Missing data in clinical trials: a data interpretation problem with statistical solutions?

Mike Kenward speaks to Laura Harvey, Assistant Commissioning Editor. Mike Kenward has been GlaxoSmithKline Professor of Biostatistics at the London School of Hygiene since 1999, with former positions at the Universities of Kent and Reading in the UK and research institutes in the UK, Iceland and Finland. His main research interests are in the analysis of longitudinal data, crossover trials, small sample inference in restricted maximum likelihood and the problem of missing data. He has coauthored three text books including *'The Design and Analysis of Cross-Over Trials'* with Byron Jones and *'Missing Data in Clinical Studies'* with Geert Molenberghs. He was formally a coeditor of *Biometrics*, and is currently an associate editor of *Biostatistics*. Over the last 25 years he has acted as a consultant in biostatistics, largely for the pharmaceutical industry, and he has given over 100 short courses worldwide on various topics in biostatistics, and has appeared as an expert witness in the US Federal District Court. He is currently writing a book on multiple imputation with James Carpenter, and preparing the third edition of *The Design and Analysis of Cross-Over Trials* with Byron Jones.

> "When data that are required for the given design are missing, this nearly always means that the original planned route to answering the study question (or questions) is no longer useable."

Q How would you assess the importance of a 'uniform approach' in dealing with missing data in terms of 'correctly' interpreting trial data and the impact of misinterpreted data?

To answer this we need to be clear, in particular, about what we mean by 'uniform approach', and 'correctly'. In any study, we plan to collect data for specific purposes, and so structure our design accordingly. The National Academy (NA) report expresses the goal of the analysis in terms of the estimand, I will use this term from now on. When data that are required for the given design are missing, this nearly always means that the original planned route to answering the study question (or questions) is no longer useable. In other words, the link between the data collected and the conclusions to be drawn about the estimand that underpinned the proposed design has been broken. This introduces inevitable ambiguity into conclusions that can be drawn about the original study goals from the data available. This is not a problem of statistical uncertainty, and does not diminish with sample size. In this sense it is not a purely statistical problem. If we are to keep the original study goals, and it can be strongly argued that this is typically what is required, then additional assumptions will have to be introduced to allow inferences to be drawn about these. Hence the conclusions drawn will depend, to a greater or lesser degree, on what assumptions we make and, most importantly, the validity of these additional assumptions will not be assessable from the data under analysis. No nontrivial conclusions can be drawn from a study with missing data without making such untestable assumptions.

Mike Kenward

Department of Medical Statistics, London School of Hygiene & Tropical Medicine, University of London, Keppel Street, London
E-mail: mike.kenward@lshtm.ac.uk

The main choice open to us then, given the study questions, is what assumptions will we make in order to draw conclusions?

Q So how would you interpret 'correctly' in this context?

I would interpret 'correctly' to imply an analysis that properly addresses the predefined estimand given what we know, first about the scientific background to the trial, second about the actual data collection process, and third the additional, untestable, assumptions made.

Q And 'uniform approach'?

The meaning of 'uniform approach' follows directly from this: it is any valid statistical analysis that properly links (through formal statistical rules) the analysis about the estimand to the goals using the additional assumptions made. Note that this will exclude a definition of uniformity that is given in terms of specific statistical methods, for example those based on analysis of covariance, or using particular imputation methods. If you follow what is known about the scientific background to the trial and the additional, untestable, assumptions (my first and third points above), then the suitability of any particular method in any given setting can be decided using formal and well established rules.

Given the necessary reliance on untestable assumptions it is sensible to accompany any primary analysis with an appropriate sensitivity analysis. Again, the requirement for a sensitivity analysis, and broad agreement on what constitutes an acceptable sensitivity analysis framework, and what does not, is also a desirable component of a uniform approach.

Q What is your opinion on the importance of a universal, principled, approach to handling the missing data perfectly?

By establishing universal approaches in terms of principles rather than particular procedures, the range of appropriate analyses and techniques is able to evolve over time as the subject develops.

