguish between (a) an immediately highly effective but toxic treatment with a short period of tolerability and (b) a nontoxic treatment with gradual outcome improvement and a long period of tolerability.

As the aforementioned discussion makes clear, there are a wide range of estimands that can be considered in a trial, and each will involve tradeoffs between the representativeness of the population of study, the ease of study design and execution, and the sensitivity to missing data.

2.2. Design strategies to minimize treatment and analysis dropouts

Recommendation 2 of the panel report [6, p. 29] states that

Investigators, sponsors, and regulators should design clinical trials consistent with the goal of maximizing the number of participants who are maintained on the protocol-specified intervention until the outcome data are collected.

The following design elements for clinical trials can help to reduce the number of participants who drop out as a result of lack of tolerability, lack of efficacy, or inability to provide the required measurement. Each of the design elements has disadvantages that should be considered.

- (1) *Include a run-in period,* after which only individuals who tolerated and adhered to therapy are randomized to a treatment. Such a design may result in a more efficient study with less missing data, but likely will not adequately estimate the rate of adverse events in the broader population, and involves loss of some external validity. The population ultimately examined in trials using a run-in is different from the population that otherwise would have been tested. For example, a run-in is usually designed to increase the difference in effect that is observed between placebo and active treatment, assuming a real difference exists, but may no longer be truly representative of the population in whom the treatment will be used. A related idea is enrichment designs, which exclude participants on the basis of initial indications that the study treatment has a weak response or is difficult to tolerate.
- (2) *Adopt flexible dosing* that accommodates individual differences in tolerability, allowing more participants to continue on the assigned treatment by reducing the frequency of dropout [14]. Flexible-dose protocols may limit the ability to assess the effects of specific doses, but giving

investigators the flexibility to increase or decrease the dose of a drug may in fact be more reflective of real-life applications. Side effects of higher doses of the test treatment might lead to unmasking of study participants.

- (3) *Select a target population not adequately served by current treatments.* Participants who are doing well at baseline on current treatments may be more likely to drop out when assigned a different drug because of lack of efficacy. Excluding such individuals may reduce treatment dropout.
- (4) *Consider 'add-on' designs where the study treatment is added to an existing effective treatment.* These designs may decrease the likelihood of missing data due to lack of efficacy. Such designs may be the only option for some conditions where use of placebo is not appropriate.
- (5) *Shorten the follow-up period.* Shorter follow-up periods may yield a reduction in dropouts, because fewer participants move out of the area, fewer develop intolerable adverse events, and the number and burden of clinical visits may be reduced. On the other hand, a short-term outcome may be less salient and may miss beneficial effects for participants who respond more slowly to study treatment. Past experience in similar trials can guide the evaluation of this tradeoff in specific situations.
- (6) *Allow a 'rescue' medication.* Allowing 'rescue' medications (alternative medications with established efficacy) for participants who do not respond to their assigned treatment could help to retain participants for the full duration of follow-up. If this design option is adopted, the estimand and associated outcome measurements need to be carefully defined in the protocol. For example, the designated 'rescue' therapy could be a component in the definition of a treatment regimen. A disadvantage of this approach is that the focus on a treatment regimen may detract from the objective of assessing the causal relationship between the new product and treatment outcome.
- (7) *Avoid outcomes that are likely to lead to substantial missing data.* Primary outcome measures such as those that require substantial invasive procedures (e.g., liver biopsies) are likely to result in significant missing data and should be avoided whenever possible. If the desired primary outcome measure requires invasive procedures, then one may consider use of a composite outcome, such as an outcome that incorporates death, or use of 'rescue' medication or surgery for initial poor response. However, composite outcomes can be difficult to interpret if individual components of the composite provide contrasting evidence about the intervention or if a weaker component dominates.
- (8) *Consider randomized withdrawal designs.* In such trials, participants who have responded to treatment are randomized either to continuation of treatment or withdrawal (e.g., placebo). Such trial designs may include a run-in phase where all participants are initially treated with the intervention under study. In cases in which loss of efficacy after withdrawal can be taken as evidence of drug efficacy, such a trial can generate long-term efficacy and safety data. This design is suitable when the goal is to assess the maintenance effect of the intervention.

