The Prevention and Treatment of Missing Data in Clinical Trials

(A National Research Council Report)

By Jeng Mah, Ph.D., Senior Principal Biostatistician

On July 19, 2010, an advance copy of a report titled “The Prevention and Treatment of Missing Data in Clinical Trials,” prepared by the Panel on Handling Missing Data in Clinical Trials, the Committee on National Statistics, and the National Research Council, was released for public viewing. An Invited Papers Section under the same title was organized by Study Director Michael Cohen of the Committee on National Statistics at the Joint Statistical Meetings on August 4, 2010. Presenters included panel chair Professor Roderick Little of the University of Michigan, panel member Professor James Neaton of the University of Minnesota, and Thomas Permutt (standing in for Robert O’Neill) of the FDA/CDER.

The Panel on Handling of Missing Data in Clinical Trials was commissioned, at the request of the FDA, to prepare “a report with recommendations that would be useful for FDA’s development of guidance for clinical trials on appropriate study designs and follow-up methods to reduce missing data and appropriate statistical methods to address missing data for analysis of results.”

With the panel’s very impressive roster and its ability to enlist knowledgeable experts to participate in the panel workshops, the result is a comprehensive report that offers:

- Definitions of general terminologies used in clinical trials, not limited to only statistical definitions.
- Trial designs that are less susceptible to missing data.

Missing Data cont. on page two...

Designing Case Report Forms: Part I - Considering Data Management Needs

By Kelsey Farley, Data Management Associate

This article is part I in a two-part series on considering data in CRF design. See part II on page three.

When designing Case Report Forms (CRFs), it is important to keep in mind the needs of Data Management. This is true for both traditional paper based and electronic data capture (EDC) studies. The Data Management team is primarily responsible for the collection and review of CRF data. The team’s ability to efficiently collect and process data relies on the consideration of certain CRF design elements, which all serve to aid in the enterability of data. These elements include, but are not limited to, consistency in formatting, uniformity in variables, and proper “go-to” instructions.

Consistency in Formatting

Using consistent CRF page formats can help facilitate form identification and accuracy in completion. It is specifically important to ensure the headers, footers, version dates, sponsor name, and subject identifiers are the same across CRFs. This will make it easy for the individuals completing the CRF, or entering the CRF data, to identify the correct form to complete or enter.

Part I: CRFs and Data Management continued on page two...
Missing Data cont. from page one...

- Methods to minimize the frequency of missing data.
- Descriptions of mechanisms for missing data and explanations on why missing completely at random (MCAR) is nearly never a justifiable assumption.
- Descriptions and evaluations of various missing data imputation methods.
- Principles and methods for sensitivity analysis and final decision making.
- Recommendations on trial conduct and analysis steps.
- A complete and up-to-date list of literature on this topic.

Even though missing data is typically viewed as a data analysis issue that is magically handled by statisticians, one impressive quality of this report (written by a panel of acclaimed statisticians) is that it does not treat missing data as a statistical problem to be solved through statistical means. The report faithfully adheres to these principal assumptions:

- There is no satisfactory solution to missing data.
- The ideal method in handling missing data is to prevent data from going missing. This should be an important objective in clinical trial design.
- Data, if not collected, can never be recovered. The goal for missing data imputations is to minimize the impact of missing data on our ability to use the existing data in drawing inferences and making decisions.
- All of missing data handling techniques ultimately rely on untestable assumptions.
- Sensitivity analysis is critical in evaluating the robustness of the trial conclusion.
- Methods for handling missing data should be pre-specified in the protocol.

The panel’s report does not cover the decision analysis approach of handling missing data, which measures the costs/losses of missing data and the efforts used in fixing it.

We at RCRI are looking forward to seeing discussions on this report and other opinions on missing data handling triggered by this report and the release of the promised FDA guidance. In the meantime, this report is a useful guidance/reference for anyone interested in the information collected in clinical trials. A free prepublication copy can be downloaded here from The National Academies Press.

For information on how RCRI can help with your study’s data needs, please contact Dr. Zengri Wang at zwang@rcri-inc.com or RCRI Vice President of Business Development Juli Denny at jddeny@rcri-inc.com

Part I: CRFs and Data Management continued from page one...

Consistent presentation of CRF questions aids in the collection of accurate information, and increases the efficiency of data entry. This is especially important when the same questions are being asked at different visits. When that is the case, it is crucial that the question text remain unchanged. For example, if an investigator is asked, “Has the subject’s health improved since last visit” at one visit, the question “Has the subject’s health deteriorated since last visit” should not be asked at the subsequent visit as this will likely lead to data collection errors. Consistently presenting CRF questions and page formats also allows the individuals completing and entering the CRFs to become proficient at locating specific items on individual pages.

Requiring responses to all questions will also lead to fewer data collection errors. For example, the question, “List any abnormalities present” may result in a blank field if no abnormalities occurred. However, it would remain unclear if, in fact, there were no abnormalities or if the field was simply left blank by accident. A better choice would be to ask “Were there any abnormalities? If yes, please specify”. This will eliminate any confusion when completing the form and will result in fewer data queries.

It is also favorable to use a primarily closed system with limited reliance on open-ended questions. A closed system consists mainly of multiple-choice questions, which aid in the overall speed and accuracy of form completion and entry. This is especially helpful for EDC studies where sites are entering data. Compounding multiple questions together also should be avoided, as this will lead to confusion and can affect the validity of responses.

