PhUSE 2016

Paper DH07

How to review a CRF - A statistical programmer perspective

Elsa Lozachmeur, Novartis Pharma AG, Basel, Switzerland

ABSTRACT

The design of the Case Report Form (CRF) is critical for the capture and subsequent analysis of subject and patient clinical trial data. It is imperative that a strong collaboration exists between the programmer and other key stakeholders who either influence or are impacted by the CRF.

CRF input has always been a fundamental and critical part of the Statistical Programmer's role. In relatively recent times, the uptake of CDISC standards has sharpened the focus for programmers to significantly contribute in the creation and review of highly standardized CRFs.

This paper will outline the importance of a well-designed CRF and its relationship and place in the clinical trial document landscape. Furthermore, the paper will provide guidance for statistical programmers to facilitate a seamless dataflow process and ultimately pay dividends for programmers during the conduct of analysis.

INTRODUCTION

A Case Report Form (CRF) is a printed, optical, or electronic (eCRF) document designated to record all of the protocol required information to be reported to the sponsor on each trial subject (source: ICH GCP, section 1.1).

For the remainder of this paper, the term CRF is used. However, eCRF could be equally considered.

The main purpose is to collect the specific data sponsors' need in order to test their hypotheses or answer their research questions:

The CRF is link between the protocol, the database and the Statistical Analysis Plan (SAP). It gives the programmers the information on how the data are captured and collected.

This paper will cover:

- Risks / consequences for poorly managed CRF process
- The big picture – Clinical trial document landscape
- Keys roles and responsibilities involved
- Responsibilities and guidelines for Statistical Programmers

In conclusion, the CDISC standard usage and its advantages will be mentioned.

RISK / CONSEQUENCES FOR POORLY MANAGED CRF PROCESS

Some of the risks (not exhaustive) due to a poorly managed CRF process can be:

- Increase the time of CRF creation
- Delayed on clinical database creation and go-live
- Collection of redundant data
- Miss collection of relevant data (i.e. question on CRF not precise enough, may lead into wrong direction)
- Increase number of queries
- Create strain on internal resources due to iterative review circles
- Increase of the likelihood of database changes and unlocks
- Increase downstream impact on programming such as remapping of variables, pre-processing before using the collected variables, complex derivations to be able to address hypothesis described in the protocol and SAP

Some general guidance could be followed to overcome these risks. Indeed good CRFs are crucial in conducting a successful clinical trial. CRFs capture data that will be used to evaluate the research questions asked in the protocol.

The review process has to involve the right person at the right time. The reviewers should focus on the following points to verify if the CRF is well-designed so good CRFs should:

1. Gather complete and accurate data that answer study questions
2. Promote accurate data entry
3. Organize data in format that facilitates data analysis

For point #1, the reviewers could check the following items:

1. Avoid duplication of data
2. Ease transcription of data onto the CRF
3. Compliance with the study protocol

For point #2, the clinical team should provide:

1. Visual cues to the person recording the data such as boxes that clearly indicate where data should be recorded
2. Clear guidance about skip patterns
3. Clean, uncrowded layout

For point #3, one way to organize the data would be to group the same form data that will be analyzed together, where possible.

THE BIG PICTURE – CLINICAL TRIAL DOCUMENT LANDSCAPE

The two tables below describe the people (Table 1) and the documents (Table 2) detailed in Figure 1:

Table 1:

ACRONYM	NAME	DESCRIPTION
CTT	Clinical Trial Team	The team responsible for the successful planning, conduct and execution of the trial. The team consists of many functions such as global trial leader, clinical managers, clinical scientist, data manager, trial statistician and statistical programmer
SL	Study Lead	Is the global trial leader
LDM	Lead Data Manager	Is the lead for all Data management activities
TS	Trial Statistician	The Stats representative on the CTT and would usually create the Statistical Analysis Plan.
TP	Trial Programmer	is responsible of all programming activities on the trial

