CLINICAL INVESTIGATION

Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation

Sensitivity analyses are commonly requested as part of the analysis of longitudinal clinical trials when data are missing. There are many ways in which such sensitivity analyses can be constructed. This article focuses on one particular approach, so-called controlled imputation. This combines two statistical ingredients, pattern-mixture models and multiple imputation. The aim is to assess sensitivity of the original conclusions to alternative assumptions about the statistical behavior of the patients' outcomes following dropout and withdrawal. Such assumptions must reflect postulated treatment compliance when intention-to-treat-like inferences are required. Many such scenarios could be considered, depending on the clinical setting. The advantage of this approach is that it makes such assumptions explicit in the sensitivity analysis and hence readily accessible to the user.

Keywords: *de facto* • *de jure* • estimand • intention to treat • missing at random • multiple imputation • 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
mike.kenward@lshtm.ac.uk

A recent article in this journal [1] provided a review of the handling of missing data in longitudinal clinical trials. One important recent development discussed there was the need for coherent sensitivity analysis in such settings. Several possible approaches to this were outlined, and it is likely that others will appear in the future. One of the approaches discussed, sometimes called controlled imputation, is becoming more widespread in its use and it is the purpose of this article to provide an overview of this type of sensitivity analysis that is based on existing developments. It is first necessary to provide sufficient background to the general problem of sensitivity analysis in longitudinal clinical trials, and for this the main points from [1] will be revisited. This article will then turn to the method itself, first providing a nontechnical description which it is hoped will clarify some of the key points. There is an almost limitless range of conditions and treatments for which trials may be run, and no single sensitivity analysis can possibly be appropriate for them all. To reflect this, the overall approach of controlled imputation provides a method for framing and conducting sensitivity analyses, which can accommodate a wide range of possible settings. Some examples will then be used to illustrate this, emphasizing the point that these may be of value for certain trials, but many other possibilities exist which may well be more appropriate in other settings. Within the range of alternatives there are some important distinctions to be made that are not always fully appreciated, and these are closely bound up with computational implementations. An attempt is made to shed light on these. Finally, reference is made to some of the currently available software implementations that may be of value to those wishing to use these methods.

It is now widely appreciated that missing data introduce ambiguity into the statistical analysis of data that are different from so-called 'statistical uncertainty' (e.g., [2,3]). One way to view this distinction is through the property that such ambiguity does not

FUTURE SCIENCE part of fsg

diminish as the sample size increases. As a consequence there are many ways to approach the problem of missing data, and the size of the total literature on missing data is now vast, and probably too great for any one individual to absorb. Even within the special setting of clinical trials the literature is very large and rapidly expanding. The appropriate handling of missing data in this setting has been of particular concern to the regulatory authorities, and in 2010 two key publications appeared. One, from the US National Research Council (NRC) of the National Academies [4], was sponsored by the US FDA. This is a comprehensive document, amounting to a monograph on the subject, that has already had considerable influence on practice. The other [5] is from the European regulators and was, in contrast to the NRC, produced by the regulators themselves. For a discussion of these two documents, in particular of their differences in viewpoint, see for example [6].