Uniformity in Variables

It is important to present variables uniformly across CRFs (e.g. date, time, question number, and code formats). Ensuring uniformity ultimately results in more accurate CRF data and less time spent on review. For example, if a date is presented in dd/mmm/yyyy

Part I continued on page three...
Part I: CRFs and Data Management continued from page two...

format on one CRF and in mm/dd/yyyy format on another, there is a greater potential that the dates will be recorded incorrectly. A similar scenario also applies for multiple choice or coded variables. If one question contains the responses, 0 = No and 1 = Yes, and another contains 1 = Yes and 2 = No, there is a greater chance that the values recorded in these fields will be improperly entered.

Proper “Go-To” Instructions

Proper “go-to” instructions are helpful in determining the appropriate questions and CRFs to complete. These instructions might include specific study protocol reminders, or an indication of what questions are to be completed next. For example, a proper “go-to” instruction for the question, “Has the subject experienced any adverse events since last visit” might be, “If Yes, complete CRF 5, Adverse Event.” Not only are these instructions helpful to the individual completing the form, but they also result in less time spent reviewing the CRF data and ultimately fewer queries to the sites.

The design of a CRF can greatly affect the data management processes. It is important that certain elements are addressed during the design process to maximize the overall efficiency of data collection, entry, and review. Interaction with the data management team will help to ensure that all the necessary elements are considered.

See the second article in this two-part series on data and CRF design below.

Designing Case Report Forms: Part II - Considering Database Development Needs

This article is part II in a two-part series on data and CRF design. The first article in the series begins on page one.

When designing Case Report Forms (CRFs), it is important to keep in mind Database Development needs. Database developers tend to view the data differently than those who deal with the patient data directly. When reviewing a CRF, developers focus on the stored data (i.e. electronic records) and how the CRF interacts with those records. This review ensures that the CRF contains the elements that facilitate the accurate and retrievable storage of the data. Developers will review, among other things, the interaction with other data, sufficient key field definitions, data entry screen design, and the degree of automation required (i.e. edit checks).

To put it simply, developers view data as a set of inputs and outputs. Inputs are defined as any information which will be or is stored in the database (e.g. CRF data). Outputs are defined as the information produced from the inputs (e.g. an Adverse Event report). Developer recommendations primarily focus on the generation of accurate and efficient output and often ask the question, “will CRF data be stored in a manner that can be easily retrieved and reported/analyzed?”.

Interaction with Other Data

Database developers also need to understand the interaction between patient data reflected on the CRF and other data such as core lab, randomization assignments, device accountability, and visit intervals. The complexity of these interactions can vary significantly depending on the study protocol requirements and database capabilities. For example, interactions can be as simple as linking a single, site-completed CRF and electronic core lab data. Conversely, implementing a visit tracking system may require the linkage of multiple CRFs and study protocol specific information. Creating a system that enables these features requires an understanding of these needs. For example, to facilitate blinded randomization data, a database developer may recommend a separate randomization CRF in order to better utilize database security.

Key Fields

Key fields are also a fundamental requirement for any effective data storage solution. Specifically, a key field allows for differentiation of unique patient records. For example, how do you set apart two follow-up records? For CRF data, the site and subject identifiers are critical components and must be on every subject record. For the example presented above, an additional identifier such as the visit date may be needed. The more complex the relationships between records, the more complex the key fields. Key fields are the most critical elements reviewed by database developers.

Part II: CRFs and Database Development cont. on page four...
Part II: CRFs and Database Development cont. from page three...

**Data Entry and Edit Checks**

A good data entry screen layout can facilitate accurate data entry. Depending on the type of study, electronic data capture (EDC) or paper, different functions and layouts may apply. The primary purpose of the review is to ensure that the layout best handles the required inputs in a manner consistent with requested outputs. For example, requesting race as ‘check all that apply’ instead of ‘check only one’ allows for reporting in a manner consistent with the regulations. Edit checks can automate the data review process by identifying data inconsistencies. This increases data accuracy by removing part of the human element from data review. Edit check definition relies on CRF layout and data formatting. If questions are asked correctly, developers can simplify the writing of these automated checks and produce fewer and more pertinent data queries. If questions are poorly designed, edit checks can be cumbersome to write and use.

**EDC**

Database developers will review CRFs differently for EDC and paper trials. While the data structure may be the same, methods that work for paper CRFs do not work for EDC. It is critical to understand for which audience the CRF is intended (e.g. site coordinator, physicians, or patients) and how the CRFs are to be used. While important in both kinds of studies, data collection requirements for site-managed EDC studies are fundamentally different than the needs of an external data entry staff.

Every database technology has different features. It is important to understand the database capabilities and limitations to ensure that features desired by the project team can be implemented. End users may not know the database capabilities and may not ask for the features which could prove useful. Interaction with the database developers will provide a bridge between study needs and database capabilities. The database developers will define the needed requirements and, with enough time, can adapt the database to do nearly anything requested by the project team.

For information on how RCRI can help with your study’s database design needs, please contact Kelsey or Ryan at kfarley@rcri-inc.com or rboulduan@rcri-inc.com or RCRI Vice President of Business Development Juli Denny at jdenny@rcri-inc.com.