EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

2 July 2010
EMA/CPMP/EWP/1776/99 Rev. 1
Committee for Medicinal Products for Human Use (CHMP)

Guideline on Missing Data in Confirmatory Clinical Trials

Discussion in the Efficacy Working Party	June 1999/ November 2000
Transmission to CPMP	January 2001
Released for consultation	January 2001
Deadline for Comments	April 2001
Discussion in the Efficacy Working Party	October 2001
Transmission to CPMP	November 2001
Adoption by CPMP	November 2001
Draft Rev. 1 Agreed by Efficacy Working Party	April 2009
Adoption by CHMP for release for consultation	23 April 2009
End of consultation (deadline for comments)	31 October 2009
Rev. 1 Agreed by Efficacy Working Party	April 2010
Adoption by CHMP	24 June 2010
Date for coming into effect	1 January 2011

This guideline replaces Points to Consider on Missing Data in Clinical Trials (CPMP/EWP/1776/99).

Keywords	*Baseline Observation Carried Forward (BOCF), Generalised Estimating Equations (GEE), Last observation carried forward (LOCF), Missing at random (MAR), Missing completely at random (MCAR), Missing Data, Mixed Models for Repeated Measures (MMRM), Missing not at random (MNAR), pattern mixture models.*

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8416
E-mail info@ema.europa.eu **Website** www.ema.europa.eu
An agency of the European Union

Guideline on Missing Data in Confirmatory Clinical Trials

Table of contents

Executive summary

It should be the aim of those conducting clinical trials to achieve complete capture of all data from all patients, including those who discontinue from treatment. Whilst it is unavoidable that some data are missing from all confirmatory clinical trials, it should be noted that just ignoring missing data is not an acceptable option when planning, conducting or interpreting the analysis of a confirmatory clinical trial. The reason for missing data and handling of missing data in the analysis represent critical factors in the regulatory assessment of all confirmatory clinical trials. The main focus of this guideline is issues associated with the analysis of the primary efficacy endpoint where patients are followed up over time.

However, by careful planning it is possible to reduce the amount of data that are missing. This is important because missing data are a potential source of bias when analysing data from clinical trials. Interpretation of the results of a trial is always problematic when the proportion of missing values is substantial. When this occurs, the uncertainty of the likely treatment effect can become such that it is not possible to conclude that evidence of efficacy has been established.

In confirmatory trials the primary analysis is commonly performed on the full analysis set as this analysis is consistent with the intention to treat (ITT) principle. If data for some subjects are missing for the primary endpoint it is necessary to specify how all randomised patients can be included in the statistical analysis. However, there is no universally applicable method that adjusts the analysis to take into account that some values are missing, and different approaches may lead to different conclusions. To avoid concerns over data-driven selection of methods, it is essential to pre-specify the selected methods in the statistical section of the study protocol or analysis plan. Unfortunately, when there are missing data, all approaches to analysis rely on assumptions that cannot be verified. It should be noted that the strategy employed to handle missing values might in itself be a source of bias. A critical discussion of the number, timing, pattern, reason for and possible implications of missing values in efficacy and safety assessments should be included in the clinical report as a matter of routine. It will be useful to investigate the pattern of missing data in previous trials in the same or similar indications for related medicinal products. This could assist in identifying additional actions to minimise the amount of missing data during the conduct of the trial, the choice of the primary analysis method and in determining how missing data will be handled in this analysis.

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. 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.. An appropriate analysis would provide a point estimate that is unlikely to be biased in favour 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. The type of bias that can critically affect interpretation depends upon the objective of the study (to show superiority, non-inferiority or equivalence). As the choice of primary analysis will be based on assumptions that cannot be verified it will almost always be necessary to investigate the robustness of trial results through appropriate sensitivity analyses that make different assumptions.

1. Introduction (background)

Missing data are a potential source of bias when analysing clinical trials. Interpretation of the results of a trial is always problematic when the proportion of missing values is substantial. This problem is only partially covered in ICH E9 (Statistical Principles of Clinical Trials).