2.3. The question of whether to continue data collection for treatment dropouts

Even with careful attention to limiting missing data in the trial design, some participants may not follow their assigned intervention all the way to the point where final outcome data are collected. An important question is then which data to collect for participants who discontinue the assigned intervention. Sponsors and investigators may believe that the participants are no longer relevant to the study and so are reluctant to incur the costs of continued data collection. Yet continued data collection, when supplemented with information about treatments after dropout, may be informative, particularly for the intent-to-treat estimand. Continued data collection also allows exploration of whether the assigned therapy affects the efficacy of subsequent therapies (e.g., by improving the degree of tolerance to the treatment through exposure to a similar treatment or by negatively impacting the efficacy of subsequent treatment as a result of drug resistance).

The correct decision on continued data collection depends on the selected estimand and study design. For example, if the primary estimand does not require the collection of the outcome after participants discontinue assigned treatment, such as with the estimand (4) (area under the outcome curve during tolerated treatment), then the benefits of collecting additional outcome data after the primary outcome is reached needs to be weighed against the costs and potential drawbacks of the collection.

An additional advantage of data collection after participants have switched to other treatments (or otherwise violated the protocol) is the ability to monitor side effects that occur after discontinuation of treatment and are not immediately apparent, although the cause of such side effects may be unclear if a participant switches to another treatment.

We are convinced, as argued in [14], that in the large majority of settings, the benefits of collecting outcomes after participants have discontinued treatment outweigh the costs. Specifically, the panel makes the following recommendations [6, pp. 30–31]:

Trial sponsors should continue to collect information on key outcomes on participants who discontinue their protocol-specified intervention in the course of the study, except in those cases for which a compelling cost-benefit analysis argues otherwise, and this information should be recorded and used in the analysis. The trial design team should consider whether participants who discontinue the protocol intervention should be provided access to and encouraged to use specific alternative treatments. Such treatments should be specified in the study protocol. Data collection and information about all relevant treatments and key covariates should be recorded for all initial study participants, whether or not participants received the intervention specified in the protocol. (Recommendations 3–5)

2.4. Power analysis in anticipation of missing data

An important and relatively neglected issue is how to account for the loss of power from missing data in statistical inferences such as hypothesis tests or confidence intervals. A common current approach is simply to inflate the initial sample size by the inverse of one minus the anticipated ‘analysis dropout’ rate, estimated from similar trials. Authors [15, 16] have developed methods for inflating sample size to account for noncompliance (treatment dropouts) to the assigned treatments. Inflation of the sample size is reasonable if the missing data are missing completely at random, but that assumption is generally too optimistic; therefore, power calculations should be based on more realistic missing at random (MAR) or missing not at random assumptions [17]. This is rarely performed, and how to do it well is an area for research; simulation studies are one possible approach.

The most worrisome effect of missing values on the inference for clinical trials is often not the reduction of power but the potential for biased estimation of the treatment effect, a problem which is not addressed by simply inflating the sample size. If the bias from missing data is similar in size to the anticipated size of the treatment effect, then detection of this effect is unlikely, regardless of the sample size. If a preliminary estimate of the potential bias can be developed, a simple strategy is to reduce the anticipated effect size by the anticipated size of the nonresponse bias and then power the study for this reduced effect size. If the adjusted effect size is too small to detect, there is a strong incentive to design the study to reduce the amount of missing data.

2.5. Application of design principles to the case studies

We now illustrate how the design recommendations might apply to the selection of estimands for the three case studies: (1) trials of chronic pain; (2) HIV treatment trials; and (3) trials of circulatory devices for heart failure.

