Review: Clinical Trial Methodology

CLINICAL INVESTIGATION

The handling of missing data in clinical trials

Clin. Invest. (2013) 3(3), 241–250

The problem of handling missing data in clinical trials is discussed, particularly in the light of two recent publications associated with the US and European regulators. The importance of assessing methods in the light of the target of the analysis – the estimand – is emphasized, and two types of estimand, *de jure* and *de facto*, are introduced. A key distinction is made between analyses in which missing data are 'defined away' and in which they represent a nuisance to be accommodated. The role of sensitivity analysis, which is acknowledged to play a major role in this setting, is considered. Finally, the recent recognition of the importance of collecting outcome data following protocol deviations is touched upon.

Keywords: *de facto* • *de jure* • estimand • intention-to-treat • missing-at-random • pattern mixture • per-protocol • protocol deviation • selection model • sensitivity analysis

Michael G Kenward

Department of Medical Statistics, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK
Tel.: +44 207 927 2472
Fax: +44 207 637 2853
E-mail: mike.kenward@lshtm.ac.uk

There is a long history of debate, not always constructive, among those involved in clinical trials about the appropriate way to deal with the ubiquitous problem of missing data. The debate has a distinctive tenor that stems from the fact that this is a problem that cannot have a definitive solution. When data are missing, any analysis that purports to produce nontrivial conclusions must rest on assumptions that cannot be wholly assessed from the data at hand. It is this author's view that the discussion around the subject has matured greatly in recent years; in particular, there have been two significant publications associated with the regulators in the USA and in Europe. In 2010, the US National Research Council (NRC) produced a report on the handling of missing data at the behest of the US FDA [1]. This was jointly written by a panel of leading international experts in the area, chaired by Rod Little. The coverage of the report is extensive, and provides a set of 18 key points, most of which will be echoed in the following. Several publications have appeared, and will be appearing in the future, that review, explore and develop the main themes in the report [2–5]. At around the same time, the European regulators produced their own set of guidelines [6]. Although comments were invited on a draft of this document, in contrast to the NRC report, this was written by the regulators themselves. While very different in style and content, the two documents do share some important themes to which we return below, in particular the need for appropriate sensitivity analysis and the importance of the collection of data following withdrawal. The NRC report reflects to a great extent, and the European Medicines Agency guidelines to a somewhat lesser degree, the change in the nature of the debate surrounding missing data that has been seen over the last few years. It has become clear that a key element of the rational consideration of alternative methods of analysis must depend on an unambiguous statement of the aims of the analysis. So increasing effort has gone into clarifying the analysis aims, and this is reflected in this review. The opening sections focus on matters of definition of the main concepts and terms to be used and on making distinctions among these. Only when these are in place are we in a position to make a constructive comparison of potential approaches. To summarize, what we see in this review is the framing of

FUTURE SCIENCE part of fsg

the problem of handling missing data, which has evolved greatly over the last few years, alongside that of specific statistical technicalities. This will be reflected in what follows, where much of the discussion will be concerned with formulating appropriate definitions and questions rather than particular details of statistical methodology. References for the latter reach back over 25 years, and remain highly relevant.

To begin with, we need to be clear by what is meant by a 'missing value'. The definition to be used now is close, but not identical, to that used in the NRC report. A missing value will be taken to mean an observation the value of which is required for the target of a particular statistical analysis, and it is noted here that this does not coincide with the more usual definition of a missing value as an observation that was not collected but had the potential to be. This rather loose definition will be tightened below when the concept on an estimand is introduced. An immediate implication of this is that we cannot even be clear about what is missing until we are clear about the aims of the analysis. Note that this will, of course, depend greatly on the aims of the trial, but the same trial can support different aims and hence different analyses, with the implication that what is missing for one analysis may not be missing for another. Given this, the focus here will be on missingness caused by dropout (variously described as withdrawal, loss to follow up, or attrition). This concept will be clarified further below. Typically this is the main source of missing data in the longitudinal trial setting, particularly given that the primary end point for analysis is nearly always a comparison at some given time at, or near, the end of the trial period. It is assumed that all baseline measurements and covariates are fully observed. We also note that death will only rarely be a source of missing data in the sense meant here. The issues raised by deaths that cannot definitively be separated from the outcome under investigation are rather different to those typically handled under the heading of missing data; see for example [7]. An exception to this would be an occurrence such as a purely accidental death, which would anyway be expected to be rare.