There are many possible reasons for missing data (e.g. patient refusal to continue in the study, patient withdrawals due to treatment failure, treatment success or adverse events, patients moving), only some of which are related to study treatment. Different degrees of data incompleteness can occur, i.e. measurements may be available only at baseline, or measurements may be missing at baseline, or may be missing for one, several or all follow-up assessments. Even if a patient completes the study, some data may remain simply unreported or uncollected. In general this document concentrates on how to handle the situation where data are missing due to patients withdrawing from a trial.

Ignoring missing data in the analysis violates the strict ITT principle which requires measurement and analysis of all patient outcomes regardless of protocol adherence. This principle is of critical importance as confirmatory clinical trials should estimate the effect of the experimental intervention in the population of patients with greatest external validity and not the effect in the unrealistic scenario

where all patients receive treatment with full compliance to the treatment schedule and with a complete follow-up as per protocol. Full set analysis generally requires the imputation of values or modelling for the unrecorded data. Even the per protocol analyses might also require the use of some method of handling missing data for patients who have no major protocol violations or deviations but, for some reason, have data not recorded. The manner in which missing data are handled can have, depending upon the amount and type of missing data, a crucial influence on the final results of a clinical trial and on the certainty with which conclusions can be drawn.

Missingness can be defined as both the existence of missing data and the mechanism that explains the reason for the data being missing. The extent to which missing values lead to biased conclusions about the magnitude of any treatment effect is influenced by many factors. Among these are the relationship between missingness, treatment assignment and outcome; the type of measure employed to quantify the treatment effect and the expected changes over time for the variables being measured.

It should be noted that the strategy employed to handle missing values might in itself constitute a source of bias and that there is no universal best approach for all situations. The acceptability of an approach will depend on the assumptions made and whether it is reasonable to make these assumptions in the particular case of interest. It is very important when designing a study that the likely pattern of missing data is taken into account when specifying the primary analysis and the predefined sensitivity analyses. The amount of missing data and the strategies selected to handle missing data can influence the required sample size, the estimate of treatment effect and the confidence with which data can ultimately be interpreted. As such, how to minimise the amount of missing data and how missing data are going to be handled in the analysis are critical issues that must be considered when planning a clinical trial.

This document is not an extensive review of all the available methods, nor can it be a guide to preferred methods in any specific experimental situation. Instead general recommendations on acceptable frameworks for handling missing data in a regulatory setting are outlined. A positive regulatory decision should not be based on an analysis that is biased in favour of the experimental agent. Hence when proposing methods to handle missing data it is important that an analysis is provided for which it is judged that it is unlikely that it is biased in favour of the experimental treatment (i.e. the method can be considered ‘conservative’).

2. Scope

This guideline provides advice on how the presence of missing data in confirmatory clinical trials should be addressed and reported in a dossier submitted for regulatory review and provides an insight into the regulatory standards that will be used to assess confirmatory clinical trials with missing data. The main focus of this guideline is issues associated with the analysis of the primary efficacy endpoint where patients are followed up over time. Typically in such trials, some patients withdraw from the study or are lost to follow-up and hence data for those patients are missing from some point at or after a baseline assessment through to the end of the study. The principles outlined apply to superiority and non-inferiority trials as well as to equivalence trials, though the same biased estimate of the treatment effect may lead to different conclusions for trials with different objectives.

3. Legal basis

The Guideline has to be read in conjunction with Annex I to Directive 2001/83/EC, as amended, and all other relevant EU and ICH guidelines. These include, but are not limited to:

CPMP/ICH/363/96, ICH Topic E9 Step 4 Note for Guidance on Statistical Principles for Clinical Trials.

4. The Effect of Missing Values on Data Analysis and Interpretation

The following problems may affect the interpretation of the trial results when some missing data are present.

4.1. Power and Variability

The sample size and the variability of the outcomes affect the power of a clinical trial. The power of a trial will increase if the sample size is increased or if the variability of the outcomes is reduced.

If missing values are handled by simply excluding any patients with missing values from the analysis, this will result in a reduction in the number of cases available for analysis and therefore normally result in a reduction of the statistical power. Clearly, the greater the number of missing values, the greater the likely reduction in power. Hence every effort should be made to minimize the amount of missing data.

