CLINICAL INVESTIGATION

INTERVIEW SERIES

Clin. Invest. (2012) 2(1), 11–17

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

"Multiple solutions might be available to achieve an acceptable outcome ... where the primary objective is to determine whether or not an acceptable evidence base exists to approve a marketing authorization application."

Robert Hemmings* & David Wright

Medicines & Healthcare products Regulatory Agency, 151 Buckingham Palace Road, Victoria, London, SW1W 9SZ, UK
*Author for correspondence:
E-mail: robert.hemmings@mhra.gsi.gov.uk

Robert Hemmings and David Wright speak to Laura Harvey, Assistant Commissioning Editor.

Robert Hemmings: Hemmings is Statistics Unit Manager at Medicines and Healthcare products Regulatory Agency (MHRA; formerly the Medicines Control Agency) and a member of the Committee for Medicinal Products for Human Use (CHMP), the body responsible for preparing the opinions of the European Medicines Agency on questions concerning medicinal products for human use. Hemmings is Chair of the CHMP's Scientific Advice Working Party with responsibility for preparing advice to the pharmaceutical industry on the appropriate tests and trials to conduct in the development of a medicine for marketing authorization. Hemmings is also a member of CHMP's Biostatistics Working Party with responsibility for giving advice on matters relating to clinical trial methodology across the EU regulatory network. **David Wright:** Wright has worked for the MHRA for 12 years as a statistical assessor. Wright is now Deputy Statistics Unit Manager and an Expert Statistical Assessor. Wright coordinates scientific advice for the MHRA, chairs the scientific advice review group, is Chair of the Biostatistics Working Party of the CHMP and is an alternate member of the Scientific Advice Working Party of the CHMP. Wright took a lead role in revising the CHMP guidelines on Missing Data in Confirmatory Clinical Trials.

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

The questions of how to avoid and how to address missing data represent two of the major challenges to the clinical trial sponsor, and to anyone wishing to interpret clinical trial data. A uniform approach is therefore beneficial in terms of promoting a considered approach to planning, conducting and, to some extent, analyzing a clinical trial. It would certainly be beneficial to have an approach that systematically encourages consideration of how to:

- Minimize missing data;
- Promote patient follow-up (whether or not the patient stays on trial treatment);
- Prospectively identify a sensible primary analysis and set of sensitivity analyses;
- Explore reasons for, and timings of, withdrawals in different treatment groups to further consider whether the planned analyses can be considered a comprehensive exploration of the problem and reliable for inference.

We would not advocate a uniform approach to data analysis in terms of the statistical methodology used or assumptions made, in this regard each trial and each

ISSN 2041-6792
10.4155/CLI.11.171

dataset should be considered uniquely both before it is generated and during analysis.

Q So in the absence of one 'perfect' statistical method to apply in the event of missing data, what is your opinion on the need for a universal, principled, approach to handling the missing data?

Considerations of how to address the missing data problem reach beyond traditional statistical methodology, from initial trial planning through to final inference. There is no statistical approach that handles missing data perfectly. Indeed, it is entirely fair to say that all methods handle missing data imperfectly. Once data are missing it is not possible to prove that a particular method of handling missing data is the most appropriate and whilst a systematic approach to the problem might be desirable to ensure that all important issues are addressed, one universal plan for statistical analysis is not obviously desirable.

In a regulatory context, considering a confirmatory clinical trial, the methodology chosen should be carefully selected and justified based on the precise (and clearly stated) question(s) that the trial is designed to address; this should include the approach taken to handle missing data. When reporting results and discussing whether efficacy is demonstrated, it is important to consider whether or not the estimated treatment effect and associated variability are likely to be biased to an important degree in favor of the test treatment (equating to a bias in favor of the null-hypothesis being rejected). Multiple solutions might be available to achieve an acceptable outcome in this regulatory setting where the primary objective is to determine whether or not an acceptable evidence base exists to approve a marketing authorization application.

Outside the regulatory setting, clinical trial data can be put to different uses. In the regulatory setting it is not necessary to make a positive decision about a new medicine if the risk of important bias is too high. In a wider context, a positive decision might have to be made, for example, about which of multiple interventions to recommend for public health benefits, or which of multiple interventions is more cost effective regardless of potential biases in the dataset. Priorities for what is considered an adequately 'reliable' analysis might differ, but even then it is not clear that any approach can be universally applied, as implied by the question.

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

The fact that there are many similarities between the two documents increases the chances that the recommendations made in these documents will be implemented in the design and analysis of clinical trials. The importance of planning and minimizing the amount of missing data is common to both documents. It is hoped that because both documents stress the importance of carefully planning and learning from what has happened in previous trials, the quality of information available to decision makers and the quality of the analyses that are provided by trial sponsors will improve.