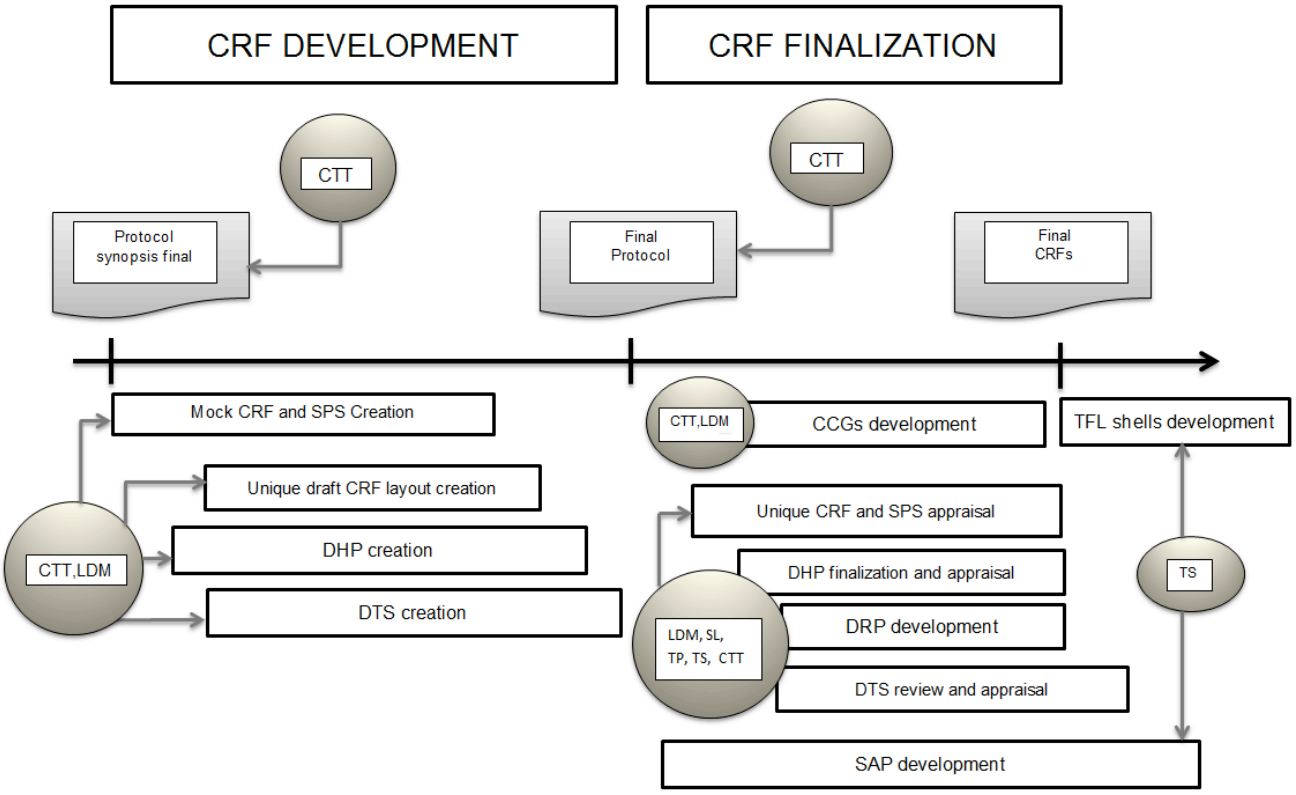
Table 2:

ACRONYM	NAME	DESCRIPTION
SPS	Study Plan Specification	Describes all domains, all variables within each domain and their attributes.
DHP or DMP	Data Handling Plan or Data Management Plan	Documents trial specific information about collection and handling of data and should be reviewed as agreed by the CTT
DTS	Data Transfer Specification	Specify the format, content, and transfer process of electronic data to the sponsor from the vendor.
CCGs	CRF Completion Guidelines	Data entry instructions
DRP	Data Review Plan	Lists all edit checks to be used to check collected data.
SAP	Statistical Analysis Plan	Defines the analysis to meet protocol objectives
TFL	Tables, Figures, Listings	List of Tables, Figures, Listings and corresponding shells

The image below depicts one approach that could be used. Different companies will have varying approaches.

It is recommended to start the CRF development when the protocol synopsis is final. As shown in the Figure 1, several documents mostly created by Data Management (DM) are developing in parallel with CRF development. All these documents are useful to understand how the database will be built, how the data are entered in the database, how the fields present on the CRF page will be translated into variables and their attributes are also described in one of these documents. All DM documents are reviewed, approved and finalized.

Figure 1:



KEY ROLES AND RESPONSIBILITIES INVOLVED

On Figure 1, we can see that during the CRF development phase, the CTT and LDM are involved in the creation and development of CRF but also on the creation of the complementary documents essential to the database construction and understanding of the database.

The CTT review the draft CRF and discuss the following:

- Consistency to protocol requirements and project standards
- Assures that the CRFs are clear and usable by the sites
- Accuracy for data capture
- Compliance with the statistical analysis plan
- Use of or agreements to standard CRFs whenever possible
- Confirms that the database can be created based on the CRFs

The TS and TP are more heavily involved during the CRF finalization. They are both key in the review of the CRF. Their primary responsibility is to ensure that the data are collected in an appropriate manner to enable successful completion of analysis.

RESPONSIBILITIES AND GUIDELINES FOR STATISTICAL PROGRAMMERS

As stated above, the TP has a key responsibility in ensuring that the analysis can be generated from the data collected. The TP acts as the link between the data and the report tables and quite often between the Statisticians and Data Managers.

In order to understand better the link between the protocol, the CRF and the SAP, an example via real case study that is outlined in Appendix 1.