Conversely, non-completers might be more likely to have extreme values (treatment failure leading to dropout, extremely good response leading to loss of follow-up). Therefore, the loss of these non-completers could lead to an underestimate of variability and hence artificially narrow the confidence interval for the treatment effect. If the methods used to handle missing data do not adequately take this into account, the resulting confidence interval cannot be considered a valid summary of the uncertainty of the treatment effect.

4.2. Bias

Bias is the most important concern resulting from missing data. If patients are excluded from the analysis, this may affect:

- The comparability of the treatment groups. A consequence of this may be a bias in the estimation of the treatment effect.
- The representativeness of the study sample in relation to the target population (external validity).

While the reduction of the statistical power is mainly related to the number of missing values, the risk of bias in the estimation of the treatment effect from the observed data depends upon the relationship between missingness and treatment and outcome (see also Section 6.1):

- In principle, missing values will not be expected to lead to bias if they are not related to the real value of the unobserved measurement (e.g. poor outcomes are no more likely to be missing than good outcomes).
- Conversely, if the unmeasured observation is related to the real value of the outcome (e.g. the unobserved measurements have a higher proportion of poor outcomes), this may lead to a biased estimate of the treatment effect even if the missing values are not related to treatment (i.e. missing values are equally likely in all treatment arms).
- Ignoring missing observations will lead to bias if related to both the treatment and the unobserved outcome variable (e.g. missing values are more likely in one treatment arm because it is not as effective).

It is not possible to establish the relationship between missingness and the unobserved outcome variable. Thus it is sensible to adopt a conservative approach, considering missing values as a potential source of bias. It should also be noted that the pattern and extent of missing data can itself provide important information for the interpretation of a trial.
The type of bias that can critically affect interpretation will depend upon whether the objective of the study is to show a difference or demonstrate non-inferiority/equivalence. If the variability of the treatment effect is underestimated, the width of the resulting confidence interval will be too narrow. In a superiority trial this could lead to a false claim of a statistically significant result. Similarly, in a non-inferiority/equivalence trial this could lead to a false claim of non-inferiority/equivalence. In a superiority or non-inferiority trial, an inflated estimate of the treatment effect is biased in favour of the experimental treatment. This too can lead to a false claim of superiority/non-inferiority being made. In an equivalence trial, an inflated estimate of difference between treatments is potentially biased against the experimental treatment. This could lead to equivalence being incorrectly rejected in situations where effects of experimental and reference treatments are the same.

5. General Recommendations

Unfortunately, there is no methodological approach for handling missing values that is universally accepted in all situations. Nevertheless there are some rules which should be considered when handling missing data.

5.1 Avoidance of Missing Data

There is no rule regarding the maximum number of missing values that could be acceptable. The quantity of missing data may be affected by a number of factors:

a) The nature of the outcome variable: the occurrence of missing values is expected to be lower when the outcome variable is mortality (e.g. in cardiovascular trials), than when the outcome is more difficult to assess and requires the active participation of patients and/or sophisticated methods of diagnosis.

b) The length of the clinical trials: the longer the follow up the greater the likelihood of missing values.

c) The therapeutic indication: missing values are more frequent in those diseases where the adherence of patients to the study protocol is usually low (e.g. psychiatric disorders).

d) The treatment modalities: e.g. surgical versus medical treatment.

Several major difficulties arise as a result of the presence of missing values and these are aggravated as the number of missing values increases. Thus, it is extremely important to avoid the presence of unobserved measurements as much as possible, by favouring designs that minimise this problem, as well as strengthening data collection regardless of the patient's adherence to the protocol. Continued collection of data after the patient's cessation of study treatment is strongly encouraged, in particular data on clinical outcome. In some circumstances, in particular where this type of "retrieved dropout" information represents the progression of the patient without (or before) impact of further therapeutic intervention, these data will give the best approximation to the Full Analysis Set and would generally be seen as a sound basis for the primary analysis. This analysis is especially important when the aim of the study is establish the efficacy of the test treatment over the whole study duration. Data should also be collected on other therapies received post dropout. Specifically full details of the type of additional (non-randomised) therapy given, including when and for how long it was used and at what dose, should be collected. This information will allow the value of any outcome data collected after withdrawal to be put into context.