Both documents also stress the importance of predefining the primary analysis and providing a justification for the choice of approach to handling missing data. The need for sensitivity analyses in virtually all analyses of clinical trials is also stated in both documents. It is hoped that this will increase the availability of sensitivity analyses at the time of regulatory submission. Not providing such analyses in the initial submission makes review of the application more difficult and can cause confusion over what should be considered the estimated treatment effect and whether sufficient evidence of efficacy has been provided.

The Committee for Medicinal Products for Human Use (CHMP) guideline aims to set out a series of principles that should be followed when planning, analyzing and interpreting a clinical trial; no CHMP guideline aims to be a textbook on a particular topic. This document should set the standards for sponsors and regulators addressing this issue without tying a sponsor to any particular methodology or statistical framework. We would therefore advocate an agreed set of principles rather than a single, 'uniform' approach.

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?

First, it is important to acknowledge the substantial degree of agreement between the two documents on many of the key issues that face sponsors when analyzing clinical trials that are part of a regulatory submission.

There are however, some differences between the two sets of guidance. Two are of particular importance.

Q Can you give details?

First, there is agreement that when designing a clinical trial an extensive search of similar studies should be performed, including finding out the amount, reason and timing for missing data in these studies. Recommendation 8 of The National Academy of Sciences report states:

"All trial protocols should recognize the importance of minimizing the amount of missing data, and, in particular, they should set a minimum rate of completeness for the primary outcome(s), based on what has been achievable in similar past trials."

The aim of minimizing the amount of missing data in a study is supported, as is the aim to reduce the amount of missing data in future studies. Setting a target would seem to address these aims, but if a study is done that fails to meet its target it could still form a successful part of a regulatory submission and this is the reason a minimum rate of completeness is not specified in the CHMP guideline.

Q And the second difference?

Second, the report states that "single imputation methods are sometimes used not as a method for imputation but rather as a convenient method of sensitivity analysis when they provide a clearly conservative treatment of the missing data."

In situations when single imputation methods are thought to clearly provide a conservative treatment of the missing data, then the methods provide a useful basis for regulatory decision making. Elsewhere in the National Academy of Sciences report it states that single imputation methods should not be the primary method of handling missing data. In the situation described where such an approach is thought to clearly provide a conservative estimate of the treatment effect its use in a regulatory submission is likely to be perfectly acceptable. It should be noted here that 'acceptability' should not be confused with being a regulatory demand – it was, and may still be, a common misapprehension that (only) single imputation methods are acceptable in a regulatory submission.

Q In another interview on this topic Mike Kenward highlighted two statements (see below) in the European Medicines Agency document as 'self-contradictory or wrong' [1] – would you care to comment?

"A positive regulatory decision must be based on an analysis where the possibility of important bias in favor of the experimental agent can be excluded."

"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."

The dialogue with statisticians from academia and industry on this topic is enjoyable and illuminating. Kenward is considering the problem within a statistical framework that is limited by the fact that the truth of the underlying assumptions cannot be verified (this, of course, is something all parties understand and agree). The CHMP document is written to guide clinical and statistical assessors in decision making, and to give advice to industry. In practice, whether or not an important bias can be excluded will be a judgement based on the particular dataset, the pattern of missing data and the statistical approaches used. As such, in terms of making a regulatory decision, which is the focus of the document, it is argued that exclusion of an important bias can often be substantiated. The broader interpretation of the sentence from the guideline is much more in line with the subsequent sentence from the authors critique, which is readily supported: "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." Indeed, a more appropriate sentence to focus on in the Executive Summary of the CHMP guideline is "An appropriate analysis would provide a point estimate that is unlikely to be biased in favor of experimental treatment to an important degree (under reasonable assumptions) and a confidence interval that does not underestimate the variability of the point estimate to an important extent." This clearly states what a company should aim for in specifying an appropriate primary analysis for a confirmatory clinical trial.

The second sentence highlighted for criticism is essentially the same as that described above and therefore our explanation above also applies. However, the highlighting of this sentence gives the opportunity to explain why this was included. This relates to a tendency in recent submissions for sponsors to use mixed-model repeated measures with only the following justification "Using mixed model repeated measures means that there is no concern over missing data as the method assumes missing at random and hence the analysis presented is unbiased." As a justification this is wholly incomplete, and the message from the guideline is that a sponsor should not focus on the properties of a method under those same assumptions that we all agree to be untestable, but on whether the estimate can be judged appropriate given all the peculiarities of the particular dataset concentrating on the patterns of, and reasons for, missing data and the consequences for the estimate and resulting inference.