The estimand

To make meaningful judgments about the competing merits of different statistical analyses it is essential that the aims of the analyses are set out unambiguously. Failure to do this has led to considerable fruitless debate in the missing value literature. In the light of this, the NRC document uses the concept of an estimand. We can think of this as the target of the analysis, it may be a quantity to be estimated, or about which a statistical test is to be conducted. Another distinction is also useful at this point: the occurrence of a missing value may be treated in two ways, leading to two very different types of estimand. First, the missing value may itself be regarded as part of the patient response. A good example of this is dropout being defined as treatment failure. In such cases the missing data have been defined away, and essentially this is no longer a missing data problem. This can be contrasted with the second situation in which the occurrence of missing data is a nuisance to be accommodated in the analysis, in other words, if all the data were available the analysis would be comparatively straightforward. This second case is the one with which we are concerned here. One implication is that the definition of a missing value depends on the aims of the statistical analysis, in other words, the estimand. To avoid ambiguity, the term recorded will be used for a measurement that has a value assigned to it in the trial database. All unrecorded measurements are missing, while a recorded value may or may not be missing. In the light of this, two types of estimand are now introduced that originate from [7]. For this the explanation from [1] is used: *"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 clear from this that the definition of a missing value requires consideration of the treatment taken by the patient. Or, more generally, whether or not there has been a protocol deviation, which will be shortened in the following to just deviation. If an intended measurement is not recorded then obviously it is missing whichever of the two types of estimand is being used. However, if an intended measurement is recorded, but not under the regime specified by the protocol, then it is missing for the *de jure* estimand but not for the *de facto* one. Another crucial implication is that when we target the *de facto* estimand, and measurements are not recorded, to address the estimand it is necessary to incorporate into the analysis the actual treatment regime, either assumed or known, to apply when those measurements should have been collected. By contrast, for a *de jure* estimand, when data are not recorded, we must assume in the analysis that the protocol was followed, whether this was the case or not. This points to an interesting contrast. When all data are recorded, we can ignore actual treatment allocation in the analysis for a *de facto* estimand but not for a *de jure* estimand. When data are not recorded the opposite is true for a *de facto* estimand: it is necessary to specify the treatment regime taken, while, by contrast, we can simply assume no deviations for the *de jure* estimand, setting to missing any measurements recorded following a deviation. We use the term scenario to refer to the supposed (or known) treatment regimens followed by patients with missing data. Any analysis that targets a *de facto* estimand when data are missing must be making assumptions about the scenario assumed, either explicitly or implicitly. One problem with simpler *ad hoc* analyses is that they do not make these assumptions clear.

Missing data mechanisms

As a final step before considering the general approach of reference-based imputation, it is necessary to link the ideas set out in the previous paragraph with Rubin's classification of missing data mechanisms [8]. These are described here from a frequentist perspective, which differs to some extent from Rubin's original

exposition. Here, and in the remainder of this article, we consider only missing data caused by dropout or first deviation, that is, monotone missing data in the longitudinal setting. From a statistical perspective, a missing data mechanism is the probability model that governs the occurrence of missing data. With dropout, as here, it is a mechanism that specifies the probability that dropout (or first deviation) will happen at some time, quite possibly unknown, following a measurement time up to and including the following time. Such a mechanism is said to be missing completely at random (MCAR) if this probability does not depend on any of the variables involved in the analysis: outcomes, treatment taken, baseline covariates and so on. Under an MCAR mechanism, those patients remaining in the trial at the end, the so-called completers, represent a genuine random sample of those randomized, and so can be validly analyzed as if they were the only subjects randomized. Unfortunately, such an assumption is very implausible. Moreover, although in practice, using only the data under analysis, MCAR can be shown not to hold when it does not, the converse is false: it cannot be shown to hold when in fact it does. The second class of mechanism, missing at random (MAR), plays a pivotal role in missing data methodology, but not necessarily because it is plausible. It can be regarded in the current setting as the least restrictive set of assumptions that allows a valid analysis from the observed data only. To define this we use the idea of the patient's history. Define the history of a patient at a given time of measurement as the collection of all observed data on that patient up to that time, including treatment allocation, baseline covariates and previous (not current) outcomes. It is assumed that an analysis would include all these variables. Then, a dropout (deviation) mechanism is said to be MAR if conditional on the history at the time, the probability of dropout is not associated with the unobserved outcomes. This can be expressed in another way that is more relevant to the following exposition. Suppose that two patients share exactly the same history up until a certain time, at which point one patient drops out (or deviates) and the other remains (does not deviate). Then, under MAR, the future conditional statistical behavior of the two patients given their common history is identical. From this, it can be seen how a valid analysis can be constructed under MAR. Because both patients share the same conditional future statistical behavior, this conditional distribution can be estimated from those who remain and applied to those who do not, or who deviate. Likelihood-based analyses are valid under MAR, provided the chosen model holds. Further, under MAR, the probability of dropout or deviation can be estimated from the observed data and incorporated into less fully parametric analyses to provide validity for these under MAR [9]. In spite of its convenience, MAR remains a strong assumption and, most importantly, cannot be confirmed from the data under analysis [10], although in very special settings it may follow from the trial design [11]. If neither MCAR nor MAR hold then the missing data mechanism is said to be missing not at random (MNAR). From the second expression of MAR using the two patients with the same history, we can see that MNAR implies that these two do not share the same future statistical behavior, and this might reasonably be what is expected in many real settings. In fact, MAR has a very strong implication here about the treatment scenario. If two patients with the same past are assumed to share the same future statistical behavior then by implication they also share the same treatment scenario, which in turn implies a *de jure* estimand. Only when there are no deviations, or no treatment differences of any sort, will such an analysis correspond to a *de facto* scenario and in these cases there is anyway no difference between *de jure* and *de facto*. This linking of the MAR assumption and the *de jure* estimand is a key component of the controlled imputation approach to sensitivity analysis.