Q How do you think the recent European Medicines Agency and NA guidelines on handling missing data will be interpreted and followed in the analysis of clinical trials?

There has in the past been an understandable desire on the part of the regulators to see analyses defined in terms of simple processes applied directly to the data and, as a consequence, acceptable analyses in the missing value setting have tended to be expressed in terms of data modification. The use of single imputation methods represents one such route that is commonly used. Unfortunately, the apparent simplicity and transparency of such an approach usually hides from clear view both the estimand implied by the subsequent analysis, and the assumptions that are required to hold for the analysis to correspond to any given estimand.

Q If these are 'hidden from view' then one must have to work backwards to clarify what the results mean?

Yes, and such 'reverse engineering' of analyses to understand their implications for the required assumptions and estimands can lead to surprising conclusions (e.g., [1]). In short, in the missing data setting, simplicity and transparency of the procedure of actual data manipulation in an analysis does not, in general, correspond to simplicity and transparency of the implied analysis (i.e., assumptions and estimand). These issues are, of course, inseparable from the clear definition of primary goal (estimand) of the particular analysis.

I see both sets of guidelines as part of the process of extending the understanding of this distinction and subsequent movement of the basis for judging analyses from the procedure used, to the principles that justify the choice of analysis in the light of the given estimand. This is not necessarily a familiar way of viewing the problem for regulators, or statisticians, who have extensive experience in these settings. The practical success of the guidelines can be judged by the degree to which this change occurs.

Q What differences have you noticed between the two sets of guidelines, do you see this potentially adding to the confusion over how to deal with missing data?

It is inevitable that the two sets of guidelines will differ considerably. The European guidelines are much more brief and has been written by the regulators themselves, with comments from industry statisticians and academics [2]. The US document are more extensive and are based on detailed input from a committee containing a very broad range of statisticians and medical researchers, chaired by a leading world expert on missing data [3].

Although the European guidelines have many improvements over the predecessor, at times the new guidlines still struggle to make clear the distinction that I set out above, between a procedural and a principled definition of an analysis defined through procedural arguments and those based on coherent principles. As a consequence some statements in the documents are simply self-contradictory or wrong.

Q Can you give examples?

Two important examples occur in the executive summary. First, "A positive regulatory decision must be based on an analysis where the possibility of important bias in favour of the experimental agent can be excluded." Excluding certain extreme and unreasonable analyses, such as worst case scenarios with binary outcomes, this is impossible to achieve. The presence, or otherwise, of "important bias in favour of the experimental agent" depends on the truth, or otherwise, of the untestable assumptions that underpin the analysis. A more reasonable requirement is that the sponsor should provide a coherent argument as to why such bias is unlikely, and make clear the assumptions on which this argument rests.

Q And the second?

The second being "The justification for selecting a particular method should not be based primarily on the properties of the method under particular assumptions but on whether it is likely that it will provide an appropriate estimate for the comparison of primary regulatory interest in the circumstances of the trial under consideration." This requirement, which is closely related to the first quote, is self contradictory for the same reason. If, as agreed later in the Executive Summary "the choice of primary analysis will be based on assumptions that cannot be verified" then clearly the assumptions proceed the properties – the latter cannot be evaluated without first defining the former.

Q What do you interpret these contradictory statements to mean in terms of the European Medicines Agency position?

While the European Medicines Agency guidelines are in many ways consistent with the points made in answers to your first question, the occurrence (even if rare) of self-contradictory statements such as the two singled out above, suggests that there remains some tension still between the 'procedural' and 'principled' view of the problem. If, overall, the guidelines cannot produce a coherent view, the message to sponsors is mixed and this is likely to inhibit progress.

Q And the NA document?