We would also pick up on the comment that some parts of the document are only correct under an intention-to-treat (ITT) type of estimand. It should be noted that typically the ITT type of estimand is precisely the question being asked in the primary analysis of a pivotal study in a regulatory submission. This is clear from other regulatory guidance documents and is the reason for the focus on addressing how to provide a realistic estimate of the treatment effect for the ITT estimand in the CHMP guideline.

Methodology

Q In terms of implementing a uniform approach to trial analysis, what would you say to the argument that, in order for uniformity across clinical trials, all analyses should rest on an agreed assumption about the unobserved data and the reason that they are missing (missing value mechanism/ missing at random assumption)?

As stated above, once there are some missing data, any approach to trial analysis is imperfect to a greater or lesser degree. Certain methods will provide results that are unbiased under certain assumptions, but those assumptions are untestable and, to our opinion, assumptions such as 'missing at random' are too often implausible and the resulting analysis will address the artificial question of the effect that might be observed if all patients had remained on treatment, in particular if the analyses cannot make appropriate use of retrieved dropout data. As such, any single approach might perform well in some circumstances, but unacceptably poorly in others and examples where each and every commonly applied method performs badly (leading to an important bias in the analysis and reported results) have been seen in regulatory submissions. Therefore, we stress again the need to consider each trial and each dataset on its own merits when deciding on an analysis plan.

Q What are the difficulties in applying this assumption to all types of trials & analyses?

Different trials are asking different questions and therefore a uniform approach to trial analysis would lose this important observation. The National Academy of Sciences makes this point by defining the term 'estimand' and stressing that the most appropriate analysis for a study depends on the estimand. It is concurred that once agreement is reached on the question being asked then this helps specify the primary analysis method. This method will make assumptions about the missing data and a clear justification for the chosen method and the appropriateness of the assumptions being made should be provided.

Q How do you think regulators should decide whether or not a particular method is acceptable?

From the point of view of the drug licensing procedure, the regulators have the difficult task of balancing the need to avoid licensing medicines based on artificially inflated estimates of efficacy whilst avoiding the creation of an architecture for developing medicines that is unnecessarily penal. The regulatory decision is based on benefits, risks and uncertainties. With regard to efficacy data, the regulators will ask whether there is adequate evidence that the treatment works and, if so, whether the magnitude of the effects is clinically important and sufficient to offset the risks associated with the treatment. Therefore, establishing the presence of a treatment effect, and obtaining a reliable estimate of the magnitude of the effect are both important.

Throughout the CHMP guidance document, the regulator is instructed to consider all aspects of the missing data problem, not only the statistical analysis method used but the amount of missing data observed, the pattern of missing data in the different treatment arms (including the timing of, and reasons for, data being missing), the robustness of results to different analysis methods and so on. Because regulators are encouraged to consider the applicability of a particular method to a specific trial, in one particular situation a range of methods might be acceptable whilst other methods are very likely to be biased in favor of the test treatment and therefore unlikely to be acceptable for positive regulatory decision. In a different setting, those methods previously found unacceptable might be considered perfectly reliable. Hence, a regulator will not judge a particular method to be universally acceptable, or universally unacceptable. The burden to justify absence of important bias in a particular dataset is primarily on the trial sponsor.

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

It is felt that a wider dissemination of the problems brought about by missing data, and the flawed methods for handling missing data, is required. Once the problems are well understood, the need for better trials and better methods of statistical analysis should also be understood and the community should be open to improved approaches.

It is anticipated that novel methodology will be more extensively used in the future and trends in this direction are already being noticed in regulatory submissions. As previously discussed, this is welcome since some commonly applied methods will have very undesirable properties in certain settings and so increased attention to this difficult problem is welcome. Of course, as also described, some of the emerging methods will also have undesirable properties in certain situations! It is anticipated that flaws in the handling of missing data will persist since some sponsors and some regulators will cling to methods that have previously brought success. However, it should be stressed that providing new methods adhere to the principles set out in the guidance document, their use should be accepted, even as a primary analysis. In addition, if a company considers the approach taken previously in a particular area can be improved on, then they are very welcome to engage in dialogue with the regulators to explore using novel methodology in a regulatory submission.