Sensitivity analysis

Although strictly not essential, most forms of sensitivity analysis in the missing data setting use the MAR assumption as an origin, and the sensitivity analysis assesses the consequences of departures from the assumptions underlying this. Given that there are many ways of varying these assumptions, there are many ways of approaching such analyses. The NRC report provides a discussion of some alternatives, and a wide range of other approaches are discussed in [3] (part V), [12], [13] (Chapter 10) and [14] (part V). A generic framework for sensitivity analysis is given in [1]:

- A clear definition of the estimand of interest;
- The assumptions under which the primary analysis is valid for this estimand;
- A nomenclature for practically relevant and accessible departures from these assumptions;
- Valid methods for assessing sensitivity to these assumptions.

The primary analysis from the second point applies when no data are missing. It is the intended analysis for complete data. For continuous outcomes, this analysis will commonly be a simple comparison of the means between randomized groups at the final, or some other key, time point, possibly adjusted

for baseline covariates. With no missing data such an analysis can be justified by the trial randomization alone. In the following, this will be used as the illustrative primary analysis, noting that the general principles to be described can be used for other primary analyses provided that they are based on a comparison of the randomization groups. Hence, the primary analysis does not reflect postrandomization treatment compliance or other deviations and, for this reason, such primary analyses will be called design-based analyses. In practice, when data are missing, alternatives to the simple design-based analysis are required, but these should, under appropriate assumptions, have the same target estimand as the primary analysis. For example, with continuous outcomes a so-called mixed model repeated measures analysis is often chosen as the primary analysis for the incomplete data.

The class of sensitivity analyses to be discussed below then is focused on design-based analyses and their target estimand, and the aim is to assess the behavior of the results from such analyses under particular departures from MAR. One of the most challenging aspects in formulating sensitivity analyses follows from the third point: the departures from MAR captured by a particular sensitivity analysis must be expressible in such a way that they can be appreciated by all those who need to make use of the results from this analysis, the majority of whom will probably not be statistical experts. The controlled imputation method to be described now addresses this need through the explicit formulation of MNAR statistical models for dropouts (or those who deviate). An important special class of controlled imputation approach, the so-called reference-based methods, constructs postdeviation/postdropout treatment scenarios from other arms of the trial. The feature that all controlled imputation methods have in common is the ability to construct the posited MNAR models using components from an MAR model, with the possible addition of fixed, known, sensitivity parameters. Thus no direct estimation of an MNAR model is required. This leads to great simplification in practice.

Controlled multiple imputation

It has been seen above that analyses based on the MAR assumption correspond to *de jure* estimands. It has also been seen that, under MAR, the future conditional statistical behavior of the two patients given their common history is identical. Hence one form of sensitivity analysis that allows departures from the MAR assumption can be formulated in terms of proposed or known deviations from protocol treatment. A particular set of assumptions about this behavior can be viewed as one possible *de facto* scenario. It is important to note that these assumptions apply to the statistical behavior of the missing data and do not affect the primary, design-based analysis. The aim of the sensitivity analysis is to examine the impact of these alternative assumptions on the results obtained when the primary analysis method is used.

The implementation of this approach in practice brings together two statistical ingredients. The first is a pattern-mixture model [3] (Chapter 16). In simple terms, this allows patients with different dropout/deviation patterns, and possibly other differences such as reason for dropout, to have different underlying outcome models. The overall model for all patients is then a mixture of such component models. For obvious reasons, the parts of the component models that correspond to missing data cannot be directly identified from the observed data. Hence, to use such models in practice, additional assumptions are introduced that allow unidentifiable parts of models to be identified or 'borrowed' from other groups of subjects. Sometimes these assumptions are expressed as constraints on the parameters of the component models. In fact, MAR can be shown to correspond to one very stringent set of such constraints. Early developments of pattern-mixture models used what could be termed 'within-group' constraints in which information is borrowed from other patients in the same randomized group. In control-based imputation methods, scenarios are formulated that define the future statistical behavior of the outcomes from each patient that deviates or drops out. These can take many forms, depending on the context. Such scenarios may, for example, modify the statistical behavior implied by the MAR assumption in some simple way, such as changing the predicted means by a given percentage. An important special class of controlled imputations consists of the so-called reference-based methods. These sensitivity analyses define the future behavior of patients with missing data according to some postulated pattern of treatments taken using information from other groups. So in contrast to 'within-group' constraints, the required behavior is 'borrowed' using a model for the missing data that are estimated from one of the other randomization groups. For example, if a subject withdraws from active treatment and subsequently takes no other, then their future behavior may be modeled on the behavior of those in a placebo group. Very commonly there will a reference group, a standard treatment for example, from which such borrowing will be made and this provides the name for the overall class of methods. In any particular instance the chosen set of mappings of behavior among randomization groups