This brings us to a further and vital distinction. It is very common in practice for the trial outcome itself to be defined by the occurrence of dropout/withdrawal. This may be done directly, such as in the situation with a binary success/failure outcome, in which withdrawal is defined as a treatment failure and assigned a score of zero. There are many variations on this theme, with potentially quite complex rules governing these definitions. The outcome may be also be defined jointly in terms of withdrawal and the response variable; an example of this is the so-called last observation analyzed (LOA) in which the outcome for a subject is the most recent value obtained before dropout/withdrawal. It is noted in passing that an LOA-based analysis is arithmetically identical to that based on the so-called Last Observation Carried Forward (LOCF); it is the interpretation that differs. In such examples the missing data have been defined away; that is, there are no missing data. The goal of the analysis is expressed in terms of these defined outcomes and so the problem has moved away from one of handling missing data to one of the appropriate clinical interpretation of the given outcomes. For a discussion of these and related issues see [8]. As we are concerned here with the problem of handling missing data, minimal consideration is given to methods in which these are eliminated through definition, and we focus rather on those settings in which we can regard missing data as a nuisance to be accommodated in the sense that the ideal analysis would be based on a complete set of data collected under the required conditions.

The missing value literature is huge, and it is impossible to accommodate more than a fraction within this article. Much that is, at the moment, largely of technical interest only within the statistical community has been omitted, with a few pointers given to potentially important areas. Of the remaining work, this article reflects the author's views on first, the important recent developments in the two key regulatory publications described above and, second, on areas of development in which the author has been involved to a greater or lesser extent. For statistical methodology that is now well established, reference has been made to standard texts and many further references can be found in these if the reader wishes to follow up details, or pursue particular avenues that are only touched on here.

The estimand & protocol deviations

We have seen above that the definition of missing value depends on the aims of the statistical analysis. The NRC report encapsulates this in the so-called estimand. This is defined as a property of the population that the trial is known or assumed to target, and refers to a quantity that reflects the differential effect of given treatment regimes. The estimand is assumed to reflect in a meaningful way the clinical goals of the trial and analysis. It can be viewed in terms of something to be estimated, or something on which a statistical test is to be conducted. Most importantly, the definition of the estimand remains the same whether or not there are missing data. When an analysis is constructed, the estimand is the target, and the success or otherwise of the analysis is judged in terms of its ability to make valid inferences about this target and, crucially, about the assumptions that are needed for such validity to hold. To clarify this, we distinguish two types of estimand, following

closely the development and terminology in [9]. We call these *de jure* and *de facto* estimands. A *de jure* estimand is one that compares the effects of treatments that are taken strictly according to the protocol, while a *de facto* estimand compares the effects of the treatments actually taken, irrespective of randomized treatment. Note that the former may well be counter-factual such as when some subjects cannot tolerate a treatment. These two estimands are clearly connected to the commonly used terms 'pre-protocol' (PP) and 'intention to treat' (ITT), but it is important not to confound them. *De jure* and *de facto* are definitions of estimands, in themselves they are not methods of estimation or analysis, while PP and ITT typically refer to groups of patients (possibly all) in a trial, and partly define what analysis is to be used. It is this disjunction between the meaning of commonly used terms like ITT, and the meaning of an estimand, together with the ambiguity in the use of ITT and PP (see for example [10]), that prompted Carpenter *et al.* to introduce these new terms [9].

We can link this definition of a *de facto* estimand to the original National Academy of Sciences report where we have the following statement [1]:

"*…confirmatory clinical trials should estimate the effect of the experimental intervention in the population of patients with greatest external validity and not the effect in the unrealistic scenario where all patients receive treatment with full compliance to the treatment schedule and with a complete follow up as per protocol.*"

If all intended observations in a trial were made, then it would be possible to estimate a *de facto* estimand in an unambiguous way; all the data needed to estimate the estimand would then be available. The same is not true for a *de jure* estimand. Suppose, however, that some subjects depart from their randomized treatment. Even if all the data were collected, we would not have a similarly unambiguous estimate. For this we would need data that we do not have. In this situation, all data following a protocol deviation would be missing as far as the *de jure* estimand is concerned, even though we do have observations for all patients at all visits. More generally we use the term protocol deviation, or just deviation, for any departure from the protocol. Examples are loss to follow up with no further information obtained, withdrawal from the study treatment, and so on.

The key points so far can be summarized as follows:

- We cannot separate the handling of missing data from the particular estimand under consideration;
- Observations are defined as missing if they are not obtained under the conditions required for the given estimand;
- Analyses with missing data incorporate implicitly or explicitly the assumed statistical behavior of the missing data, in the light of what has been observed. One consequence of this is that any analysis that addresses a particular estimand when data are missing must incorporate information about the treatment taken (assumed or known) following deviations.

Broadly, the problem of handling missing data cannot be separated from the assumptions about the treatments taken. Any analysis that purports to target an estimand without making such assumptions explicit must be doing so implicitly. The paper by Little and Yau is an early recognition of these points [11], whilst more recently, there has been recognition of this and developments based on it [9,12–15].