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

It is expected that different analysis methods, based on different assumptions, will give different results and sponsors should not be unduly penalized for this. Having an acceptable primary analysis, supported by a range of sensitivity analyses based on different assumptions (of note, that is not simply multiple analyses based on variations of the same assumptions), is important where the level of missing data is non-negligible. When different analyses give substantially different results this provides important information on deciding the robustness of the results. If the different analyses are considered to make reasonable assumptions about the missing data and consistent evidence of efficacy is seen for the different analyses, this suggests that the evidence of efficacy is sufficient, enabling a decision on the benefit–risk trade-off. When considering the range of sensitivity analyses presented in support of an acceptable primary analysis, the change in the estimated treatment effect is more important than whether or not statistical significance is maintained.

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

Both the CHMP guidance document and the National Academy of Sciences document recommend collection of 'retrieved dropout' information. Retrieved dropout information has the advantage that it provides the measurement of an end point for a particular patient after they stopped taking the treatment to which they were allocated. This is of importance if, for example, the treatment effect at the end of the study is of key interest, and this is commonly the case. The concept of a patient discontinuing allocated treatment, but remaining 'on-study' should be promoted and can be considered consistent with the regulatory interpretation of the ITT principle.

The difficulty with this type of information is whether, and if so how, to include retrieved dropout information into the statistical analysis. This depends on a number of factors, again including the precise question that the study is designed to answer, but also whether or not the subject goes on to receive another treatment. If a patient receives no treatment after discontinuation of allocated study treatment then response could be said to reflect the 'failure' of study treatment and hence be an appropriate reflection for the primary analysis. If patients discontinuing treatment go on to receive another treatment (as is often ethically and clinically mandated), it may be difficult to decide whether to include such information in the primary analysis because of the changes in therapy recieved after dropout. However, the alternative of not having this information and therefore having to impute or model the likely measurement brings with it uncertainty, and hence having the retrieved dropout information to use, at least as one of a range of sensitivity analyses, is strongly preferred in many circumstances.

Design

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

The primary reason that difficulty remains is that there is no correct answer to the problem! Once a data point is missing, an observation is not measured, but 'guessed' (whether modeled or imputed). Because of this, any and every method can be open to criticism. Historically, the problem has not been well understood or well explored by trial sponsors, regulators or academics and methods were chosen which would, on many occasions, have extremely undesirable properties, and were only selected because they were easy to implement and (mis)interpret. The fact that they had undesirable properties was either not known or was overlooked.

Q And moving on from this situation?

Beyond this, there have been different views on what is trying to be achieved when handling missing data. Some would say that a method should try to estimate what would have happened if the patient had stayed on trial and on trial treatment as this represents the best estimate of the efficacy of the medicine in conditions of 'perfect use'. Others would argue that this question is fundamentally wrong and that we should derive an estimate of the efficacy of treatment that reflects the 'imperfect use' that will be seen if the medicine is used in practice (albeit the estimate of efficacy is made within the artificial setting of a clinical trial) and that therefore reflects patients discontinuing, and no longer benefiting from, treatment. This is addressed in the CHMP guideline, which expresses a preference for the latter approach in 'pivotal' trials for drug licensing.

Q How do you see the situation improving postguideline publications?

Hopefully the situation should improve following the publication of these two guidelines, which both stress the value that this extra work and preparation at the design stage can play in minimizing the amount of missing data, help pre-specify an appropriate primary analysis and a range of sensitivity analyses that can explore deviations from the expected pattern of missing data observed in the trial.

The guidance documents include principles that, although not formulated into a 'go to' list of questions, should, if followed, leave the trial sponsor with fewer problems when it comes to drawing inferences from the trial data.

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

The ambition should be to avoid rather than to accommodate for missing data. Trials should be structured to collect as much key information as possible so that the proportion of missing data, and thus the difficulty of needing to handle this, is minimized. A clear statement of the question to be addressed is critical and should be understood and agreed by all interested parties. The amount and pattern of missing data should be considered in defining the statistical analysis methodology, case report forms, trial conduct and monitoring. It can be prudent also to consider the missing data problem in the sample size calculation; not only that 'x% of patients will withdraw and therefore an additional x% of patients will be recruited', but the likely impact of an acceptable analysis methodology on effect size, variability and, hence, on statistical power.

Q In your opinion how do you know when a certain amount of missing data renders a study invalid? Do you think that there should be guidelines in place such that researchers know when their collected data is not suitable for use?

Various interested parties stated that the CHMP guideline should specify the amount of missing data that renders a study invalid. The guideline does not state a cut-off as it is not possible to do so. A study with missing data is, arguably, never invalid but as the proportion of missing data becomes greater, the certainty that the results of the study provide a reliable measure of efficacy reduces.