defines the scenario, in the sense of the word used above. There are very many ways in which such scenarios can be constructed and the choice in practice depends critically on the particular clinical setting. Some examples are described below, both reference based and others, that have been found useful, but many others could be considered. All scenarios share a common property however. Each provides a statistical model for the conditional behavior of the missing data for each subject who deviates or drops out. This model can differ among randomized groups, and among reasons for, and times of, deviation or dropout. This model can be estimated from the observed data.

Having constructed a model for the entire set of outcomes, both observed and missing, it is necessary to use this in the sensitivity analysis. For this the second statistical ingredient is required: Multiple Imputation (MI) [13], see also [14] for a recent text. The pattern-mixture model allows missing outcomes to be imputed under the chosen scenario and in this way can be used to 'complete' the data set. The primary (design-based) analysis can then be applied to this completed data set. However, to produce a valid analysis it is necessary to account for both the fact that the imputation model has been estimated from the observed data and to reflect properly the loss of information due to the missing data. Without this a valid comparison cannot be made with the MAR-based *de jure* analysis, or other possible *de facto* scenarios. It turns out that the use of MI with pattern-mixture models corrects for both of these requirements. It is important at this point to be clear about the precise nature of the sensitivity analysis being used. Under the alternative scenario on which the pattern-mixture model is based, and which corresponds to a nonrandom missing data mechanism, treatment compliance in at least one group will not follow the original protocol. Yet, in keeping with the randomization, and in accordance with intention to treat principles, the original design-based analysis is retained, as would be used if the data were complete. Hence, it is important for this class of sensitivity analysis that the primary analysis does not reflect the postulated data generating mechanism as specified by the chosen scenario. The aim is to assess the robustness of the conclusions from the primary analysis as the underlying scenarios are altered. This should be contrasted with an analysis in which both the postulated data generating mechanism and the subsequent analysis are changed; the latter being modified to match the former. The actual postulated treatment adherence will be incorporated into such an analysis. Such an analysis is well defined and valid, but answers a very different question. This distinction is made as follows in [7]: *"We distinguish between two forms of sensitivity analysis, both of which have a role, and both of which can be formulated using the proposed multiple imputation framework. In the first, one assesses the impact of alternative postulated behaviors of the missing data on the conclusions from the original method of analysis. There is no full likelihood equivalent of such an analysis. It answers the question: how robust are the inferences from our analysis to misspecifying the assumptions about the behavior of the missing data (in the chosen directions)? In the second type of sensitivity analysis, the alternative postulated behavior is used both to impute the missing data and is acknowledged in the analysis model. Hence this answers the question: what would be the consequences of constructing the entire analysis under the chosen alternative assumptions? In this approach the imputation and analysis models are congenial, and there will be an equivalent analysis based on full likelihood."*

Failure to appreciate this difference can lead to mistaken attempts to assess the properties of one approach through the behavior of the other. See for example [15] and the response [16].

A reference-based example of the controlled imputation approach to sensitivity analysis was first described in [17], with some extensions given in [18]. A comprehensive account linking the approach to *de jure* and *de facto* estimands and introducing several alternative scenarios was provided by Carpenter *et al.* [7]. Discussions of one particular scenario (the so-called jump to reference) are provided in [19] and [20]. Two recent texts on handling missing data in clinical trials, [21,22], also discuss the approach. At present, controlled imputation methods for other outcomes, principally count, binary and event-time, are under development. A comparatively simple method for counts is described in [23].