Missing data mechanisms

Much writing on the handling of missing data has taken as its framework the classification of missing data mechanisms introduced by Rubin (1976) [16]. Here we introduce this framework with sufficient detail for its use in this review, but avoiding an exhaustive development. The aim is to show how the previous discussion of *de jure* and *de facto* estimands fits into this framework, and to allow reference to the mechanisms below when these are needed. We express these classes in terms of dropout, but could equally well refer to a protocol deviation that generates missing data in the sense used above. Data are said to be missing completely at random (MCAR) if the probability of a subject dropping out in an interval between two measurement occasions does not depend on any observations from that subject either already collected, or that might be collected in the future. In the dropout, or monotone setting, data are said to be 'missing at random' (MAR) if, having allowed for any observed observations in the past, including baseline covariates and treatment taken (together called the subject's history at this time), the probability of dropout does not depend on future observations. An alternative, and arguably more accessible, way of saying the same thing is that, given a subject's history, the statistical behavior of future observations is the same whether the subject drops out or not. Formally, by 'statistical behavior' we mean the conditional distribution of the observations given the history. Finally, missing not at random (MNAR) corresponds to all cases that are not MCAR or MAR: given a subject's history, the statistical behavior of future observations would not be the same if the subject continued compared with the behavior that would occur if the subject dropped out. Most importantly, in nearly all cases, the data under analysis cannot distinguish between MAR and MNAR, because this depends on the behavior of the unseen data. One exception to this occurs when subjects are withdrawn because their observations have drifted beyond prescribed limits. This is a special case of MAR,

sometimes called deterministic MAR [17].

MCAR implies random dropout in the intuitive sense: the subjects remaining at the end of the trial, sometimes called the completers, are a genuine random sample of those who started, and any analysis valid for the full data set would be valid for those who remain. There is of course loss of efficiency due to the consequent reduction in sample size. Unfortunately, MCAR is usually a very implausible assumption for at least some of the subjects in a trial. Formally, the MCAR assumption can be assessed (e.g., rejected) using the observed data [12,13].

MAR occupies a very special position in the missing data setting, not because it is especially plausible, but because it is in a certain sense a minimal assumption under which analyses can be constructed in which the actual missing data mechanism can be ignored, or in which the mechanism can be estimated from the observed data and so incorporated explicitly, leading in both cases to valid conclusions, given of course that the models used in the analysis are correct. One of the most important ways in which MAR is exploited in a statistical analysis follows from the definition stated earlier: 'given a subject's history, the statistical behavior of future observations is the same whether the subject drops out or not'. It follows that future statistical behavior for those who dropout (or deviate) can be estimated (or 'borrowed') from those who do not. However, if statistical behavior is being borrowed in this way then by implication the treatment taken is also being 'borrowed' from those who remain and applied to those who dropout (by implication the randomization group is the same) and, excepting rather unusual circumstances, this implies a *de jure* estimand. As mentioned earlier, this may imply a counterfactual situation: if a subject is unable to tolerate the randomized treatment, then the *de jure* estimand is still defined in terms of that subject receiving that treatment throughout the trial. It follows at once, by contrast, that the scenarios captured by a *de facto* estimand will, in most cases, correspond to MNAR mechanisms. This follows directly from the above definition. If those who deviate change to a treatment the effects of which differ, their future behavior cannot match those who remain on the randomized treatment. This has a very interesting implication. A *de facto* analysis with completely observed data can ignore the actual treatment taken, but as soon as data are missing we need to take account of this.

Simple *ad hoc* methods of analysis

It is probably true, although hard to measure in practice, that the great majority of trial analyses continue to use very simple *ad hoc* methods to accommodate the problem of missing data. The restriction to completers is one such *ad hoc* analysis, and we have seen above that this will only be strictly valid under the implausible MCAR assumption and will, even when valid, be potentially inefficient compared with methods that use the incomplete sequences of data from those patients that drop out. Another very commonly used *ad hoc* technique is LOCF, mentioned above, and its close relative 'last baseline carried forward'; for example, see [12,13]. Before commenting on these methods we need to be clear that they are not being used in the 'definitional' sense introduced earlier; for then it is strictly not a missing data problem. In fact, it is likely that, along with other simple imputation methods, this is by far the most common intention when such methods are used, and it would then be helpful if terminology like LOA could be used instead. This still does not absolve such methods of criticism; however, there remain major concerns about the interpretation of the results of such analyses. A particularly damning criticism of the use of LOCF in the neurosciences is given in [18].