By contrast the NA document makes very clear that the starting point must be the goal of the analysis, as enshrined in the estimand, and that the steps that lead from this to the primary analysis must be coherent, based on conventional statistical and scientific argument and, in the course of this, make explicit the underlying untestable assumptions. Many different techniques may potentially be used and the document is careful not to place excessive weight on any particular statistical approach. In such a long document it is inevitable that one can disagree with detailed points; however, the overall message is clear and largely consistent. In this sense it provides the basis for progress. A key issue in this setting is the need for sensitivity analysis.

Q How do you think the two documents cover this issue and what do you see as important steps towards drawing up useful guidelines on this?

Both documents agree on this. But in spite of large academic literature, with many possible approaches developed, from a practical perspective the subject is still in its infancy. Much more experience is required before useful specific guidelines can be formulated, and it is not to be expected that either document will be the last word on this. It can be expected, however, that some agreement should be reached on what is meant by a sensitivity analysis. It is clearly not, for example, a collection of 'wrong analyses'. Without having a clear view on the best way to proceed in any given setting, there are still broad guidelines that appear sensible. Examples include the requirement that the sensitivity analysis is tied to the estimand and that it should reflect the impact of changes in underlying assumptions in a coherent and clinically relevant and transparent way. This excludes the commonly seen presentation of the results from a seemingly arbitrary collection of analyses addressing different estimands under an *ad hoc* collection of unclear assumptions. The NA guidelines contain a discussion of the appropriate framework for sensitivity analysis which acknowledges the early stage of the development of these. Much of this is couched in a fairly formal model-based statistical framework and some illustrations are provided.

Q And the European Medicines Agency discussion?

This discussion is rather less coherent, tending to be ambiguous. For example, "Compare the results of the full set analysis to the complete case analysis" (page 11, 'What are These Analyses?') and "Utilise

6   www.future-science.com   future science group   fsg

fsg future science group   **Clin. Invest.** (2012) 2(1)   7

retrieved dropout data if not already done for the primary analysis. If a patient has received other therapies after withdrawing from the study, a positive value for the primary end point at the end of the trial could be due, in part at least, to the switching of therapies for this patient. Analyses that downplay the positive outcome to give a more realistic view of the product being evaluated should be conducted" (page 11), which makes sense only with an intention-to-treat (ITT) type of estimand. The European guidelines are tending to move straight to procedures without first setting down the overall framework that should be guiding the sensitivity analysis, whatever the detail of the subsequent methods used. Once again, the estimand must be the starting point.

The most important impact of both sets of guidelines is the agreement that appropriate sensitivity analysis should be included as part of the overall analysis. While we can agree on some broad principles for these, and the NA document is rather better for this than the European Medicines Agency one, there is still a long way to go in formulating specific approaches that will have wide applicability. Much more experience is needed for this.

Methodology

Q How do you think regulators should decide whether a particular method is acceptable?

This should be based on the coherence of the statistical justification, given the clinical setting and estimand. An argument should be made as to why the proposed analysis targets the given estimand, together with a clear statement of the assumptions required for this argument, and why these assumptions are thought to be clinically relevant and appropriate. The specified sensitivity analysis should then address assumptions that are most likely to be the subject of disagreement and/or potentially influential on the results. Again, this choice needs to be justified.

Q Should the scientific community be more open to novel methodology and what is the best way to introduce new methods to data analysis?

If research is to develop, novel methodology must be introduced. Effective routes for this include demonstration, for example at meetings and conferences; education, through training workshops; in the subject matter literature, example and exposition as well as software development (through the production of macros and procedures in widely used packages [e.g., R, SAS, Stata]).

Q What happens if different analyses give different results in terms of estimated effects or in terms of statistical significance?

Statistical analysis is not a branch of accountancy! This is no different to the situation met by regulators when an overall submission is assessed. Many pieces of evidence are considered, and the decision is made from the totality of these, often containing inconsistent aspects. No decision should be based on the statistical significance from a single trial. If the analyses considered come from part of a coherent sensitivity analysis then the different results are informative, and contribute to overall understanding. Only if the analyses do not have a logical connection, that is, do not constitute a proper sensitivity analysis in the sense described above, will they potentially serve to confuse the picture.