Some example scenarios

We now describe in a nontechnical way some simple examples of scenarios that have been proposed for control-based imputation. These represent a mixture of reference-based methods (M1, M2, C1) and others (M3, C3, C4). For this we closely follow the development in [7]. It is assumed that there are two treatments under comparison, a new treatment that is being compared with a reference. In some settings, the reference treatment might be a standard, established treatment, in others, a placebo. In more complex settings, with more than two treatments for example, the basic scenarios illustrated here can be extended in a variety of natural ways, depending on the setting.

(M1) Jump to reference

It assumed that, following dropout, a patient's mean follows a profile that is the same as that of a patient

from the reference group. An obvious example when this might be considered is when a patient ceases taking treatment and the reference is placebo. This latter is sometimes called 'placebo imputation,' although in practice this more commonly refers to the conditional analogue of this, which we call 'copy reference' (see below).

Such a change may be seen as extreme: the patient immediately reverts, following dropout, to the reference profile, losing any gain that might have been made under the randomized active treatment. Hence, using the reference group in this way might be used as a worst-case scenario in terms of reducing any treatment effect because patients on active treatment who have withdrawn will lose the effect of their period on active treatment. It may well not be an appropriate scenario for treatments with a long lasting, or even permanent, impact.

Hence, this scenario might be considered for treatments with only short-term effect.

(M2) Copy differences in reference

This scenario provides a contrast to the extreme effect of jump to reference by assuming that in the future a dropout continues from their established position, but the subsequent changes in mean profile follow that of the reference arm. So, for example, if a particular rate of decline in mean were seen under the reference treatment this rate would then be applied to future outcomes for the dropout. In other words, the patient profile following dropout tracks that of the reference arm, but starting from the benefit already obtained. In contrast to the previous scenario, this would be more relevant when changes on outcome were not subject to rapid alteration under treatment change.

Such a scenario might be used, for example, in an Alzheimer's study where treatment halts disease progression, but after stopping therapy the disease continues to progress [26].

(M3) Last mean carried forward (marginal)

There has been much criticism of the *ad hoc* method of handling missing data known as 'Last Observation Carried Forward (LOCF).' See for example [1,24–27]. A more principled statistical approach that provides one interpretation of the rationale underlying LOCF can be constructed using the current approach. It is not an example of a reference-based imputation because it does not, at least in its simplest form, apply models from one arm to imputations in another. It is assumed that, following dropout, the patient follows a mean profile that does not change with time, in other words, an average level is established that is maintained to the end of the trial. In contrast to LOCF, individual imputations incorporate appropriate random variation, and this provides the proper underpinning of the method. In the MI/pattern-mixture version of the approach, it is the marginal profile that remains constant. There is a conditional analogue that is introduced below.

(M4) Marginal delta method

The so-called delta method can be used in various ways, with and without copying from other groups. The common feature of such methods is the addition of a chosen increment (often labeled delta, hence the name) to the marginal mean following dropout. This can be done once immediately following dropout to provide a one off 'kick' to the profile, or can continue for future times, perhaps increasing as time passes. These increment(s) can be added to various scenarios. For example, when added to a constant mean, as seen in the last mean carried forward scenario, a linearly increasing/decreasing profile can be constructed. Alternatively, the increment(s) can be added to the mean profile from those from the same group who do not dropout. Or, a proportion of the difference between the active and reference profiles can be added as an increment, generating a scenario that lies between MAR and jump to reference. In each case, an attempt is being made to capture a dropout behaving in a systematically different way to those who remain, with this behavior captured by one or more parameters, the delta's. This means that the delta method has a quantitative aspect that is absent from the other scenarios, and the size of delta can be treated as a sensitivity parameter, in other words, it can be varied across a range of values. Many alternatives exist within the delta approach and appear to be gaining some popularity in practice.

An example where this has been used is a trial on the treatment of pulmonary arterial hypertension, with a 6-min walking distance as an outcome. Dropouts were expected to do steadily worse on this score as time passed. Results from different values of delta formed the basis of a tipping point analysis [28].

We finish by considering three conditional scenarios.