2.5.1. Trials for chronic pain. Potential choices for the causal estimand include the following:

- (a) (*Difference in pain relief in all participants* [e.g., degree of pain relief at 12 weeks in all patients in whom the treatment intervention is initiated; estimand (1)]. This estimand may shed only limited light on whether the treatment therapy is effective if a high proportion of participants switch from the assigned treatment and receive effective alternative therapies. Thus, to the extent possible, treatment dropouts should be managed according to a policy described in the protocol that reflects current clinical practice. It is important to limit treatment dropouts, and this may be accomplished by a flexible dose design that allows dose reduction when treatment is not tolerated or dose increase when treatment has an inadequate response. The advantages of a flexible dose design need to be weighed against the benefits of fixed dosing regimens, such as determining a minimal effective dose. For this estimand, it is important to continue to collect data throughout the trial in all patients, including those who choose to switch therapies, and use these data in the analysis.
- (b) (*Difference in pain relief in tolerators* [e.g., degree of pain relief in patients who tolerate and choose to receive 12 weeks of therapy, that is, estimand (2)]. This option addresses a key regulatory question, long-term efficacy in patients who will take the drug, but it fails to address other key questions, especially how well and often the drug is tolerated and its efficacy in the total population, including those who do not take it for 12 weeks. One design that limits missing data is a *randomized withdrawal (discontinuation) design*, in which patients are treated with the test treatment open-label for some time, and participants who tolerate and have adequate response are

randomized to (a) continue on the drug or (b) switch to placebo. Participants are then followed-up for the remainder of the trial. Another design to limit missing data uses an *active control run-in period* followed by placebo washout, with randomization of those patients who tolerated the active control to active treatment or placebo. These designs may limit missing data problems but do not address safety and efficacy in the broader population (i.e., external validity is compromised).

- (c) (*Difference in treatment success rate* [e.g., proportion of patients who can tolerate therapy, remain in study, and achieve adequate pain relief over 12 weeks; estimand (5)]. This addresses an important regulatory question and avoids missing data by defining a composite primary outcome. However, classifying all patients as either a treatment success or not may ignore important information, such as the extent of success or cause of failure. Also, counting patients who cannot tolerate therapy as failures may strongly weigh against drugs that are excellent in patients who tolerate them, even if there are significant subsets of patients who cannot tolerate them.

2.5.2. *Trials for treatment of HIV.* Possible choices for a causal estimand include the following:

- (a) *Virologic response* (e.g., the percentage with an HIV RNA level of less than 50 copies/mL after 48 weeks) in all participants randomized and managed according to standard practice, whether or not study treatment is discontinued [intent-to-treat or estimand (1)]. This approach compares two treatment policies (e.g., starting with a regimen using drug A plus background treatment and starting with drug B plus the same background treatment). It is not often used in a regulatory setting because of concerns that estimation of the differences between drug A and drug B could be affected if more participants on one treatment group than another were switched to a virologically more potent regimen before 48 weeks.
- (b) *Virologic response in tolerators* [e.g., the percentage with an HIV RNA level of less than 50 copies/mL, among participants who were able to tolerate the treatment for 48 weeks (note that switches in ART due to lack of efficacy need to be differentiated from switches due to side effects or lack of tolerability)]. This approach addresses the efficacy in participants who will take the drug, a key regulatory question. However, it fails to address other key questions, for example, efficacy of the drug in the total population receiving it, including those who do not take it for 48 weeks. A run-in period is usually not practical because of concerns about HIV drug resistance. An analysis that excludes those who do not tolerate the study treatments may lead to biased estimates of treatment efficacy.
- (c) *Treatment success rate* (e.g., the proportion of all randomized participants who stay on assigned treatment, remain in the study, and achieve an HIV RNA level of less than 50 copies/mL at 48 weeks). This approach predominates in regulatory settings, addresses an important regulatory question, and avoids missing data through the use of a composite outcome, treating treatment dropouts as failures. However, the use of the composite outcome gives equal weight to missing data, deaths, intolerance, and lack of virologic efficacy. Thus, the approach may mask important treatment differences and hence yield misleading results. For example, outcomes labeled as virologic failures may in fact reflect toxicity or losses to follow-up. Furthermore, counting participants who cannot tolerate therapy as failures may overly weigh against drugs that have excellent virologic efficacy in patients who do tolerate them, even if there are significant subsets of patients who cannot tolerate them. The collection of data after treatment discontinuation to the end of follow-up may permit assessment of the consequences of treatment failure before 48 weeks due to intolerance or lack of virologic efficacy (e.g., the development of HIV drug resistance associated with virologic failure). Continued follow-up allows a separate assessment of each component of the composite outcome at or before 48 weeks (e.g., summaries of numbers assigned to each treatment who failed virologically). Treatment policies, as in estimand (1), can also be compared.