5.2 Design of the study and relevance of predefinition

It is very important when designing the study and specifying the statistical methods to be used, to anticipate the proportion of missing values likely to be observed in the trial. Experience from exploratory trials and from other trials in related indications should inform expectations for missing data when planning the trial. Careful planning will help specify a plausible approach to handling missing data and also help to specify a range of sensitivity analyses to explore the impact of departures from the expected missing data pattern. As part of this process, an upfront investigation of different missing data handling methods under different assumptions, ranging from pessimistic to realistic to optimistic, using clinical scenario evaluations should be carried out. Indeed, an estimate of the predicted (and unavoidable) amount of missing data is highly recommended: firstly because this may have repercussions for the variability and the expectations of the effect size and hence the sample size calculation, secondly because proper planning should minimise the risk that the strategy for missing data handling itself introduces bias, and thirdly because the uncertainty introduced in interpreting the results increases (and hence the number of sensitivity analyses required may need to increase – See Section 7) as the number of missing values increases.

There is no universally applicable method of addressing the fact that data for some patients are missing at specific timepoints, and different approaches may lead to different results. To avoid concerns over data-driven selection of methods, it is essential to pre-specify the selected methods in the statistical section of the study protocol or analysis plan and to pre-specify which method will be used for the primary analysis. This section must include a detailed description of the selected methods and a justification of why the methods to be applied are expected to be an appropriate way of summarising the results of the study and allow assessment without an important degree of bias in favour of experimental treatment. The sensitivity analyses envisaged should also be pre-specified.

It is considered of particular importance to consider whether the selected method is likely to be a conservative approach (under circumstances expected to occur in the study), which is not expected to favour the study's working (alternative) hypothesis (e.g. demonstration of superiority to placebo or demonstration of non-inferiority to active control) either because of a bias in the point estimate or because of an under-estimate of the variability. The process of imputation or modelling might be relevant to not only the main variables, but also to the baseline variables and covariates, and key secondary efficacy and safety variables with a view to including all patients in all important analyses.

5.3 Final Report

A detailed description of the pre-planned methods used for handling missing data, any amendments of that plan and a justification for those amendments should be included in the clinical study report. If the pattern of missing data is different to that envisaged at the design stage, and the planned sensitivity analyses are inadequate, further sensitivity analyses should be provided that are tailored to the missing data pattern observed (see section 7). To properly address the missing data problem it may be insufficient to only follow pre-specified methods, without consideration of the actual data collected.

A critical discussion of the number, timing, pattern, reason for and possible implications of missing values in efficacy and safety assessments should be included in the clinical report as a matter of routine. Graphical summaries (e.g. Kaplan-Meier plots) of the dropout patterns should be provided so that it can be clearly seen if there is a differential dropout pattern between treatment groups. These graphical summaries should identify the recorded reason for dropout.

Data explorations and accompanying explanations that investigate missing data imbalance in all relevant factors and whether patients with and without missing values have different characteristics at baseline can also be informative. Data presentations supporting the efficacy results should be such that it helps to quantify the contribution of each patient to the statistical analysis. For example, if single imputation methods are used (see section 6.3.1) the imputed values must be listed and identified. Elucidation of the missing data pattern is just as important to regulatory assessment as the choice of missing data handling method as it helps to understand the likely direction of any bias in the analyses.

When patients drop out of a trial, full reporting of all reasons for their discontinuation should be given where possible. This should allow identification of the most important reasons that caused them to discontinue and may influence how these subjects are treated in the missing data analysis. Any follow-up information collected post dropout could be helpful in justifying how these patients are handled in the analyses. The final report must include documentation of any deviation from the expected number of missing values, a discussion of whether the pre-defined analysis is still sensible, plus appropriate sensitivity analyses.

As stated before, sensitivity analyses should investigate the robustness of the conclusions of the study, and it is essential that an analysis can be identified which is assessed not to be biased to an important degree in favour of experimental treatment. Also the confidence interval for this analysis should appropriately reflect the uncertainty associated with the estimated treatment effect.