When we consider LOCF, and related simple imputation analyses; that is, methods in which each missing value is replaced by a single number, as methods for dealing with missing data as a nuisance, in the sense intended in this review, we run into a host of problems. The NCR report is particularly critical of its use [1], and the European Medicines Agency guidelines do express suitable caution [6]. The problem with assessing such methods is that they are defined not in terms of the aims of the analysis, or estimand, but procedurally; that is, in terms of what is 'done to the data'. There is a sense in which such methods need to be reverse-engineered to determine what estimand they were intended to target in the first place. This can muddy discussions of their use in practice. Focusing on LOCF, different authors can have different aims for their particular LOCF analyses, which may be explicit or implicit, and it is no surprise that there are many critical assessments; examples are [19–28]. One overall message is that LOCF analyses can be shown both to create apparent treatment effects when none exists and hide them when they do. One defense that has been put forward is that there are circumstances in which LOCF is known to be conservative [6]. Such claims do need to be made with care, however. First, we need to define the estimand to which we refer (this is usually not done). Second, we need to define 'conservative'. This could mean conservative in estimate or in inference or both. Third, such a claim can only be made in the light of strong assumptions about the nature of the actual treatment profiles both before and after dropout and the actual dropout mechanism, most of which is unknown. In nearly all such cases the strength of assumptions required to make such a claim would be more than enough to construct a more statistically principled analysis. A recent publication establishes the

conditions under which an LOCF analysis does indeed provide a valid estimate and subsequent inference [29], see also the discussion in [30]. Unsurprisingly, the conditions for this correspond to a MNAR mechanism; we should expect this, since subjects who drop out have explicitly different behavior from those who remain, even if they share the same history. It turns out that this LOCF MNAR mechanism has a particular property, known as future dependence [31]. This means that the probability of dropout, given the history and the value of the (possibly unobserved) following observation, still depends on observations further into the future. Such mechanisms can exist, although it would be unusual to deliberately formulate a process for practical use with such counterintuitive properties. The key point here is not whether such a mechanism is possible or not in practice, rather that such a complex and counterintuitive missing value mechanism has emerged from such a simple procedure. This reflects a general point concerning missing data and simplicity. Apparently simple analyses such as LOCF can imply underlying assumptions that are neither simple nor transparent and, vice versa, simple and transparent assumptions can imply analyses that appear complex.

Model-based analyses & multiple imputation

If we are to avoid simple *ad hoc* methods when these are not appropriate, we need to consider more statistically principled approaches, and typically these will be based around some form of statistical model. These have the advantage of making the assumptions more explicit. There are many alternatives available and there is a large literature surrounding these. As yet, however, only a small fraction of these are used in a routine way in a clinical trial setting. Primary analyses that use explicit statistical models are nearly always based on the MAR assumption and because, as we have seen above, a MAR-based analysis implicitly borrows treatment behavior from those who stay for modeling the statistical behavior of those who drop out, this will usually imply that a *de jure* estimand is the target. Typical analyses for continuous outcomes are based around unconstrained forms of the multivariate linear model (sometimes called, in this context, the Mixed Model Repeated Measures analysis); see for example [12,13]. Analogous analyses for other outcomes, particularly binary, introduce additional complications associated with the lack of a suitable flexible multivariate distribution and the use of a nonlinear scale for the treatment effect. For a discussion of suitable approaches see [12,13]. None of these make great demands, however, from a technical perspective. An alternative, important class of analyses known as 'Inverse Probability Weighted' have received much attention (see for example [32,33]). These play a leading role in causal inference for observational data, but their use is as yet limited in the routine trial setting. This may be in part due to the technical demands of much of the literature and the lack of standard software implementations, but the situation could well change in the future. These methods do have the advantage of relaxing the strong modeling requirements of most of the alternatives that typically rely on likelihood-based methodology, but disagreement remains about their comparative advantages, especially in finite samples.

There has been increasing interest in the use of so-called multiple imputation (MI) for handling missing data in trials. MI was originally introduced by Rubin for large sample surveys with missing data [34], and it is particularly well suited to such settings. Its use has subsequently spread to many other areas. One of its great strengths is that it can be used to formulate analyses when data are missing using analysis tools that would have been applied had the data been complete. Two models are defined for MI. First, the substantive model is the model that would have been used to make inferences about the estimand had the required data been fully observed. This provides a direct and helpful connection to the definition of an estimand and the aims of the analysis. For a continuous outcome this might be, for example, a baseline-adjusted analysis of covariance of the observations made at the final time. Second, the imputation model is a model for the distribution of the missing data conditional on the observed data, or in our longitudinal setting, the conditional distribution of future observations given the history. In the MI analysis, several (possibly many) completed data sets are constructed using appropriate Bayesian draws from the imputation model and the substantive model is fitted separately to each of the completed sets. Rubin provided simple rules whereby the results from these analyses are combined to provide an overall estimate and measure of precision, together with inferential tools, that have broad validity and properly acknowledge, in a statistical sense, the incompleteness of the original data. There are now several implementations of MI in mainstream statistical packages [35,36].