2.5.3. *Trials for mechanical circulatory devices for severe symptomatic heart failure.* We assume that, in such studies, four outcomes are of interest: death, disabling stroke, criteria met for implanting LVAD (e.g., based on previous trials in patients ineligible for transplant), and a self-administered quality-of-life assessment using a standard instrument (alternatively, or in addition to, a functional measure test, such as a 6-min walk time, could be used as an alternative or additional measure).

Three potential choices of estimand for evaluating health status are:

- (a) *Difference in quality of life between treatment groups for all randomized patients.* The quality-of-life comparison could be performed at a time point earlier than 2 years (e.g., at 6 months) to

maximize the number of patients in each treatment group who are followed up. Patients who die or who are unable to complete the questionnaire for health reasons could be given a ‘worst rank’ score so that they are included in the analysis. This latter strategy would likely affect the power of the resulting test statistic as some deaths would be expected to be unrelated to the treatments.

- (b) *Difference in quality of life among survivors.* This addresses the effect of the LVAD on a composite outcome of death, disabling stroke, or progression to prespecified criteria for LVAD implantation. Missing data may not be MAR, so an analysis method such as pattern-mixture modeling that does not assume MAR may be appropriate.
- (c) *Area under the quality-of-life curve while alive,* which has the advantage of taking into account both quality of life and duration of survival.

Irrespective of which estimand is used, it is important to assess health status as objectively as possible. Because the intervention in trials of this sort cannot be masked to patients or the investigators caring for the patient, the use of independent, trained, and masked evaluators (i.e., those not involved in the care of the patient) should be considered. In addition, both patient self-report and clinician evaluations of health outcomes should be considered. A working group of the Heart Failure Society of America has summarized considerations in measuring and assessing health status in such trials [18].

3. Trial conduct to limit missing data

The incidence of missing data varies greatly across clinical trials. Some of this variation is context specific, but we feel that careful attention to limiting missing data in trial conduct can substantially reduce the problem. An important step towards limiting missing data is to recognize the problem explicitly in the study protocol. Thus, in Recommendation 6 [6, p. 43], the panel states that

Study sponsors should explicitly anticipate potential problems of missing data. In particular, the trial protocol should contain a section that addresses missing data issues, including the anticipated amount of missing data, and steps taken in trial design and trial conduct to monitor and limit the impact of missing data.

The panel offered some practical suggestions on how clinical trial designers and managers can limit missing data in the conduct of the trial. These suggestions are quite practical, go a long way to prevent analysis problems, and deserve serious attention:

- (1) *Set realistic target and acceptable rates of missing data in the study protocol.* One way to set target rates and maximally acceptable rates for missing data would be to examine what happened in similar trials. Other trial experience, however, does not necessarily imply a different trial can do no better, and just because a previous trial had a certain amount of missing data does not mean that amount is acceptable. The amount of acceptable missing data will depend on many characteristics of the trial, including the medical condition under study, whether the assumption that the missing data are MAR is reasonable, the size of the anticipated effect of the intervention under study, and the likelihood that a sensitivity analysis would render the results of the trial inconclusive. Applied research is needed on this topic, and techniques will need to evolve on the basis of that research. Once targets for missing data are established, performance against these goals can be monitored, and the goals can be used to motivate investigators.
- (2) *Make the Data and Safety Monitoring Board (DSMB) for a trial aware of the trial’s target for missing data and have investigators report to and discuss with the DSMB progress relative to the target.* For trials with DSMBs or data monitoring committees, such committees can play an important role in monitoring missing outcome data during trial conduct. The DSMB should know about and help to set acceptable trial targets, but the primary responsibility for ensuring that missing data are kept to a minimum resides with the investigators, the protocol team, and the sponsor.
- (3) *Limit the participants’ burden and inconvenience of data collection.* Possible approaches to limiting participant burden are to minimize the number of visits and assessments, to collect only information directly relevant to the main study aims, to design user-friendly case report forms, to define outcomes that do not require attendance at a clinic visit whenever feasible, and to allow a relatively large time window for completion of each follow-up assessment.
- (4) *Provide effective treatments to participants after the trial,* such as continued access to effective study treatments or extension protocols until the treatment is licensed.

- (5) *Select investigators with a good track record of enrolling, following, and collecting complete data on participants in previous trials.*
- (6) *Provide training on negative impacts of missing data.* Training needs to emphasize the importance of complete data collection and the difference between discontinuing the study treatment and discontinuing data collection. Training should stress the value of collecting data after a participant discontinues the assigned treatment. As discussed previously, many trial sponsors and investigators mistakenly assume that there is little reason for additional data collection when participants discontinue study treatment. But as we emphasize, the continued collection of data is important in clinical trials.
- (7) *Train research staff and investigators on the informed consent process as a tool for encouraging complete data.* As part of the consent process, research staff need to convey to participants the importance of remaining in the trial regardless of the treatment to which they were assigned. Trainers need to emphasize that withdrawal of consent is a decision made by the participant, not the investigator or trial sponsor. When participants are dissatisfied with the conduct of the trial but have not yet withdrawn, the investigator should make an effort to address their concerns and retain them in the trial. In doing so, investigators must be careful that their efforts do not cross over into coercion.
- (8) *Provide financial incentives for completeness of data collection.* Payments to investigators should reflect follow-up work (e.g., payment per form, visit, or procedure completed) as well as the number of participants enrolled. In addition, it is acceptable and generally advisable to link a final payment to completion of the study closeout visit, providing the visit does not entail significant additional risks to the participant that create an investigator conflict of interest. If there are minimal risks associated with data collection to the participant, the investigator may receive financial incentives to continue to collect data, whether or not the participant continues treatment.
- (9) *Monitor missing data during the trial.* Missing data and missed visits should be assessed in real time by site personnel and summary reports made available to investigators at regular meetings and on study websites. This creates a climate that encourages investigators to avoid missing data. Identification of poorly performing sites can help identify the need for remediation, such as additional training, site visits, or even site closure.

Investigators and site personnel can also act in several ways to reduce the amount of missing data:

- (1) *Emphasize in the informed consent process the importance of participation for the full duration of the trial* Recommendation 7 of the panel states [6, p. 47] that

Informed consent documents should emphasize the importance of collecting outcome data from individuals who choose to discontinue treatment during the study, and they should encourage participants to provide this information whether or not they complete the anticipated course of study treatment.

Trial procedures should allow for informed withdrawal of consent. However, if participants discontinue study treatment, they should be made aware of the importance of continued follow-up for data collection. The following [6, p. 42] is an example of language for withdrawal that was used in the Development of Antiretroviral Therapy trial [19]

- I no longer wish to take trial anti-HIV drugs but I am willing to attend follow-up visits.
- I no longer wish to take trial anti-HIV drugs and do not wish to attend further visits. I agree to my medical records being consulted in future to obtain clinical information for the Development of AntiRetroviral Therapy (DART) in Africa.
- I no longer wish to take trial anti-HIV drugs and do not wish to attend further visits. I do not agree to my medical records being consulted in future to obtain clinical information for DART.