Because of the unpredictability of some problems, it may be acceptable to allow in the study protocol the possibility of updating the strategy for dealing with missing values in the statistical analysis plan at a later point in time (e.g. during the blind review of the data at the end of the trial). Relevant deviations from and amendments of the pre-specified plan should be clearly documented and justified. In addition, the time at which these deviations and amendments were decided and implemented in relation to the blinding of the data must be clearly identified. Methods for the documentation of these changes can be found in ICH E9. If unexpected missing data patterns are found in the data, it will be necessary to conduct some *post hoc* sensitivity analyses in addition to those predefined in the statistical analysis plan (see section 7). In this case the reasons why these analyses have been conducted should be carefully explained and thoroughly justified, and the initially planned analysis should also be presented. Proper planning will minimise the number of such *post hoc* analyses, avoiding concerns over data-driven selection of methods. As with any retrospective analysis, implementation will be closely scrutinised. It is not envisaged that these *post hoc* analyses can be used to rescue a trial which otherwise fails. They should only be used to investigate whether positive results remain so when the unexpected missing data pattern is appropriately reflected in the sensitivity analyses.

6. Handling of Missing Data

6.1 Theoretical Framework

A framework in the literature for the applicability of the different methods to handle missingness is based on a classification according to the following missingness mechanisms:

- For the dependent variable (conditional on the covariates in the model), if the probability of an observation being missing does not depend on observed or unobserved measurements then the observation is Missing Completely At Random (MCAR). A typical example is a patient moving to another city for non-health reasons. Patients who drop out of a study for this reason could be considered a random and representative sample from the total study population.

- Conditional on the covariates in the model, if the probability of an observation being missing depends only on observed measurements then the observation is Missing At Random (MAR). This assumption implies that the behaviour of the post dropout observations can be predicted from the observed variables, and therefore that response can be estimated without bias using exclusively the observed data. For example, when a patient drops out due to lack of efficacy reflected by a series of poor efficacy outcomes that have been observed, it would be appropriate to impute or model poor efficacy outcomes subsequently for this patient.
- When observations are neither MCAR nor MAR, they are classified as Missing Not At Random (MNAR), i.e. the probability of an observation being missing depends on unobserved measurements. In this scenario, the value of the unobserved responses depends on information not available for the analysis (i.e. not the values observed previously on the analysis variable or the covariates being used), and thus, future observations cannot be predicted without bias by the model. For example, it may happen that after a series of visits with good outcome, a patient drops out due to lack of efficacy. In this situation the analysis model based on the observed data, including relevant covariates, is likely to continue to predict a good outcome, but it is usually unreasonable to expect the patient to continue to derive benefit from treatment.

As already stated in section 4.2., it is impossible to be certain whether there is a relationship between missing values and the unobserved outcome variable, or to judge whether missing data can be adequately predicted from the observed data. It is not possible to know whether the MAR, never mind MCAR, assumptions are appropriate in any practical situation. A proposition that none of the data missing in a confirmatory clinical trial are MNAR seems implausible.

Because it cannot be excluded that some data may be MNAR, the properties (e.g. bias) of any methods based on MCAR or MAR assumptions cannot be reliably determined for any given dataset. Therefore the justifications for the methods chosen should not depend primarily on the properties of the methods under the MAR or MCAR assumptions but on whether it is considered to provide an estimate that is acceptable for regulatory decision making in the circumstances of the trial under consideration.

6.2 Complete Case Analysis

Complete case analysis cannot be recommended as the primary analysis in a confirmatory trial. In this approach incomplete data are ignored and the statistical analysis is performed using only cases with complete data. Some problems associated with this approach are discussed in section 4. In general, complete case analysis violates the intention to treat principle and is subject to bias. However, the approach may be considered in other circumstances, e.g.

- In exploratory studies, especially in the initial phases of drug development.
- In confirmatory trials as a secondary supportive analysis (sensitivity analysis) to illustrate the robustness of conclusions.

6.3 Methods of Handling Missing Data

Factors that affect the acceptability of individual methods include:

- differences between the treatment groups in the proportion of patient withdrawals,
- differences between the treatment groups in the timing of withdrawals,
- the reason for the patient withdrawals,
- the direction of any spontaneous changes over time.

