SciAnnews
TOPICS IN CLINICAL RESEARCH

Techniques for Designing Case Report Forms in Clinical Trials

Considerations for Efficient Data Management and Statistical Analysis

Kyung-hee Kelly Moon

1. Abstract

A case report form (CRF) is a data collection tool used in clinical trials to support investigators and coordinators in capturing all protocol-required information. A well-designed CRF facilitates data collection and entry, and directly benefits other facets of data management and statistical analysis. An informative and structured CRF simplifies database design and data validation processes as well as manipulation of data during statistical analysis. This paper explores CRF design techniques that consider efficient data management and statistical analysis.

2. Introduction

A case report form (CRF) is a printed or electronic document used in clinical trials to collect and record protocol required information.[i] The data from the CRFs is compiled as a dataset into a database and validated prior to statistical analysis. Some CRFs are designed to maximize the number of fields per page in order to provide site personnel convenience in completing and retrieving CRFs. This reduces printing costs by reducing the number of pages in a CRF booklet. However, such a design can lead to increased costs in other the areas such as data query management and statistical analysis of clinical trials. If more than 2 different sets of protocol-required information are grouped onto one CRF page, the complexity of database design increases. By positioning too many fields on the same page, the CRF becomes cluttered, potentially making data fields illegible and difficult for data entry. This in turn can lead to an increased frequency in data queries. Moreover, due to the complexity of the database design, statistical programming involves further data preparation and data manipulation steps before the data can be analyzed.

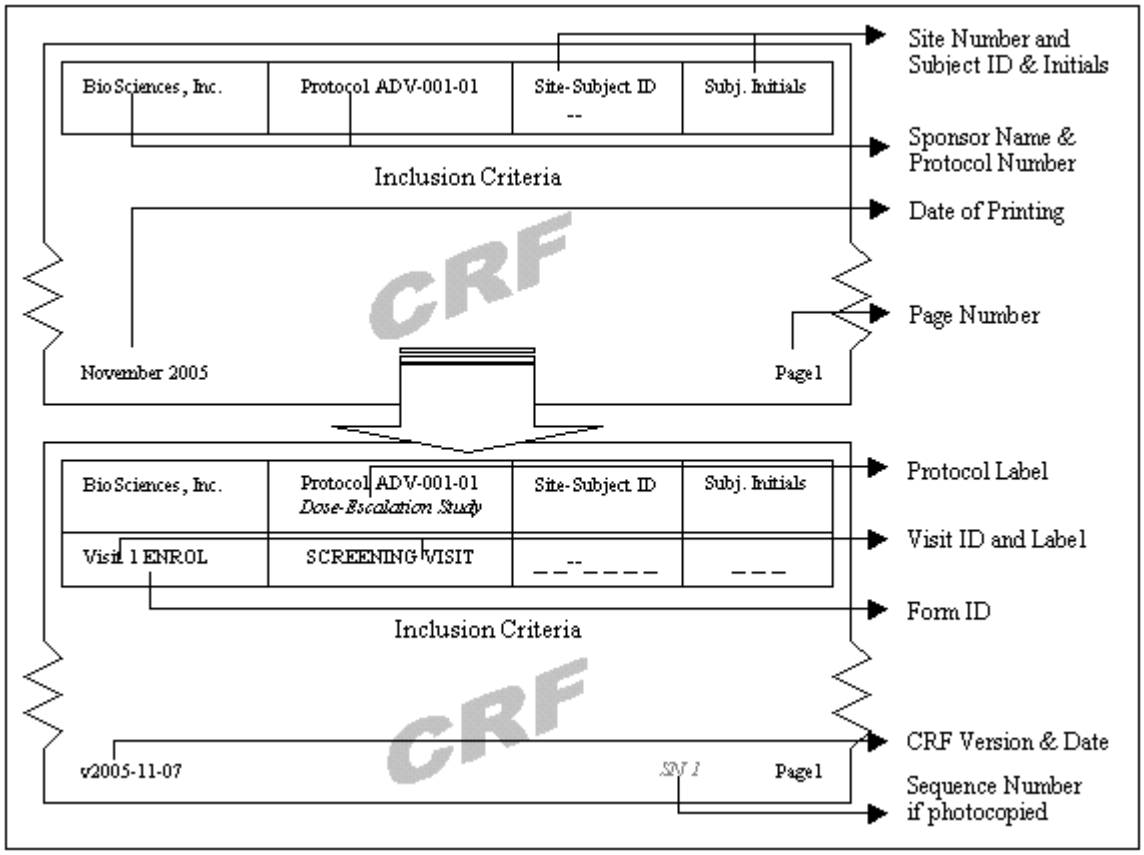
A well-designed CRF is informative and structured that the electronic database design and data manipulation for statistical analysis can be simplified. It also promotes capturing legible, consistent, and valid data, which lessen the load on data entry error reconciliation and query generation. This paper explores CRF design techniques in terms of CRF layout, field organization, and response type design, and suggests including CRF completion guidelines for efficient data management and statistical analysis.