- (2) *Provide participants with financial incentives for remaining in the trial.* Paying for voluntary participation in a clinical trial is generally regarded as ethical provided that the responsible Institutional Review Board (IRB) ensures that the compensation is neither coercive nor at the level that would present undue influence (21 CFR 50.20) [20, 21]. Providing cash is generally not viewed as being coercive, as it is a benefit. Most IRBs allow a small proportion of the cash payment to be retained as an incentive for completion, but, generally, payments accrue as a study progresses in payment for participation activities that are completed. Payments for return visits of participants who have stopped taking medication are virtually always considered ethical, because the risk to the participant is probably minimal.

- (3) *Collect information on which participants are at risk for dropping out and why.* Formal ‘intent-to-attend’ questioning may help to identify reasons for dropout [22] and may yield useful covariates for missing data models. Potential factors influencing decisions to participate include time and duration of visits, the need for assistance with transportation or child care, need for reminders, problems in relations with the staff, problems with blood drawing or other procedures, side effects, and perceptions of intervention efficacy.
- (4) *Educate participants on the importance of continued engagement in the trial* to ensure that important scientific knowledge is generated. Mechanisms for such education include the production of a study newsletter, maintenance of a regularly updated website for trial participants, and providing access to interim papers and presentations on study progress and findings. IRBs generally require approval for communications with study participants.
- (5) *Provide non-monetary incentives for participant engagement and retention.* Examples include study-branded gifts; regular expressions of thanks; birthday and holiday cards; social networking; reminders before a visit and after missed visits; development of a friendly staff and welcoming environment; study practices that respect participants’ time and schedules; on-site diversions for small children; solicitation of input regarding relevant issues of study conduct; and valued education at the site.
- (6) *Provide transportation and child care costs.*
- (7) *Keep participants’ contact information up to date.*

4. Conclusion

Missing data undermine the benefits of randomization. These render the interpretation of the trial findings more difficult, because missing data analysis methods inevitably rely on assumptions that are not subject to empirical test. Thus, it is important to design and conduct clinical trials in a way that limits missing data and not assume that a statistical analysis, no matter how sophisticated, will fix the problem. This article has focused on ways to do this. A companion article focuses on principles of analysis of clinical trial data with missing values.

The distinction between treatment dropout and analysis dropout is important. There are trials in which substantial treatment dropout is understandable, but there is less reason for analysis dropouts. Even when estimands take into consideration an expectation that some trial participants will discontinue treatment, the benefits of retaining participants in the study can be substantial, including support for an analysis of effectiveness (comparison of treatment policies) and the ability to monitor side effects that occur after discontinuation of treatment.

In addition to trial design, aspects of trial conduct can also substantially reduce the amount of missing data. Section 3 highlights the importance of specifically addressing missing data in the trial protocol, monitoring the extent of missing data from the design stage throughout the conduct of a trial, and taking steps if the level is too high for some trial sites. This section also outlined specific practical techniques that should be considered to limit missing data in the conduct of the trial. Serious consideration of the ideas in the design and conduct of trials may substantially reduce the burden of handling missing data in the analysis and lead to more robust study findings.

References

1. O’Neill RT. Missing data in clinical trials intended to support efficacy and safety of medical products: the need for consensus. *30th Annual Conference of the International Society for Clinical Biostatistics*, Prague, Czech Republic, 2009. Available at: <http://www.iscb2009.info/RSystem/Soubory/Prez%20Tuesday/S18.3%20O’Neill.pdf>.
2. International Conference on Harmonisation. Efficacy Guideline E9. *Dose-Response Information to Support Drug Registration*, 1994; 1–36. Available at: <http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html>.
3. European Medicines Evaluation Agency. Guideline on Missing Data in Confirmatory Clinical Trials. *Committee for Medical Products for Human Use*, 2010; 1–12. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.
4. International Conference on Harmonisation. Efficacy Guideline E4. *Dose-Response Information to Support Drug Registration*, 1994; 1–10. Available at: <http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html>.
5. International Conference on Harmonisation. Efficacy Guidelines E10. *Choice of Control Group and Related Issues in Clinical Trials*, 2010; 1–29. Available at: <http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html>.
6. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The National Academies Press: Washington, DC, 2010.