All of these factors must be comprehensively displayed and their influence discussed when the methods used to handle missing data are justified. This highlights the importance of proper planning (see section 5.2) and proper evaluation of unanticipated occurrences (see section 5.3).

The acceptability of any method will be judged by the same standards regardless of its complexity or its assumptions. As described in section 6.3.1 and 6.3.2, a primary analysis can only be accepted if is considered that an important bias in favour of the experimental treatment can be reasonably excluded and if it can be verified that the associated confidence interval does not underestimate the variability of the estimated treatment effect to an important extent. In the event that the proportion of missing data is non-negligible, it is likely that no single method will provide a comprehensive solution to the missing data problem.

6.3.1 Imputation Methods

Single imputation methods replace a missing data point by a single value and analyses are conducted as if all the data were observed.

To cope with situations where data collection is interrupted before the predetermined last evaluation timepoint, one widely used single imputation method is Last Observation Carried Forward (LOCF). This analysis imputes the last measured value of the endpoint to all subsequent, scheduled, but missing, evaluations.

Only under certain restrictive assumptions does LOCF produce an unbiased estimate of the treatment effect. Moreover, in some situations, LOCF does not produce conservative estimates. However, this approach can still provide a conservative estimate of the treatment effect in some circumstances.

To give some particular examples, if the patient's condition is expected to deteriorate over time (for example in Alzheimer's disease) an LOCF analysis is very likely to give overly optimistic results for both treatment groups, and if the withdrawals on the active group are earlier (e.g. because of adverse events) the treatment comparison will clearly provide an inappropriate estimate of the treatment effect and may be biased in favour of the test product. Hence in this situation an LOCF analysis is not considered appropriate. Indeed in Alzheimer's disease, and other indications for diseases that deteriorate over time, finding a method that gives an appropriate estimate of the treatment effect will usually be difficult and multiple sensitivity analyses will frequently be required.

However, in other clinical situations (e.g. depression), where the condition is expected to improve spontaneously over time, LOCF (even though it has some sub-optimal statistical properties) might be conservative in the situations where patients in the experimental group tend to withdraw earlier and more frequently. Establishing a treatment effect based on a primary analysis which is clearly conservative represents compelling evidence of efficacy from a regulatory perspective.

Baseline Observation Carried Forward (BOCF) is another single imputation approach that is sometimes used. The use of BOCF may be appropriate in, for example, a chronic pain trial where when a patient withdraws from treatment it may be reasonable to assume that pain return to its baseline level and that the patient does not, in the long-term, derive benefit from treatment.

Other simple approaches for single imputation of missing data are to replace the unobserved measurements by values derived from other sources. Possible sources include information from the same subject collected before withdrawal, from other subjects with similar baseline characteristics, a predicted value from an empirically developed model or historical data. Examples of empirically developed models are unconditional and conditional mean imputation, best or worst case imputation (assigning the worst possible value of the outcome to dropouts for a negative reason (treatment failure) and the best possible value to positive dropouts (cures)), regression methods and hot-deck imputation. These approaches may sometimes be useful as sensitivity analyses.

An attractive approach for imputing missing data may be to employ a different pre-specified imputation technique for each different reason for withdrawal, rather than the same technique for all patients. While this would represent a relatively novel approach, there is no objection to this in principle. The strategy has more flexibility in handling different reasons for and timings of withdrawal and consequently the possible relationship between missing data and the outcome of interest. If used appropriately, it may better address the question of primary regulatory interest. The method also offers an intuitive framework for conducting a range of sensitivity analyses.

A potential disadvantage of single imputation methods is that these methods risk biasing the standard error downwards by ignoring the uncertainty of imputed values. Therefore, the confidence intervals for the treatment effect calculated using single imputation methods may be too narrow and give an artificial impression of precision that does not really exist. This possibility should be addressed when results from these analyses are presented.