3. Well-Referenced CRFs

One of the components of well-referenced CRFs is an informative header and footer. Typical headers and footers contain information such as Sponsor ID, Protocol number, Subject ID, Subject initials, date of printing, and page number, all of which uniquely identify a CRF page. Although these fields are sufficient in managing data, additional information such as protocol

label, visit ID and label, form ID, CRF version and date printed, as well as sequence number make database design easier and a better link to the study database. Using this information, *Figure 1* shows how header and footer contents of a CRF template can be improved. The protocol and visit labels are informative features that provide brief descriptions of the study and the schedule of assessments. The form ID is used to identify how the CRF page is linked to the database. The visit ID identifies how observations are archived in the database, and also acts as a unique identifier of time dependent data when it is compiled into a dataset as per schedule of assessments. The CRF version number is a critical field that not only prevents using an incorrect CRF page but also confirms that there is no change made to the page. If a different version number is used, it warns users about possible changes within the page and the database should be redesigned to accommodate the change.

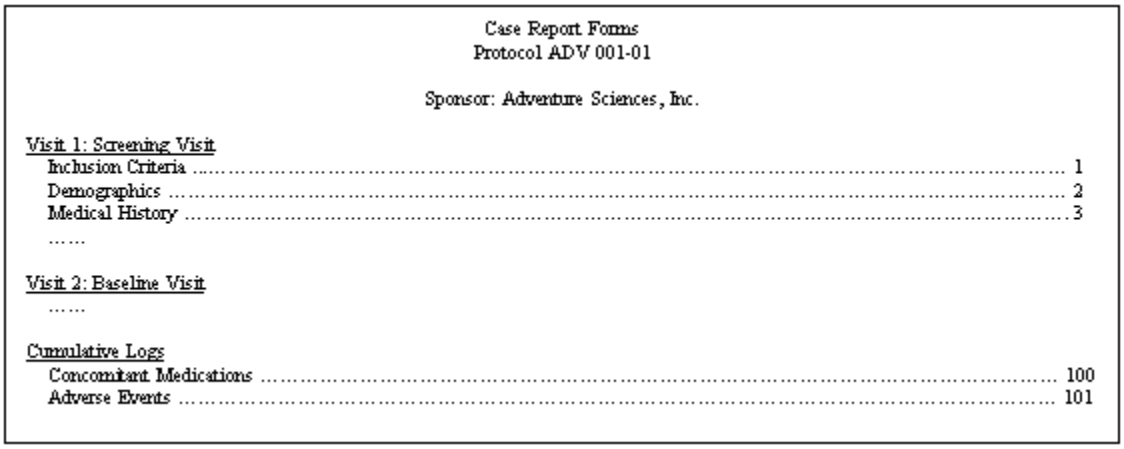
Figure 1. How to Improve Header and Footer Designs of a CRF



All pages of the CRF booklet should be numbered in sequential order. This helps in identifying queries from data validation and manual review. In case of only one unscheduled assessment or cumulative log page being pre-printed in the CRF booklet, a sequence number field can be inserted in the footer to identify the sequential order of photocopied pages. The sequence number is useful in the retrieval of CRFs and constructing the database. Alternatively, multiple pages can be pre-printed and placed in 'Unscheduled' or 'Cumulative Log' sections of the binder or stored in a separate binder to be used and retrieved if and when necessary.

It is essential to organize CRF pages in a structural order, reflecting the schedule of assessments specified in the protocol. A table of contents acts as a reference for the sequential order of CRFs. Not only does it provide site personnel with a quick reference to specific pages (for recording the study data) but it also helps define the database in a structural manner. *Figure 2* illustrates an example of the table of contents that includes information about Visit ID, label, CRF contents and CRF sequential order, specified by a page number.

Figure 2. An Example of a Table of Contents for CRFs



Case Report Forms
Protocol ADV 001-01

Sponsor: Adventure Sciences, Inc.