Although MI is a flexible and very widely used approach for handling missing data, we need to be clear about its potential role in the current trial setting in which we are assuming that only outcome data are missing. It turns out that when this is true, the imputation and substantive models coincide for subjects who drop out or deviate. Consequently, we obtain essentially the same analysis whether we use conventional MI or use a model-based likelihood analysis, perhaps with suitable small sample correction. If MI is to have a role in this setting then, it must be in other situations in which these two models do not coincide – in this setting this

is known as uncongeniality [37]. Three such examples are described in [38]:

- It may happen that measurements collected after randomization are predictive of dropout/deviation. We cannot condition on these in the subsequent analysis because, as is well known, such conditioning potentially leads to bias. This implies that we cannot include them in the substantive model. However, such measurements can be incorporated in the imputation model if required. An example is a measure of noncompliance. Suppose that those with greater observed noncompliance are more likely to drop out. If we have a *de facto* estimand then it is important that the analysis reflects this behavior. A MAR-based analysis, such as one in which the substantive and imputation model both ignore the noncompliance, would treat dropouts as though compliance of the dropouts matched those remaining in the study, and so would potentially exaggerate the treatment effect; that is, would bias it from a *de facto* perspective. The noncompliance would have to be included in the imputation model to avoid this.
- For the reasons given earlier, there are technical issues surrounding the longitudinal modeling of non-normal outcomes, such as binary, under MAR. One commonly used measure of treatment effect, called marginal or population averaged, is awkward to estimate using a likelihood approach, and nonlikelihood methods, such as generalized estimating equations are typically used. In their simple form, such analyses are valid only under MNAR, not MAR. Two principal corrections can be used for validity under MAR. The first is inverse probability weighted and the second is MI. The problem with direct application of MI in this setting for a marginal model is that the awkward nature of a likelihood analysis carries over to the imputation model. However, as a referee points out a nonlikelihood method based on the so-called 'Fully Conditional Specification' might be considered as an alternative to this, a convenient, but uncongenial, imputation model can be used, provided that it is sufficiently rich. Examples are log–linear and sequential logistic regression models. Details are given in [12].
- We have seen that in the presence of patient deviations, *de facto* estimands require us to use missing data mechanisms that depart from MAR. These can be encapsulated in appropriate imputation scenarios and so form the basis of sensitivity analyses. We return to this important use of MI below.

Sensitivity Analysis

With a few exceptions, most of the commonly used model-based analyses for longitudinal trial data with dropout rest on the MAR assumption and, in general, any such analysis will require untestable assumptions. Moreover, as we have seen above, such analyses will typically correspond to a *de jure* estimand, which may not be the relevant one for the problem. As a consequence, the need for sensitivity analysis has long been recognized in the missing data literature. Indeed, much of the text by Daniels and Hogan [39] and Molenberghs and Kenward [12] is devoted to this. Similarly, both the NRC report [1] and the European Medicines Agency guidelines [6] stress the need for sensitivity analysis. While there is much agreement on the need, there is, however, less agreement about how this should be done. Many alternative routes are possible. It is clear, however, that we need to move away from *ad hoc* sensitivity analyses that consist of rather arbitrary collections of alternative analyses, which lack a clear overarching rationale. A proper sensitivity analysis is not a collection of arbitrary and *ad hoc* analyses, rather we need a structure along the following lines:

- A clear definition of the estimand of interest;
- The assumptions under which the primary analysis is valid for this estimand (noting, as emphasized earlier, that the choice of estimand reflects the aims of the analysis in the given trial setting);
- A nomenclature for practically relevant and accessible departures from these assumptions;
- Valid methods for assessing sensitivity to these assumptions.

There is no need to be too prescriptive or proscriptive about which technical approaches should be considered for this; clearly the details must depend critically on the substantive setting. In one way or another, many of the proposed methods assess sensitivity to departures from MAR. From a statistical perspective two broad routes can be distinguished.

In the first, the sensitivity analysis is framed in terms of the missing value mechanism itself. The assumptions are expressed in the ways in which dropout might be influenced by outcomes, past and future, and other information that may be available. This is sometimes called a selection model approach, and is particularly convenient for expressing certain forms of departure from the MAR assumption; for example, see [12,40–42]. The term 'selection' is taken from the econometric literature [43] and refers to the explicit modeling through a probability model that describes the way patients are 'selected' for dropout. One potential drawback of such methods is the need to provide an accessible description of the precise assumption whose sensitivity is being assessed, and this is often expressed in terms of rather indirect quantities such as log-odds ratios in logistic regression models for the probability of dropout/deviation.

The second broad approach attempts to provide a more directly accessible representation of sensitivity. Essentially, it is based on the alternative expression for MAR used earlier: under MAR, the statistical behavior of future observations is the same whether the subject drops out or not. Departures from MAR can then be formulated in terms of alternative future behaviors, which can be described in simple terms and even plotted in various ways to aid communication. This is known as the pattern-mixture approach [12,31,39]. In this, a potentially different statistical model is used for those patients who drop out/deviate at different times and for those who complete, and so allows explicit modeling of the behavior of dropouts under different *de facto* scenarios, and as shown in [44], such a pattern mixture framework can be used to incorporate expert opinion into sensitivity analyses.