The risk of underestimating the variance of treatment effect when imputing can be reduced by proper implementation of techniques such as multiple imputation. Multiple imputation methods generate multiple copies of the original dataset by replacing missing values using an appropriate stochastic model, analyse them as complete sets and finally combine the different parameter estimates across the datasets to produce a unique point estimate and standard error taking into account the uncertainty of the imputation process. The seed of the pseudorandom number generator used to randomly generate imputations for the missing values should be pre-specified in the protocol to avoid bias being introduced by *post hoc* selection of the seed and therefore the random numbers to be imputed.

6.3.2 Mixed models and generalised estimating equations

Some statistical approaches to handling missing data do not employ explicit imputation. For continuous responses, linear mixed models are sometimes used to handle missingness when a series of outcomes are measured repeatedly over time (mixed-effect models for repeated measures (MMRM)). For categorical responses and count data, the so-called marginal (e.g. generalized estimating equations (GEE)) and random-effects (e.g. generalized linear mixed models (GLMM)) approaches are in use. Likelihood-based methods (MMRM and GLMM) and some extended GEE (i.e. weighted GEE) models are applicable under MCAR and MAR assumptions.

In many cases, there is a variety of different settings for each method which could lead to different conclusions (e.g. type of variance-covariance matrix for MMRM and assumptions to model the unobserved measurements used in the MNAR methods, etc.). Therefore, the precise option settings must be fully justified and predefined in advance in detail, so that the results could be replicated, if required, by an external data analyst and so that it can be established that the choice has not been made *post hoc*.

Under the MAR assumption, the methods above provide an unbiased estimate of the treatment effect that would have been observed if all patients had continued on treatment for the full study duration. However, since it is likely that some data are MNAR, these methods will, in certain circumstances, overestimate the size of the treatment effect likely to be seen in practice, and hence introduce bias in favour of the experimental treatment. In light of this, the point estimates obtained can be similar to those from a complete cases analysis, and therefore may introduce important bias. This is problematic in the context of a regulatory submission as confirmatory clinical trials should estimate the effect of the experimental intervention in the population of patients with greatest external validity, and not in the unrealistic scenario where all patients receive treatment with full compliance to the treatment schedule and with a complete follow-up as per protocol. It is not acceptable to make an assumption that data are MAR and therefore to present such models as a comprehensive solution to the missing data problem.

Generally, MNAR data is difficult to rule out, and it is not clear whether even a small amount of MNAR data could have an impact on the study results in a particular experiment. Therefore, approaches that investigate different MNAR scenarios such as a pattern mixture model (PMM), a selection model (SEM) and a shared parameter model (SPM) may be useful. A combined strategy incorporating several methods for handling missingness (e.g. assume dropouts due to lack of efficacy and adverse events are MNAR and lost to follow-up are MAR) may also be considered. As described above, methods that do not assume MCAR or MAR such as PMMs may offer a flexible framework to explore the impact of treating different types of missing data as MNAR and evaluating the impact different modelling strategies have on the estimated treatment effect.

6.4 Responder Analysis

In some circumstances, the primary analysis of a continuous variable is supported by a responder analysis. In other circumstances, the responder analysis is designated as primary. How missing data are going to be categorised in such analyses should be pre-specified and justified. If a patient prematurely withdraws from the study it would be normal to consider this patient as a treatment failure. However, the best way of categorisation will depend on the trial objective (e.g. superiority compared to non-inferiority).

In a situation where responder analysis is not foreseen as the primary analysis, but where the proportion of missing data may be so substantial that no imputation or modelling strategies can be considered reliable, a responder analysis (patients with missing data due to patient withdrawal treated as failures) may represent the most meaningful way of investigating whether there is sufficient evidence of the existence of a treatment effect.

6.5 Survival Analysis

When the outcome measure is time to event, survival models which take into account censored observations are often used. Specialist considerations are outside the scope of the current document. However, some standard survival methods assume that there is no relationship between the response and the missing outcome. Violations from this assumption could lead to biased results especially when data are missing due to withdrawal (rather than data for an interim visit being missing). Therefore whether it is reasonable to assume non-informative censoring should be discussed in the protocol or the study report. This is particularly of importance in situations where the amount of missing

data/patient withdrawals could influence whether the treatment effect is established or could influence the size of the treatment effect. A range of sensitivity analyses should be provided. Further considerations for handling this type of data are outlined in Appendix 1 to the CHMP guideline on the evaluation of anticancer medicinal products in man (EMEA/CHMP/EWP/27994/2008).