Visit 1: Screening Visit
Inclusion Criteria 1
Demographics 2
Medical History 3
.....

Visit 2: Baseline Visit
.....

Cumulative Logs
Concomitant Medications 100
Adverse Events 101

A well-designed CRF regarding laboratory data collection captures key parameters that provide a link to the central laboratory. In a multi-center study, a central lab may be used to analyze the blood samples and provide the results in an 'analysis-ready' dataset. It minimizes transcription errors and ensures data quality control. Although it may be considered superfluous to collect sample information on the CRF, it is an important factor for validating the lab data and ensuring that all records are in the dataset. In order to facilitate the data validation process, the CRF should collect the sample date and time, fasting information, and accession number for each sample taken. In addition, these data can be used to construct an integrated database. Thus, a well-design CRF contains these well-referenced CRF components that make database design and data validation processes more efficient.

4. CRF Design Layout

There are three types of data: non-time dependent, time dependent, and cumulative data. *Figure 3* describes CRF Design layout strategies based on data types and preferred database structure. The CRF design layout strategy should be determined considering a CRF clustered level and the time and frequency of a data review.

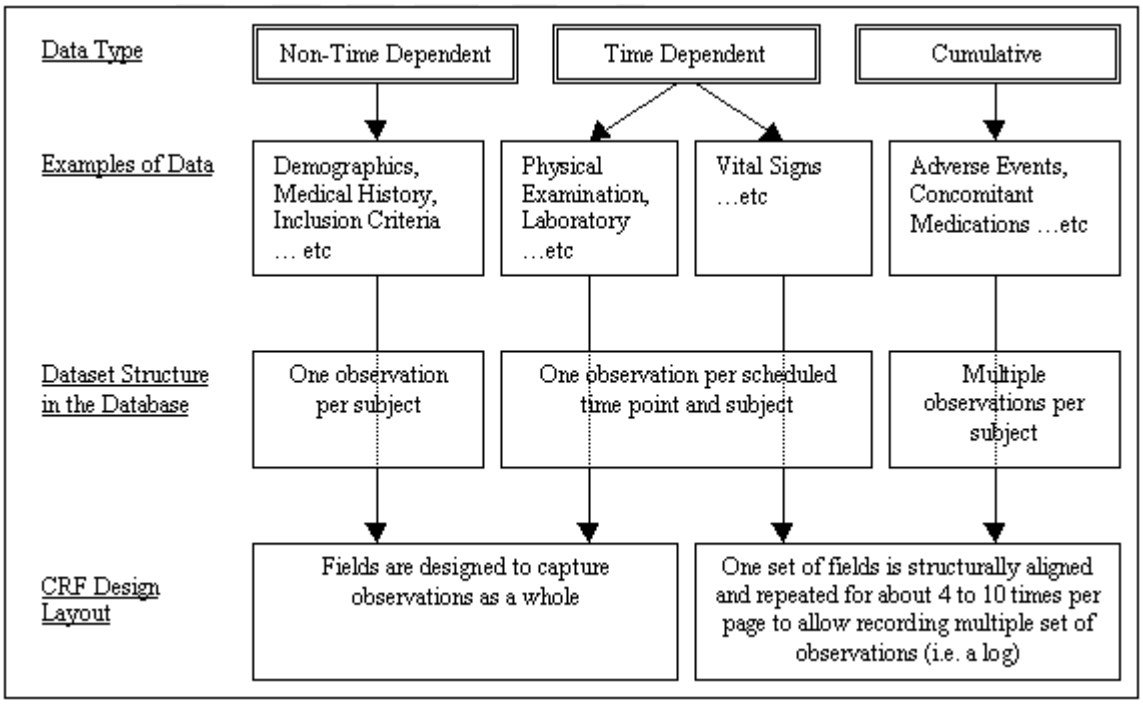
Non-time dependent data: Non-time dependent data is the data collected at a snapshot in time. Such data include subject demographics and medical history.

Time dependent data: Time dependent data is data collected repeatedly over time. A typical example is vital signs recorded at multiple visits. With time dependent data, there are 2 options to the CRF layout: Single page, per visit or a cumulative log. With the first approach, the data is represented at each visit while the second approach is a single page with multiple records representing the "repeated" time measurements. Both approaches have pros and cons. The "per-visit" approach more accurately reflects the schedule of assessments but could lead to a larger

CRF booklet. The "cumulative log" approach saves number of pages in a CRF booklet and makes variables structured in the same way as the CRF page in the database. However, since it does not allow the CRF retrieval as "per-visit", it may be inconvenient for investigators to frequently flip over pages from the log section to the actual visit section. It also restricts retrieval of data for a plan of data review by scheduled visits. Furthermore, it is more likely to yield data entry errors if too many fields are combined into one page and become cluttered. The "per-visit" approach is preferred for assessments such as physical examination or laboratory data as they involve many parameters. The "cumulative log" approach may be preferred for groups of assessments such as vital signs, which involve a fewer number of parameters.