Although not essential, a particularly convenient method for using the pattern mixture approach for sensitivity analysis is MI. For this, the alternative scenarios for future behaviors correspond to different imputation models, while the substantive model remains the same as that used in the primary analysis. The necessary steps are set out in [31] and [45]. The key question in the use of this approach then is the choice of alternative scenarios. An important advance was the recognition that these might be 'borrowed' from different treatment groups, depending on the known, or assumed, post-deviation treatment. Little and Yao applied this to the setting in which it is assumed that subjects deviating from active treatment move on to placebo, so that the alternative future behavior for the active dropouts is 'borrowed' from the placebo group [11]. This is an example of an MNAR setup, corresponding to a particular *de facto* scenario. As such it can be considered to be a primary analysis in its own right, or as part of a sensitivity analysis in which different post-deviation treatment behaviors are considered for a *de facto* estimand. It has one great technical advantage: it requires only that the MAR model be fit to the data. The subsequent imputations made under the MNAR *de facto* scenarios can be constructed in relatively simple steps from this MAR model. Within this general MI/pattern-mixture framework a large range of possible scenarios can be considered. Two, including a principled form of LOCF, Last Mean Carried Forward, are described in [14], while a fairly comprehensive treatment of the overall approach is given in [9], and a SAS macro, written by James Roger, which accommodates a range of alternative scenarios, is freely available from [101]. Within this overall framework alternative scenarios can be subject dependent, so this provides a natural and convenient way of incorporating the reasons for dropout/deviation into the analysis. Depending on the reason for dropout, different subjects may be associated with different scenarios.

These scenarios constructed in the manner described in the previous paragraph correspond to a *de facto* estimand. We have seen that if data are missing we cannot construct a *de facto* analysis without specifying in some way the future statistical behavior of subjects who deviate from the protocol. The scenarios defined in the pattern mixture framework correspond to different assumed behaviors and, as such, can form a component of a well-structured sensitivity analysis. This means that the analyst is forced to make decisions about the future behavior of those who deviate, which ensures transparency. The apparent disadvantage of this is that it may both be difficult to decide what treatment use is appropriate, and that there may be little or no information available on the statistical behavior under the selected treatment. This, however, reflects the difficulty of the problem, not a drawback of the analysis. Any analysis for a *de facto* estimand in which data are genuinely missing, that is for which missing data are not defined away, will necessarily define explicitly or implicitly such statistical behavior. The pattern-mixture approach simply makes this explicit. It also raises useful questions about scenarios that are meaningful for a *de facto* estimand, and in some settings raises questions about the value of a blanket use of *de facto* (or ITT type) estimands [46]. However, this takes us beyond the purely missing data problem.

Collection of post-deviation data

It is clear from the above that no amount of statistical expertise can make up for the absence of real data. We have also seen that for *de facto* estimands we need observations under whatever treatment (within reason) a randomized subject actually takes. This points to the potential value of the collection of post-deviation data, a point made at some length in the NRC report (indeed in its title) [1], and emphasized as well in the European Medicines Agency guidelines [6]. There are clearly cost and logistical issues to be addressed in this, and it may be that a successful compromise can be reached in which a reduced collection strategy is used that ensures that the most relevant information is obtained. A recent paper, considers strategies that can be used for this from a statistician's perspective [47]. A discussion of this is also provided in [5], and further details are given in [48].

Future perspective

Three main themes can be expected to play an important role in practically relevant developments in the statistical handling of missing data over the next few years. First,

the implications of the necessary linkage between the estimand and treatment taken will be explored and developed, and is likely to underpin significant new developments. Second, there needs to be a thorough exploration and assessment, from a practical perspective, of alternative approaches to sensitivity analysis, again reflecting the linkage between estimand and treatment actually taken. Third, issues surrounding the collection of post-deviation data will need to be explored from both practical and statistical perspectives, but this practice is expected to increase. It would also be beneficial if, as understanding becomes more widespread of the properties of simple imputation methods like LOCF, the use of these declines. Finally, in the interests of clarity, it is to be hoped that the distinction between defining away missing data through outcome definition, and the handling of genuine missing data, can be maintained in discussions on the subject.

Acknowledgements

Much of the author's thinking on missing data has been shaped by discussion with J Carpenter, G Molenberghs and J Roger.

Disclaimer

Opinions expressed in this article are those of the author alone, and not of any medical society or professional association.