(C1) Copy reference

This scenario is the analogue of jump to reference, but the conditional profile given the history is copied from the reference group not the marginal. It is less extreme than jump to reference and, in a certain sense, allows the patient to continue from the level achieved under the active treatment as though it had been achieved under the reference treatment. One consequence is that if a patient on active treatment is above the reference mean then this positive residual

will feed through into subsequent observations, to a degree determined by the correlation pattern in the reference arm. Hence, the patient's profile will slowly decay back toward the mean for reference at later times. Whether this makes sense in any given setting depends wholly on the clinical context. There are clear examples where it is not appropriate. It is computationally particularly convenient to implement however under the conditional MI algorithm, sometimes called the regression method. One difference between copy reference and the marginal scenarios becomes particularly apparent when the implied marginal profile for dropouts is recovered. This is seen, necessarily, to be a function of the dependence parameters (variances and correlations) as well as the marginal means, and this has implications for the corresponding estimand. We might ask when it is appropriate for the estimand to involve these parameters. This is not necessarily an argument against using copy reference, rather that, when used, it should be clear what question is being answered in the resulting sensitivity analysis.

(C2) Last mean carried forward (conditional)

This scenario follows the same principle as M3 above except that it is the patient's *conditional* mean that is carried forward. This means that to a certain extent the mean carried forward will reflect how successful or otherwise the individual patient has been up to the point of dropout. Those who have done well up to that point, for example, will retain that advantage (on average) to the end of the trial.

(C3) Conditional delta method

Again we take an approach originally applied to the marginal version, except that the same manipulations are applied to the conditional means. In this case, it follows that the delta increment is applied to the conditional not the marginal mean. And as we have seen with C1 above, it is important to understand what this implies about the resulting scenario. In particular, when the marginal implications of the conditional model are assessed it is important that these reflect what is required. Consider a simple example in which a one-off 'kick' is applied following dropout, with no further impact. In the marginal scenario only the mean following dropout is affected. However, when this is applied in the current condition scenario, the one-off delta increment influences subsequent marginal means, and the degree of influence depends on the variance and correlations among the outcomes. Similar, but more complex, implications for the implied marginal profiles follow from the use of other conditional delta methods. As with scenario C1, we see that the consequences on the implied marginal profiles of making these modifications are a function of the dependence structure of the data. And the implication of this is also the same: in using these scenarios it is important to be sure that these reflect what is required.

Further points

An important feature of these methods is that the grouping for scenarios need not be based just on time of dropout or deviation. The choice of scenario could also depend on the reason for dropout, for example. Information on this is usually collected in trials, categorized in some way. It could well be that certain reasons are likely to be quite unrelated to the trial and patient outcome, such as dropout associated with a change in home location of the patient. An MAR scenario might be appropriate for such a dropout. Other reasons, such as lack of efficacy, might well require an alternative, nonrandom scenario. This potential for incorporating reasons for dropout into the analysis is another valuable feature of this approach. As yet, there are few other statistical methods available to the practitioner that allows this information to be accommodated.

Some have used the availability of the parameter delta in scenarios M4 and C3 to construct the so-called 'tipping point ' analyses [28,29]. In these, the value of the sensitivity parameter, here delta, is increased in steps until the statistical significance, or non-significance, of the of the primary analysis is reversed. The behavior of the dropouts implied by this value of delta can then be examined from a clinical perspective to assess its plausibility.

One great advantages of the MI-based approach to sensitivity analysis is the availability of the imputations themselves for inspection, in contrast to purely model based methods. They can be plotted as means over groups of patients to provide a graph of average behavior from different dropout/deviation groups, or as means from each patient to give individual profiles. In this away, the actual behavior implied by the different scenarios can be visualized, greatly helping their assessment from substantive perspective. Such an approach is usefully combined with the tipping point method, for example.

It should also be noted that the borrowing of information from other arms does not avoid the issues of selection bias which are ever-present in missing data problems. For example, those who continue in the reference arm are not a random sample of those originally randomized to that arm and might, in some settings, represent less sick patients. Hence, it is an important aspect of the resulting sensitivity scenario

that the statistical behavior being borrowed reflects those who do indeed continue in the reference arm. If this does not match the required scenario, this can be modified by techniques like the delta method. In this overview, we have also confined discussion of the reference method to deviations confined to the active arm. In principle, many other scenarios can be constructed, for example, with deviations and dropouts in the reference arm as well. A referee asks whether the problem of missing data should be handled first in the reference arm before imputing for the active arm. There can be no 'correct' answer to this; the different options represent different scenarios which need to be considered in the context of the trial setting.