6.6 Count Data

For count data (e.g. the number of exacerbations) a weighted approach with time-in-study as an offset variable (e.g. Poisson regression) is sometimes used. Although this approach is intuitively appealing it should be noted that it assumes there is no relationship between the response and the missing outcome i.e., the method assumes that event rate after withdrawal from trial is the same as event rate on study treatment. For this reason it will often not be an appropriate primary analysis of count data in the presence of missing data. It is recommended that data are collected after withdrawal from study treatment as outlined in section 5.1 and a primary analysis that takes into account post withdrawal information is adopted. Baseline count imputation could also be a plausible approach if information after withdrawal from study treatment is not available.

7. Sensitivity Analyses

In this context, sensitivity analyses can be defined as a set of analyses where the missing data are handled in a different way as compared to the primary analysis. It should be noted that obtaining similar results from a range of methods that make similar or the same assumptions does not constitute an adequate set of sensitivity analyses. The sensitivity analyses should show how different assumptions influence the results obtained.

In submissions with non-negligible amounts of missing data, sensitivity analyses should be presented as support to the main analysis. Because the performance of any analysis presented (in terms of bias and precision) cannot be fully elucidated, presentation of trial results without adequate investigation of the assumptions made for handling missing data cannot be considered comprehensive.

When the results of the sensitivity analyses are consistent with the primary analysis and lead to reasonably similar estimates of the treatment effect, this provides some assurance that neither the lost information nor the methods used to handle missing data had an important effect on the overall study conclusions. In this situation the robustness of the results is clear and the missing values will not generally be considered to be a serious source of concern. The consistency of results from a broader range of sensitivity analyses will give greater reassurance on the robustness of the trial results. Conversely, for a study with a statistically significant primary analysis, whilst not all sensitivity analyses must necessarily give statistically significant results, if they produce inconsistent results (e.g. a markedly smaller estimate of treatment effect), their repercussions on the conclusions of the trial must be discussed. In certain circumstances, the influence of missing data is so great that it might not be possible to reliably interpret the results from the trial.

The sensitivity analyses required will need to be defined on a case-by-case basis, though they will usually comprise the analyses already described in Section 6 above.

Some ways of performing a sensitivity analysis are:

- Compare the results of the full set analysis to those of the complete case analysis.
- As discussed in section 6.2, it is not possible to guarantee that at least some of the missing data are not MNAR. Therefore, when missing data are particularly problematic, further sensitivity analyses that treat certain types of missing data as MNAR should be provided. It may be appropriate to treat data missing for different reasons in different ways. A range of analyses should be provided that explore these possibilities. For each of these analyses, a clear explanation of what values have been imputed should be given. This may be done using multiple imputation methods incorporating pattern mixture approaches.
- Compare the impact different model settings have on the results. If different results are obtained from models using the same missing mechanism assumption (e.g. multiple imputation versus MMRM both assuming MAR) full details of the differences between these models that explain the different results obtained should be provided. In any case the impact different settings of a model have on the results obtained should be explained in detail.
- Utilise 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 endpoint 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.

- A primary analysis of a continuous variable could be supported by a responder analysis that treats all missing values as failures, or treats only missing values due to a certain reason (e.g. due to adverse events) as failures.
- A worst case analysis: assigning the best possible outcome to missing values in the control group and the worst possible outcome to those of the experimental group. If this extreme analysis is still favourable then it can be confidently concluded that the results are robust to the handling of missing data.

Each sensitivity analysis should be designed to assess the effect on the results of the particular assumptions made to account for the missing data. The sensitivity analysis should be planned and described in the protocol and/or in the statistical analysis plan and any changes must be documented and justified in the study report.

References

Appendix 1 to the CHMP guideline on the evaluation of anticancer medicinal products in man (EMEA/CHMP/EWP/27994/2008).

ICH E9 Statistical Principles of Clinical Trials (CPMP/ICH/363/96).