Missing Data: Discussion Points from the PSI

Missing Data Expert Group

Alan Phillips, June 2010

- **Background**
  - PSI Discussion Group

- **Discussion Topics**
  - Minimising the extent of missing data and understanding the pattern of missing data
  - Defining the principles for handling missing data
  - Understanding the assumptions underlying different analysis methods

- **Key Messages**

- Adopted by the Committee of Health and Medicinal Products (CHMP) in December 2001.

- September 2007 the CHMP issued a recommendation to review the document, with particular emphasis on the following.
  - Summarising and critically appraising the pattern of drop-outs.
  - Use of sensitivity analysis or the justification for their absence.
  - Explaining the role and limitations of the "Last Observation Carried Forward" (LOCF) method.
  - Describing the CHMP's cautionary stance on the use of mixed models.

Credits

- Tomasz Burzykowski   MSOURCE Medical Development
- James Carpenter   London School of Hygiene and Tropical Medicine
- Corneel Coens   EORTC
- Daniel Evans   Pfizer
- Lesley France   AstraZeneca
- Mike Kenward   London School of Hygiene and Tropical Medicine
- Peter Lane   GlaxoSmithKline
- James Matcham   Amgen
- David Morgan   Ipsen
- Alan Phillips   ICON Clinical Research
- James Roger   GlaxoSmithKline
- Brian Sullivan   Statistical Solutions
- Ian White   MRC Biostatistics Unit
- Ly-Mee Yu   Centre for Statistics in Medicine, University of Oxford

Clinical Research

- Minimising the extent of missing data and understanding the pattern of missing data.

- Defining the principles for handling missing data.

- Understanding the assumptions underlying different analysis methods.

- Proactively plan for missing data

- Unambiguously state
  – The objectives of the study
  – The patient population of interest
  – How missing data may impact the inference to be made

- Protocols/SAPs rarely discuss the expected pattern of missing data or consider the impact of the potential patterns on the overall scientific validity of the trial.

- "A sample size of 64 in each group will have 80% power to detect a difference in means of 4.000 assuming that the common standard deviation is 8.000 using a two group t-test with a 0.050 two-sided significance level."

- A total of 71 patients will be recruited, assuming a 10% drop out rate

- **Consider a two-step withdrawal process for patients**
  - Withdrawal of consent for treatment
  - Withdrawal of consent from observation

  - May be a challenge to explain value of continuing to observe patients when not being treated for some diseases (eg pain control).
  - Practice is already standard in others (eg Oncology)

  - Reduce amount of data after withdrawal to improve compliance

- **Tighter control of patient population**
  - Select patients who are more likely to complete study
  - Approach reduces the generalizability of the trial

- Understanding of patient withdrawal patterns starts with the collection of relevant information.

- Standard withdrawal or discontinuation Case Report Forms (CRFs) are often employed
  - Prescribed standard lists for reasons for withdrawal
  - Not enough thought is given to the customisation of these CRFs for the disease under consideration or the study objectives
  - How often are disease- or study-specific reasons included?

- Identify potential predictors of missing data, both to facilitate the collection of relevant data, and for potential inclusion in the analysis.
    - eg Asthma clinical trial, FEV1 is often used as the primary endpoint.
    - "Asthma exacerbations" widely recognised as an important endpoint.
    - When such events occur a patient may visit their health care professional, who in turn may advise the patient to withdraw from the trial.
    - Subsequently it may be important to collect data on "asthma exacerbations".

- Drug development spans many years and comprises an ordered program of clinical trials.

- Little effort is made to understand missing data in the earlier phases of drug development.

- Tend to start considering "missing data" during late Phase II and Phase III
    - When such issues can affect the "approval" of the final package.

- Missing data mechanisms need to be considered when making go/no-go decisions at the end of Phase I (in selected therapeutic areas; oncology) and early Phase II and how this may impact later phase clinical study design.

Q3. What are the relative merits of the 1) *Plotting raw data and inspection of the data? 2) Analysis by pattern of missing data (drop-out cohort)? 3) Logistic regression of drop-out on earlier data?*

Clinical Research

- Graphical display is one of the most important tools available when trying to understand the causes of missing data.

- Although analytical methods exist for exploring missing data, a large amount of information can be ascertained by simply plotting the data:
    - Kaplan-Meier plots to look at time to withdrawal
    - Plots of treatment means against time for cohorts of subjects with similar follow up times

- The key to success is thinking through the question of interest and intelligently plotting the data.

- The principles for minimising and understanding missing data should not change for safety data

- Challenges may be very different.