7. Hogan JW, Cohen ML, Little RJ, Molenberghs G, Rotnitzky A, Scharfstein D. The analysis of clinical trials with missing data. To be submitted to *Statistics in Medicine*.
8. Meinert CL. Toward more definitive clinical trials. *Controlled Clinical Trials* 1980; **1**:249–261.
9. Dworkin RH, Turk DC, Peirce-Sandner S, Baron R, Bellamy N, Burke LB, Chappell A, Chartier K, Cleeland CS, Costello A, Cowan P, Dimitrova R, Ellenberg S, Farrar JT, French JA, Gilron I, Hertz S, Jadad AR, Jay GW, Kalliomäki J, Katz NP, Kerns RD, Manning DC, McDermott MP, McGrath P, Narayana A, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Reeve B, Rhodes T, Sampaio C, Simpson DM, Stauffer JW, Stucki G, Tobias J, White RE, Witter J. Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations. *Pain* 2010; **149**:177–193.
10. Kim Y. Missing data handling in chronic pain trials. *Journal of Biopharmaceutical Statistics* 2011; **21**:311–325.
11. U.S. Food and Drug Administration. Guidance for industry: antiretroviral drugs using plasma HIV; RNA measurements—clinical considerations for accelerated and traditional approval, 2002. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070968.pdf>.
12. Squires K, Lazzarin A, Gatell JM, Powderly WG, Pokrovskiy V, Delfraissy JF, Jemsek J, Rivero A, Rozenbaum W, Schrader S, Sension M, Vibhagool A, Thiry A, Giordano M. Comparison of once-daily atazanavir with efavirenz, each in combination with fixed-dose zidovudine and lamivudine, as initial therapy for patients infected with HIV. *Journal of Acquired Immune Deficiency Syndrome* 2004; **36**:1011–1019.
13. Miller LW, Pagani FD, Russell SD, John R, Boyle AJ, Aaronson KD, Conte JV, Naka Y, Mancini D, Delgado RM, MacGillivray TE, Farrar DJ, Frazier OH. Use of continuous-flow device in patients awaiting heart transplantations. *New England Journal of Medicine* 2007; **357**:885–896.
14. Lavori PW, Brown CH, Duan N, Gibbons RD, Greenhouse J. Missing data in longitudinal clinical trials part A: design and conceptual issues. *Psychiatric Annals* 2008; **38**:784–792.
15. Lakatos E. Sample sizes based on the logrank statistic in complex clinical trials. *Biometrics* 1988; **44**:229–241.
16. Shih JH. Sample size calculation for complex clinical trials with survival endpoints. *Controlled Clinical Trials* 1995; **16**(6):395–407.
17. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**:581–592.
18. Normand SL, Rector TS, Neaton JD, Pina IL, Lazar RM, Proestel SE, Fleischer DJ, Cohn JN, Spertus JA. Clinical and analytical considerations in the study of health status in device trials for heart failure. *Journal of Cardiac Failure* 2005; **11**:396–403.
19. DART Trial Team. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomized non-inferiority trial. *Lancet* 2010; **375**:123–131.
20. Emmanuel EJ. Undue inducement: nonsense on stilts? *American Journal of Bioethics* 2005; **5**(5):9–13.
21. Committee on Assessing the System for Protecting Human Research Participants. Responsible Research. *A Systems Approach to Protecting Research Participants*. Institute of Medicine. National Academy Press: Washington, DC, 2002.
22. Leon AC, Hakan D, Hedeken D. Bias reduction with an adjustment for participants' intent to dropout of a randomized controlled clinical trial. *Clinical Trials* 2007; **4**:540–547.