Financial & competing interests disclosure

M Kenward is GlaxoSmithKline Professor of Biostatistics at the London School of Hygiene and Tropical Medicine. He has consulted for GlaxoSmithKline, Novartis and Pfizer in the last 12 months. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

Executive summary

- All nontrivial analyses for missing data rest on assumptions that cannot be assessed from the data under analysis.
- A comparison of different approaches for handling missing data cannot be separated from the aims of the analysis.
- The National Research Council report encapsulates such aims in the definition of the estimand.
- It is vital to distinguish between approaches in which missing data are 'defined away' by incorporating them into the definition of the subject outcome; for example, by defining dropout as a treatment failure, and in which they are a nuisance to be accommodated: such approaches have different estimands and so it is not meaningful to compare them directly.
- Two classes of estimands are the so-called *de jure*, in which treatments are compared assuming subjects follow their randomized treatment according to the protocol, and *de facto*, which compares the effects of the treatments actually taken irrespective of randomized treatment.
- When data are missing, any analysis that targets a *de facto* estimand must make assumptions about treatment use following dropout/deviation. This may be done explicitly or implicitly.
- Given the necessary uncertainty about the validity of assumptions underlying missing data analyses, including potentially treatment use following dropout/deviation, appropriate sensitivity analyses are strongly recommended.
- Such sensitivity analyses should be constructed in a coherent and relevant way given the estimand.
- An arbitrary and *ad hoc* collection of alternative analyses with no such underlying rationale does constitute a meaningful sensitivity analysis.
- There are many potential approaches to constructing sensitivity analysis and no broad agreement as yet about the most appropriate routes to take.
- Sensitivity analyses combining pattern mixture models with multiple imputation provide a transparent linkage between the estimand and the assumptions being targeted.
- The collection of post-deviation outcomes can greatly reduce the reliance on assumptions in *de facto* targeted analyses and so can, in principle, greatly reduce the problems surrounding missing data. This is likely to be of increasing importance in the future.

References

Papers of special note have been highlighted as:

- ▪ of interest
- ▪▪ of considerable interest

1 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, USA (2010).

▪▪ **The comprehensive report from the US National Research Council (NRC). It should be regarded as essential reading for anyone considering the problem of handling missing data in clinical trials. Its breadth reflects the wide range of views and expertise of the panel of authors.**

2 O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clin. Pharmacol. Ther.* 91, 550–554 (2012).

3 Ware JH, Harrington D, Hunter DJ, D'Agostino RB. Missing data. *N. Engl J. Med.* 367, 1353–1354 (2012).

4 Little RJA, D'Agostino RB, Cohen ML *et al.* The prevention and treatment of missing data in clinical trials. *N. Engl J. Med.* 367, 1355–1360 (2012).

▪ **Essentially a summary of the main points from the NRC report, valuable for its focus on the key points.**

5 Little RJA, Cohen ML, Dickersin K *et al.* The design and conduct of clinical trials to limit missing data. *Stat. Med.* 31(28), 3433–3443 (2012).

6 Committee for Medicinal Products for Human Use. *Guideline on missing data in confirmatory clinical trials*. European Medicines Agency, London, UK (2010).

■■ **Set of guidelines from the European regulators. Although it is far less comprehensive than the NRC report, not fully consistent in the views put forward, and less comfortable with the more technical areas of the subject, it marks a great improvement over previous guidelines. It is important as an aid to the understanding of the current position of the regulators in Europe.**

7 Kurland B, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* 6, 241–258 (2005).

8 Mallinckrodt CH, Kenward MG. Conceptual considerations regarding choice of endpoints, hypotheses, and analyses in longitudinal clinical trials. *Drug Inf. J.* 43, 449–458 (2009).

9 Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviations: a framework for relevant, accessible assumptions and inference via multiple imputation. *J. Biopharm. Stat.* (2013) (In Press).

■ **Introduces the concepts of *de jure* and *de facto* estimands as a foundation for formulating primary and sensitivity analyses. A general framework, with accompanying software implementation, for the construction of such analyses is developed using pattern-mixture models with multiple imputation.**

10 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomized controlled trials. *Br. Med. J.* 319, 670–674 (1999).

11 Little RJA, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics* 52, 471–483 (1996).

12 Molenberghs G, Kenward MG. Missing data in clinical studies. *Biometrics* 65(3), 993–994, (2009).

13 Carpenter JR, Kenward MG. *Missing data in clinical trials - a practical guide.* Birmingham: National Health Service Co-ordinating Centre for Research Methodology (2008).

14 Kenward MG, Carpenter JR. Chapter 21: Multiple Imputation. In: *Longitudinal data analysis*. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (Eds). Taylor & Francis, Oxford, UK (2008).

15 White IR, Horton NJ, Carpenter JR, Pocock SJ. Strategy for intention to treat analysis in randomized trials with missing outcome data. *BMJ.* 342, d40 (2011).

16 Rubin DB. Inference and missing data. *Biometrika.* 63, 581–592 (1976).

17 Murray GD and Findlay JG. Correcting for the bias caused by drop-outs in hypertension trials. *Stat. Med*. 7, 941–946 (1988).

18 Molnar FJ, Hutton B, Fergusson G. Does analysis using 'last observation carried forward' introduce bias in dementia research? *Can. Med. Assoc. J.* 179, 751–753 (2008).