This points raised in the previous paragraph are related to a further query from a referee, who suggests that the *de facto* scenarios might, in some settings, make better foundations for primary analyses than the MAR-based analyses, in those circumstances when MAR is clearly an implausible assumption. While this does appear to be a sensible route for some settings, it does not follow that the controlled imputation approach is then necessarily the appropriate method for such a primary analysis. We can separate the construction of a *de facto* scenario from the analysis that incorporates it. A key feature of controlled imputation is the deliberate mismatch between the data generating model and the analysis model. It can be argued that for a primary analysis these two models should coincide, and the analysis model should properly incorporate the mechanisms leading to deviation and dropout. Such modeling approaches are under development, but take us away from the controlled imputation method.

Resources

As yet, a full range of controlled imputation methods has not been implemented in commercial statistical software. However, the general approach has one great computational advantage. The various scenarios can be introduced into a conventional MI analysis through modification of the imputations generated under MAR, provided the chosen package allows access to the necessary ingredients. These are then manipulated to produce the required scenario-based imputations. This is certainly possible using the SAS procedure MI [30] and the Stata command MI [31], and it is not especially difficult to construct bespoke analyses for particular scenarios using these tools. More flexible tools have also been implemented as macros that have been built on existing facilities. The most flexible among these is a set of SAS macros, the development, implementation and dissemination of which were led by James Roger and which is freely available from [32]. This covers most of the scenarios described above and, in addition, allows the choice of scenario to be patient-specific, that is, it allows each individual scenario to be chosen according to other collected information, such as to reason for dropout. It is accompanied by examples and good documentation. A second, also well-documented, SAS macro by Bodhana Ratich and Michael O'Kelly provides analyses for the C1 and C3 scenarios and also includes a tipping point analysis. This can be obtained from the same website as above, under the heading DIA working group.

Concluding remarks

The rationale behind, and the basic principles underlying, reference-based imputation methods of sensitivity analysis have been described. These allow sensitivity analyses to be constructed from alternative postulated treatment scenarios following dropout and deviation. Several possible scenarios have been described that have been found useful in practice. They have already formed part of successful regulatory submissions. We reiterate here the important point that this is a general approach and is not tied to these particular scenarios. Many other possibilities exist, and a key feature of the overall approach is that it can be tailored to very specific, and potentially very different, settings. Computationally, analyses can be constructed readily by those with experience of building macros SAS or Stata. For others, there exist two macros in SAS that can be used for a range of scenarios and others are in development.

Future perspective

At the moment it does seem very likely that the use of controlled imputation will be used increasingly as a method for sensitivity analysis in regulatory submissions.

While the overall approach of the controlled imputation sensitivity framework is now well established, there are many technical details that need filling in. The extension of controlled imputation to other outcomes, such as counts, binary data and event-time data, is currently ongoing. Settings such as these in which the treatment effect may be a nonlinear function of the outcome means raise additional issues when constructing pattern-mixture models, some of which may carry over to the controlled imputation setting. It is also important to be able to assess the performance of newly developed statistical methods and, for this, a clear definition of 'expected behavior' is needed. In the current sensitivity setting, in which the data generating and analysis models are not the

same, special care is needed. In particular, to justify the use of controlled imputation, we need to be clear about the required behavior of Rubin's MI variance formula in this very special setting. Some progress has already been made on this (see [18] for example) and further work is also ongoing on this problem.

As important as the technical developments is the continued accrual of practical experience of using these methods. As emphasized throughout this article, controlled imputation is a framework, not a particular analysis. Up to now, simpler settings have tended to be explored in more detail. However, there is an enormous range of clinical settings in which such methods might be considered, some with considerable complexity, and experience gained from these that can be communicated widely will be invaluable in providing guidance for the future use of controlled imputation.

A further major issue for the future was briefly discussed earlier: there is increasing pressure to use *de facto* scenarios as the basis for primary analyses. However, there does not yet exist agreement on how this should be done, and an argument against using controlled imputation for this was outlined above. It is possible that other viewpoints might undermine this objection though. Or, it may be that more formal model based analyses will be required in which both the data generating and analysis models coincide. These will almost certainly have to incorporate more information on the processes presumed or known to lie behind deviation and dropout. Work is ongoing on such models and subsequent analyses and is likely to be the topic of much development in the immediate future.

Disclaimer

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

Acknowledgements

Much of my thinking on missing data has been shaped by discussion with J Carpenter, G Molenberghs and J Roger. I am particularly grateful to J Roger for providing me with information on available computer software for these methods.