Cumulative data: Cumulative data is data collected over time but not linked to a specific visit. Adverse events and concomitant medications are typical examples. The usual approach to designing a CRF for cumulative data is the "cumulative log" approach described in the previous section.

Figure 3. CRF Design Layout Strategy as per Data Type

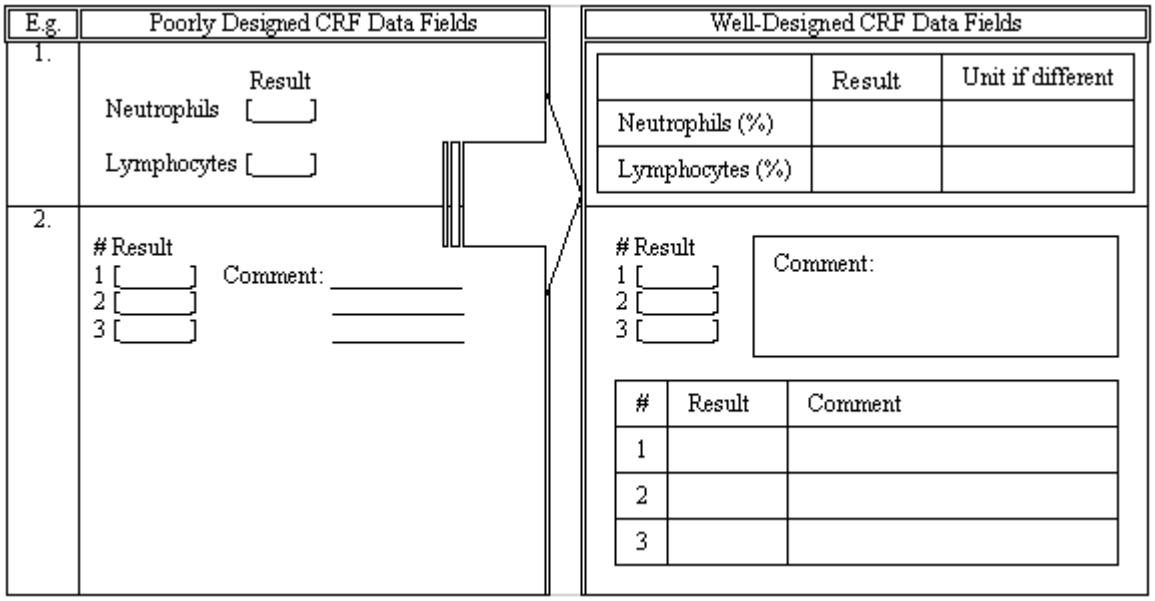


5. Organizing CRF Fields

Well-aligned and structured CRF fields provide a clear direction for data collection and annotating CRFs. *Figure 4* displays examples of how to improve CRF data field structures by comparing poorly-designed with well-designed CRF fields. The first example in the *Figure 4* illustrates advantages of organizing similar fields together by using the example of specifying default units for each laboratory parameter. A test result for Neutrophils can be specified in either a conventional unit or SI unit, especially in multi-centre studies. When the default unit is specified as "%", it is easier to perceive the expected response as a number in the unit of "%". When the result is assessed in a different unit, a conversion can be performed for statistical analysis using the response entered in the field of "Unit if Different". The laboratory data fields that expect a similar data format can be grouped to enforce the efficiency of data manipulation.

The second example, the "Comment" field in the *Figure 4*, illustrates how ambiguous field designs can be improved so that users perceive whether only one comment is expected or whether a comment is expected for each result. Well-organized CRF fields help to prevent from misinterpretation of required responses.