- Careful consideration needs to be given as to how information will be collected about specific events of interest in Phase III, and the impact of missing data on the inferences to be drawn during the design phase.

- Increased use of graphical displays and more in-depth analyses are required for safety data.
  - Interpretation should be linked to the Risk Management Plan.

Defining the principles for handling missing data.

Q5. Regulators have stated on numerous occasions that missing data from patients who drop out are different from other types of missing data. What are the principles for handling different types of missing data?

Clinical Research

- It is important to understand the mechanism causing the missing data.

- The proposed method of analysis, and associated handling of missing data, regardless of whether the patient discontinued or not, must be directly linked and properly reflect the objectives of the study and trial design assumptions.

- Specifically for patients who withdraw
  - "what information needs to be collected for patients who discontinue"

- The cost of running additional trials to investigate the effect of missing data far outweighs the cost of collecting the appropriate information in the first instance.

- Two important principles exist when considering sensitivity analyses: transparency and relevance of the assumptions.

- Clearly describe the original assumptions when designing the study so that all stakeholders can assess their relevance.

- The assumptions underlying any sensitivity analyses should be divergent from the original assumptions.

- A series of "wrong" analyses does not properly constitute a sensitivity analysis.

Q7. Regulators seem to be favouring a requirement for sponsor companies to monitor patients after withdrawal. How should post-withdrawal data be handled in the statistical analysis?

Clinical Research

- Collecting data after withdrawal seems to be a critical one from the regulatory perspective.

- The issue reinforces the need to clearly define the objectives of the study.

- When clearly and precisely defining the objectives it will become apparent whether collecting data from patients who withdraw is necessary to address the question of concern.

Understanding the assumptions underlying different analysis methods

Q8. What are the underlying assumptions of the a) Last Observation Carried Forward (LOCF), b) Mixed Model for Repeated Measures (MMRM) and c) Multiple Imputation (MI) methods for handling missing data in a longitudinal clinical trial with dropouts or withdrawals?

Clinical Research

- **LOCF**
  - Single-imputation method
  - Makes an implicit assumption that the patients would sustain the same response seen at an early study visit for the entire duration of the trial.
  - Assumption is untestable and potentially unrealistic.
  - Uncertainty of imputation is not taken into account
    - Method results in systematic underestimation of the standard errors.

Clinical Research

- MMRM and MI both make the assumption that data are "Missing At Random".

- MMRM
  - Information from the observed data is used via the within-patient correlation structure to provide information about the unobserved data
  - The missing data are not explicitly imputed.
  - Uses all the available data to provide the information about the unobserved data.
  - Estimates the treatment effects assuming the withdrawn patients mimic those who continued.

- MI
  - Imputation step is separate from the modelling step
    - Additional flexibility to explore different assumptions about the nature of the missing data.
  - If flexibility is not used then it may in some circumstances essentially mimic an MMRM, and so offer no advantages over that method.

- If the underlying mechanisms that cause missing data are non-informative the resulting impact on the statistical analysis is far easier to handle, compared to informative missingness.

- The data being analysed, however, cannot provide evidence to distinguish between these two situations.

- Expert group tabulated scenarios when it might be appropriate/inappropriate to use LOCF, MMRM or MI techniques
  - Longitudinal clinical trial with dropouts or withdrawals.

- MMRM and MI in their most basic form, both assume the multivariate normal distribution when providing information about the missing data.
  - Invalid inferences can be drawn when the assumption is not met.
- Generalizations and modifications of these approaches exist which are valid under other distributional assumptions.

- Key issues
  - True missing data mechanism will always be unknown and not testable from the data.
  - No amount of clever modelling can overcome this.
  - If the mechanism for missingness is informative then it will not be possible to fully evaluate the impact of the treatment of missing data in the analysis

- Biostatisticians tend to react to missing data.

- Comprehensive, proactive planning is rarely undertaken when designing trials.
  - Precise objectives of the trial need to be documented
  - Potential impact of missing data thoroughly considered during the planning phase.
  - Missing data mechanisms for a trial need to be considered.
  - Sensitivity analyses investigating the robustness of the inferences to the different assumptions made should be considered.

- Biostatisticians need to better understanding of the pattern of missing data observed
  - Plotting of data; for example, use of Kaplan-Meier curves of time to withdrawal.

- Handling of missing data is a difficult area.
  - If the mechanism for the missing data is non-informative then the issue can be addressed by using relatively straightforward statistical techniques.
  - If the mechanism for the missing data is informative then the issues are complex, and appropriate sensitivity analysis is called for.