Financial & competing interests disclosure

M Kenward is GSK Professor of Biostatistics at the London School of Hygiene and Tropical Medicine. He has consulted for GlaxoSmithKline, Pfizer and Otsuka 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 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.
- There are many potential approaches to constructing sensitivity analysis and no broad agreement as yet about the most appropriate routes to take.
- In sensitivity analyses based on reference-based approaches, the assumptions are formulated in terms of treatment scenarios that follow dropout and deviations. Many such scenarios are possible.
- Such sensitivity analyses provide a transparent linkage between the estimand and the assumptions being targeted.
- Macros for these analyses have been developed and are freely available.

References

1 Kenward MG. The handling of missing data in clinical trials. *Clin. Investig. (Lond.)* 3, 241–250 (2013).

2 Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Wiley, Hoboken, NJ, USA (2002).

3 Molenberghs G, Kenward MG. Statistics in practice. In: *Missing Data in Clinical Studies.* Wiley, Chichester, UK (2007).

4 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).

5 Committee for Medicinal Products for Human Use. *Guideline on Missing Data in Confirmatory Clinical Trials*. European Medicines Agency, London, UK (2010).

6 Kenward MG. Missing data in clinical trials: a data interpretation problem with statistical solutions. *Clin. Investig. (Lond.)* 2(1), 5–9 (2012).

7 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.* 23, 1352–1371 (2013).

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

9 Tsiatis AA. *Semiparametric Theory And Missing Data*. Springer, NY, USA (2006).

10 Molenberghs G, Beunckens C, Sotto C, Kenward MG. Every missing not at random model has got a missing at random counterpart with equal fit. *J. Roy. Stat. Ser. B.* 70, 371–388 (2008).

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

12 Daniels MJ, Hogan JW. *Missing Data In Longitudinal Studies: Strategies For Bayesian Modeling And Sensitivity analysis*. Chapman & Hall/CRC, London, UK (2008).

13 Rubin DB. *Multiple Imputation For Nonresponse In Surveys*. Wiley, NY, USA (1987).

14 Carpenter JR, Kenward MG. *Multiple Imputation and its Application*. Wiley, Chichester, UK (2013).

15 Seaman SR, White IR, Leacy FP. Comments on analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J. Biopharm. Stat.* 23, 1352–1371 (2014).

16 Carpenter JR, Roger JH, Cro S, Kenward MG. Response to comments by Seaman et al. on comments on analysis of longitudinal trials with protocol deviation: a framework for elevant, accessible assumptions, and inference via Multiple Imputation. *J. Biopharm. Stat.* 23, 1352–1371 (2014).

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

18 Kenward MG, Carpenter JR. Multiple imputation. In: *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods.* Davidian M, Fitzmaurice G, Verbeke G, Molenberghs G (Eds). Chapman & Hall/CRC, London, UK, 477–500 (2008).

19 Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharm. Stat.* 12, 337–347 (2013).

20 Ayele BT, Lipkovich I, Molenberghs G, Mallinckrodt CH. A multiple imputation based approach to sensitivity analyses and effectiveness assessments in longitudinal clinical trials. *J. Biopharm. Stat.* 24, 211–228 (2014).

21 Mallinkrodt CH. *Preventing and Treating Missing Data in Longitudinal Clinical Trials*. Cambridge University Press, Cambridge, UK (2013).

22 O'Kelly M, Ratitch B. *Clinical Trials with Missing Data: A Guide for Practitioners*. Wiley, Chichester, UK (2014).

23 Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharm. Stat.* 13, 258–264 (2014).

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

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

26 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).

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

28 Fritsch A, Curram J, Kenward MG. A missing data sensitivity analysis of the Phase III study PATENT-1 in pulmonary arterial hypertension (PAH). Presented at: *34th Annual Conference of the International Society for Clinical Biostatistics.* Munich, Germany, 25–29 August 2013.

29 Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials. *J. Biopharm. Stat.* 19, 1085–1098 (2009).

30 SAS Institute Inc. *SAS® 9.3 System options: reference (2nd Edition)*, SAS Institute Inc, Cary, NC, USA (2011).

31 StataCorp. *Stata statistical software: release 13*. StataCorp LP, College Station, TX, USA (2013).

32 Missing data www.missingdata.org.uk