Q What role does retrieved dropout information play in minimizing the impact of unobserved data?

I would not express this in terms of 'minimizing the impact of unobserved data,' I would say rather that such information is potentially related to the assumptions that must be made when considering how to include the data from such subjects in the analysis. In this sense it is potentially very valuable.

Design

Q Given the comparatively low cost for extra work/preparation at the trial design and analysis stages (compared to trial costs as a whole), why do you think there is still difficulty in this area and what role do you think regulatory bodies can play in providing clinicians with a go to list of questions to approach missing data?

There is a natural conservatism in this setting and, if approaches have proved successful in the past, a sponsor will understandably stick with them. A second concern lies with statistical methodology, which will be unfamiliar and, to some, more demanding technically. However, the same process of evolution has accompanied the application of statistical methodology in all its uses from its beginning and is a necessary part of its development. Many techniques currently regarded as routine were at one time novel and challenging.

The regulatory bodies are important instruments for change. The requirement should be for a coherent justification for the proposed analysis with respect to the given estimand, including a transparent and clinically relevant discussion of the underlying assumptions. In

the same way, a coherent justification can be required for the sensitivity analysis. A potentially wide-range and well-developed set of statistical approaches are available for this.

Q How can clinical trials be adapted/structured at the design stage to accommodate for missing data?

I am sure that there are many ways that this can be done, but again it is very context specific. Both sets of guidelines are agreed on the importance of attempting to collect key data (if not all) from subjects who dropout/withdraw. This can have a major impact on the quality of ITT-type analyses in certain settings.

Understanding, at least broadly, the mechanisms behind withdrawal can be of value. Some of these, such as length of trial or assessment procedure may permit some variation, and so allow some modification in the chance of this leading to withdrawal. Others, such as the consequence of side effects may be less amenable to this.

Note that the precise role of withdrawal may depend on the estimand, and this needs to be taken into account when adaptation is considered. In particular there are situations where withdrawal can legitimately be regarded as part of the outcome (possibly a failure) in which case the trial needs to reflect as well as is possible, for pragmatic estimands, the patterns of withdrawal match those expected in the population. Again this emphasises the need for care in the definition of 'missing'.

Terminology

Q How do we move to a situation where we have a shared and agreed terminology to use in trial design and data analysis? What should this terminology look like?

This is important, but difficult. It first requires an

agreement on what this terminology should be. One route to this might be a broad based committee along the lines of the authorship of the NA report. Our understanding of the problem needs to develop in a constructive and rational way.

Q A lot of current terminology has many different meanings and interpretations (e.g., ITT). Do you think this affects implementation of a shared terminology?

This is a serious difficulty. One consequence is that there has been much fruitless debate on the subject. The NA report is valuable in this respect. Meaningful debate in this area is particularly dependent on clear definitions of key concepts. Examples are an ITT/per protocol distinction, but also terms such as withdrawal, dropout and deviation, and what is meant by 'missing data'. For example, an observation actually obtained may be regarded as 'missing' if it has not been made under the conditions (e.g., treatment) required by the definition of the estimand.

Q Do you think that trial analysis methods being transparently stated in write-ups would open discussion on how data is interpreted and get the ball rolling on a uniform approach?

I would have thought that this was anyway an essential part of the design and analysis for any trial. It is hard to see how progress can be made without this.

Reference

1 Kenward MG, Molenberghs G. Last observation carried forward: a crystal ball? *J. Biopharm. Stat.* 9, 872–888 (2009).

2 European Medicines Agency Committee for Medicinal Products for Human Use (CHMP). Guidline on missing data in confirmatory clinical trials. *EMA/CPMP/EWP/1776/99 Rev. 1* (2010).

3 Panel on Handling Missing Data in Clinical Trials, National Research Council. *The Prevention and treatment of missing data in clinical trials.* National Academies Press, Washington, DC, USA (2011).

8 www.future-science.com future science group

future science group Clin. Invest. (2012) 2(1) 9