■ **Provides a well-argued case against the use of 'last observation carried forward' in the neurosciences. The criticisms apply much more widely, and are indicative of the problems that can arise with the use of analysis techniques that are not well founded.**

19 Heyting A, Tolboom JTBM, Essers JGA. Statistical handling of drop-outs in longitudinal clinical trials. *Stat. Med*. 11, 2043–2061 (1992).

20 Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Stat. Med*. 14, 1913–1925 (1995).

21 Siddiqui O, Ali MW. A comparison of the random-effects pattern-mixture model with last-observation-carried-forward (LOCF) analysis in longitudinal clinical trials with dropouts. *J. Biopharm. Stat.* 8, 545–563 (1998).

22 Mallinckrodt CH, Clark WS, Stacy RD. Type I error rates from mixed-effects model repeated measures versus fixed effects analysis of variance with missing values imputed via last observation carried forward. *Drug Inf. J.* 35, 1215–1225 (2001).

23 Mallinckrodt CH, Clark WS, Stacy RD. Accounting for dropout bias using mixed-effects models. *J. Biopharm. Stat*. 11, 9–21 (2001).

24 Mallinckrodt CH, Carroll RJ, Debrota DJ *et al.* Assessing and interpreting treatment effects in longitudinal clinical trials with subject dropout. *Biol. Psychiatry*. 53, 754–760 (2003).

25 Mallinckrodt CH, Scott CW, Carroll RJ, Molenberghs G. Assessing response profiles from incomplete longitudinal clinical trial data with subject dropout under regulatory conditions. *J. Biopharm. Stat*. 13, 179–190 (2003).

26 Molenberghs G, Thijs H, Jansen I *et al.* Analyzing incomplete longitudinal clinical trial data. *Biostatistics*. 5, 445–464 (2004).

27 Carpenter JR, Kenward MG, Evans S, White IR. Last observation carry-forward and last observation analysis. Letter to the Editor. *Stat. Med.* 23, 3241–3244 (2004).

28 Beunkens C, Molenberghs G, Kenward MG. Tutorial: direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clin. Trials*. 2, 379–386 (2005).

29 Kenward MG, Molenberghs G. Last observation carried forward: a crystal ball? *J. Biopharm. Stat.* 19, 872–888 (2009).

30 Little RJA. Selection and pattern-mixture models. In: *Advances in Longitudinal Data Analysis*. G Fitzmaurice, M Davidian, G Verbeke, G Molenberghs (Eds). CRC Press, London, UK 409–431 (2008).

31 Kenward MG, Molenberghs G, Thijs H. Pattern-mixture models with proper time dependence. *Biometrika*, 90, 53–71 (2003).

32 Scharfstein DO, Rotnizky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *J. Am. Stat. Assoc.* 94, 1096–1146 (1999).

33 Tsiatis AA. *Semiparametric Theory and Missing Data*. Springer, New York, NY, USA (2006).

34 Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley, Chichester, UK (1987).

35 Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am. Stat*. 61, 79–90 (2007).

36 Harel O, Zhou X-H. Multiple imputation: review of theory, implementation and software. *Stat. Med*. 26, 3057–3077 (2007).

37 Meng X-L. Multiple-imputation inferences with uncongenial sources of input. *Stat. Sci.* 9, 538–558 (1994).

38 Kenward MG, Carpenter JR. Multiple imputation: current perspectives. *Stat. Methods Med. Res.* 16, 199–218 (2007).

39 Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall/CRC, Boca Raton, FL, USA (2008).

40 Kenward MG. Selection models for repeated measurements with nonrandom dropout: an illustration of sensitivity. *Stat. Med*. 17, 2723–2732 (1998).

41 Carpenter JR, Pocock S, Lamm C. Coping with missing values in clinical trials: a model based approach applied to asthma trials. *Stat. Med.* 21, 1043–1066 (2002).

42 Yun S, Lee Y, Kenward MG. Using hierarchical likelihood for missing data problems. *Biometrika* 94, 905–919 (2007).

43 Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann. Econ. Soc. Meas*. 5, 475–492 (1976).

44 White IR, Carpenter JH, Evans S, Schroter S. Eliciting and using expert opinions about non-response bias in randomized controlled trials. *Clin. Trials* 4, 125–139 (2007).

45 Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-mixture models. *Biostatistics*. 3, 245–265 (2002).

46 Keene O. Intent-to-treat analysis in the presence of off-treatment or missing data. *Pharm. Stat*. 10, 191–195 (2011).

47 Hughes S, Harris J, Flack N, Cuffe R. The statistician's role in the prevention of missing data. *Pharm. Stat*. 11, 410–416 (2012).

48 Mallinckrodt C. *A Practical Guide for Preventing and Treating Missing Data in Longitudinal Clinical Trials*. Cambridge University Press, Oxford, UK (2013).

Website

101 London School of Hygiene and Tropical Medecine. MissingData website. www.missingdata.org.uk