Figure 4. Organizing CRF Data Fields



6. CRF Field Designs

Well-designed CRFs use different types of icons for different response formats. This helps obtain consistency in data format. They provide a clear picture of what is expected in the field for investigators and coordinators to help report valid data. *Figure 5* illustrates the use of different types of icons for different types of data fields such as categorical, text, date, and numeric fields. A square symbol is used to indicate multiple responses being allowed in the field whereas a circle symbol is used to indicate that a single-response is expected in the field. Any category in a multiple or single response field can have an "open-ended" field to elicit additional information. A typical example is the "Race" field where "Other" is an option to be checked. If "Other" is checked, then you can have the site personnel define the specification in an "open-ended" field. A pre-coded response (i.e. 1=Mild, 2=Moderate, or 3=Severe) is primarily used to aid data entry and statistical analysis. It makes data manipulation easier during a statistical analysis process. Providing an example and expected format for a field can reduce data misinterpretations and also reduce the number of un-necessary queries that would be issued to clarify the data. For instance, a date, written as "02/03/05", for instance, can be misread as either "March 02, 2005" or "March 05, 2002" when there is no specification of date format. This affects the validity of data. In order to avoid this, the date format, "dd/MMM/yyyy", can be used as shown in the *Figure 5*. Entering data with correct format assists in establishing validity and consistency of the data. Unnecessary data queries, interpretation and data manipulation can be reduced during the data validation and statistical analysis processes.

Figure 5. Examples of CRF Data Field Designs

Data Field Types	CRF Data Field Examples
CATEGORICAL • Single-Response • Multiple-Response • Open-ended Multiple Response • Coded Response	◯ Yes ◯ No ☐ Key(s) ☐ Button(s) ☐ Key(s) ☐ Button(s); If Button(s) is used, please explain: _____ 1◯ Mild 2◯ Moderate 3◯ Severe [___] 1=Mild; 2=Moderate; 3=Severe
TEXT	Site (e.g. left shoulder): _____
DATE • Standard Format	___/___/____ dd/MMM/yyyy ☐☐ ☐☐☐ ☐☐☐☐ D D M M M Y Y Y Y
NUMERIC	Number of Key(s) being used [___] (e.g. 0, 1, 2 …etc)

7. CRF Completion Guidelines

CRF completion guidelines can be inserted to provide study-specific data collection procedures under industry regulatory guidelines *ICH Guidance E6: Good Clinical Practice: Consolidated guideline*. The guidelines help to bridge the gap between the study protocol and the users in regards to CRF completion, correction, signing and handling procedures. Data formats for appropriate response fields, a data correction guide, how to handle unknown or unavailable data, and a retrieval schedule for completed CRFs can be outlined. For instance, a procedure for completing subject initials in the header can be included. The guidelines can remind investigators that their initials cannot be changed even if a subject's name is changed as a result of marriage or divorce during the study. They can also instruct how to handle missing or unavailable data. If a required piece of information or entire section cannot be retrieved, the use of "NA", "ND" or "UNK" can be defined to avoid ambiguous responses. To handle an unknown or unavailable date response, an imprecise date format can be suggested in the guideline. When 'day' and/or 'month' of date are unknown, unknown day and month format can be defined as "UK/UNK/2000" and forms that do not allow an imprecise date (eg. Study Drug Dosing) can be listed. Moreover, data correction rules can be specified such as drawing a single line through the original entry with a signed, dated correction. The guidelines can also demonstrate how to complete CRF pages for unscheduled assessments or in the cumulative log. Furthermore, CRF retrieval procedures including how to handle CRFs for subjects who have discontinued during the study can be listed. In order to enhance the efficiency of CRF completion procedures, there should be a study-specific CRF completion guidelines.

8. Conclusion

A well-designed CRF booklet can help define a structured database and collect valid and consistent data in a clinical trial. It reduces time on data query management and increases the efficiency of statistical analysis and output generation. This paper presents CRF design techniques and considerations for efficient data management and statistical analysis. Firstly, it introduces well-referenced CRFs using a table of contents and informative header and footer to show data collection and tracking procedures, and to assist in constructing a structural database. In addition, the well-referenced CRFs employ CRF fields, used in reconciliation of central laboratory data. Secondly, it illustrates CRF layout, field organization, and response type design techniques. The layout of a CRF should be determined by data types and the structure of the dataset in the database while considering a cluster level of fields in one page as well as a plan of data review by scheduled visits. The well-organized CRF fields give users a clear direction of required responses. The response format should be specified in the field to ensure accurate data transcription and consistency. Finally, the paper suggests providing CRF completion guidelines which contain detailed instructions for completing and correcting CRFs, signing procedures, and handling of completed CRFs under the industry standards. All of these techniques and considerations contribute in developing well-designed CRFs under "best practices".

[i] ICH Guidance E6: Good Clinical Practice: Consolidated guideline