Q Yes, however, there must be a point at which the uncertainty due to missing data renders study conclusions unreliable. Could the study not then be said to be 'invalid'?

A trial demonstrating a clinically and statistically 'small' effect (however defined) would need only a relatively small proportion of missing data before the results were questioned, conceivably as low as 5–10%, although the figure depends on the distribution of missing data between treatment groups, timings of and reasons for missing data, the validity of the statistical analysis method selected and the robustness of results to different approaches. On the other hand, a trial with a clinically large and statistically extreme effect could be robust to a higher proportion of missing data. Different clinical indications and trial populations have their own characteristics and expectations for levels of missing data. If the amount of missing data is clearly higher than observed in previous similar trials the quality of the trial would undoubtedly be questioned. This, in particular for a non-inferiority or equivalence trial, could cast doubt on the conclusion where poor trial quality could bias the results in favor of showing similar effects of two treatments.

Further to this, two studies could have similar observed effects and amount of missing data, but if the pattern of missing data in terms of timing and reason for dropout were similar between treatment groups in one trial and different in another trial it is this difference that could cast doubt on the conclusions drawn from one of the studies.

Q Do you think guidelines to help trial analysis, on when the extent of missing data in a trial could present an interpretation problem, would be helpful to trial investigators?

Basic guidance on what amount of missing data might be accepted would represent an over simplification of the problem. Detailed guidance could be written of course, and indeed the CHMP guidance document outlines important principles for trials to be used as pivotal in regulatory submissions without tying sponsors to particular methodology, which might not be appropriate for their trial. The content of this guideline, and indeed the National Academy of Sciences guideline, could be propagated more widely.

Terminology

Q How do you see the fact that a lot of current terminology has many different meanings & interpretations (e.g. intention to treat), thereby affecting implementation of a shared terminology?

It is of considerably greater concern that trial sponsors do not try to clearly define the question(s) that a particular trial will try to answer, and do not routinely try to discuss difficulties of individual trial datasets or the pros and cons of different analysis methods when reporting trial results. In the absence of these important discussions, the usefulness of a shared and universally agreed terminology is reduced.

Q So would you say that increased and frequent communication between trial co-ordinators & regulatory bodies would be a crucial step towards a helpful agreed terminology?

Potentially yes, if these discussions were held more routinely then lessons learned and misunderstandings encountered could be formulated into guidance documents that discuss and promote a consistent definition to important terminology.

There is a focus in statistical literature on one statistical framework for the missing data problem, attempting to describe the pattern of missingness in relation to the chosen statistical model (e.g., missing at random). This has been somewhat abused as some regulatory submissions will limit discussion to whether a particular method gives unbiased estimates of effect under certain assumptions and not on the validity of those assumptions. More importantly, it is of considerably greater interest to understand the pattern of missing data and the impact of the imputation or modeling approach on the estimated treatment effect and associated variability, than whether a particular analysis method fits into a particular statistical framework.

Q So, in conclusion would you say that work on universal agreement to the approach to conducting and analyzing a trial on its merits will better serve the clinical trial community than efforts towards an agreed terminology?

It is not clear what aspects of terminology need to be immediately improved to implement, or to further discuss, the principles and framework outlined above and therefore, yes, in our opinion, an agreement on principles to guide a framework for design, conduct and analysis is more important than an agreement on terminology, but the two cannot be separated entirely.

Q What compromises do you foresee and what changes/actions would you deem necessary in implementing a shared approach?

This is a topic on which trial sponsors, academia, learned bodies and societies and the regulatory system will continue to learn and develop. It is an area where we expect guidance to be 'living' as preferred approaches emerge in different therapeutic areas and as newer statistical methodology gains wider applicability and wider use.

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?

Regardless of whether such reporting gets the ball rolling on a uniform approach, it is considered essential that when the primary analysis is reported in a clinical trial report, or in published literature, the method of handling missing data should be clearly stated and reasons why this approach was chosen and why it leads to reliable conclusions should be given. Without this the reader is not properly informed and the report should be considered deficient. Ideally, the paper should also describe key sensitivity analyses so that the reader can appreciate the robustness of the conclusions drawn to the assumptions made in the different analyses along with a discussion of the impact of missing data on the reliability of the trial conclusions. This is one aspect in which the sponsor's final position, and indeed the final regulatory assessment of a clinical trial, can differ markedly to the published version of the same trial and that situation cannot be optimal.

Financial & competing interests disclosure

R Hemmings and D Wright have no 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. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Reference

1 Kenward M. Missing data in clinical trials: a data interpretation problem with statistical solutions? *Clin. Invest.* 2(1), 5–7 (2012).