GUIDANCE FOR STATISTICAL PROGRAMMERS

The design of the CRF is vital to the success of a clinical trial. A good CRF design can reduce the confusion during the data collection which ensures the accuracy of reporting in the end. Since statistical programmers have an in-depth knowledge of each data point utilized in the programming, their involvement can help to ensure that critical data points are collected efficiently via CRF. Another point illustrated with the example in appendix 1 and in the figure 1, is that the clinical documents are not only linked to each other but also mutually dependent. With that in mind, we can then draw up a kind of check list for statistical programmer to help in their review:

- Ensure collected data answer the protocol questions and satisfy the analysis requirements
- Page layout: check the instance/fields that can affect the dataset structure.
- Consider the datasets that would be created for these tables
- Quite often, similar subsequent studies will not go through as thorough a review as the original to optimize operational effectiveness. However, when reviewing a CRF, it is important to establish the differences from previous studies before relaxing any such reviews given the importance of the document and its potential impact downstream.
- All code lists displayed in the CRF use or map to current published CDISC Controlled Terminology
- All collected fields have a Controlled Terminology associated
- Variable names have not more than 8 characters (Health Authorities requirement)
- Ensure consistency across associated documents such as non-CRF data collection specifications or Data Management Plan
- Ensure alignment with Therapeutic Areas specificities
- Ensure data collected can be easily usable by their programs to generate programming requirements

In addition, statistical programmers are also often expected to play a significant role in mapping each and every CRF data point in to the statistical programming analysis database. To perform this activity, they should be familiar with

industry quality standards, guidelines and procedures. For example, a good knowledge of the Study Data Tabulation Model (SDTM), which defines a standard structure for study data tabulations that are to be submitted as part of a product application to a regulatory authority such as the United States Food and Drug Administration (FDA), is definitely helpful. With this knowledge they can ensure usage of data collection standards at every step of drug development and avoid any deviation.

CONCLUSION

The CRF review is a process where statistical programmers should be involved in an early stage as they will ultimately use the data to conduct the analysis. With accurate and high quality documentation such as a well-designed CRF, protocol, SAP and data collection specifications, the data from the database to the analysis will flow more seamlessly.

In addition, the usage of CDISC standards can have the following advantages:

- Allow efficient data acquisition and efficient data monitoring at sites
- Eliminate duplication of collected data
- Reduce transcription errors and improve the quality of data
- Improve site monitoring to minimize the need for cross-reference data in multiple sources
- Make it easier for investigators to conduct clinical research
- Facilitate the inspection and reconstruction of clinical investigations by FDA
- Maintains quality and consistency of data capture and reporting

As the standard CRF pages should have been agreed by all line functions, the data structure is known in advance and then the programmer can use the CDISC standard dataset structure as well and be compliant with the health authorities' requirement. Thus, it enables greater efficiency by shortening the development time of the CRF application and saving programming resources. Also, it allows the focus on those data that have been identified as key to supporting the successful registration and marketing of the study drug.

REFERENCES

- https://www.federalregister.gov/articles/2015/06/26/2015-15644/source-data-capture-from-electronic-health-records-using-standardized-clinical-research-data
- 2013 FDA Guidance for Industry: Electronic Source Data in Clinical Investigations
- CDASH: a must for EDC trials, author: Joris de Bondt, data management coordinator , sgs life science services
- PhUSE 2009 DH05 - This study is essentially just like that other one, Louise Mazzeo, Quanticate, Hitchin, UK

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Elsa Lozachmeur
NOVARTIS Pharma AG
Work Phone: + 41 6 132 40 110
Email: elsa.lozachmeur@novartis.com

APPENDIX 1 – REAL STUDY CASE

- From the protocol:

The primary objective is to demonstrate that at least one dose regimen of treatment xxx in ambulatory sporadic inclusion body myositis (sIBM) patients will increase the distance traveled as measured by change from baseline at week 52 of the 6 minute walking distance test relative to placebo. The 6 Minute Walking Distance test is done as following: the results of each test are recorded into an electronic device, including the physical length of the course, the number of completed laps that the subject walked during the test, and the distance of any non-completed lap by the subject during the test. The electronic device will calculate the total distance walked in meters. In table 1, the schedule visit assessments are described.

Table 1:

EPOCH	SCREENING		TREATMENT														
Visit name/ study week	Screening	BL	Day 1	Wk2	Wk4	Wk8	Wk 12	Wk 16	Wk 20	Wk 24	Wk 28	Wk 32	Wk 36	Wk 40	Wk 44	Wk 48	Wk52/ EoT
Day	-28 to -6	-5 to -1	1	15	29	57	85	113	141	169	197	225	253	281	309	337	365
6-minute Walking Test	X	X				X		X		X		X		X		X	X

- From Statistical Analysis Plan (SAP):

The primary variable is the change from baseline to Week 52 in 6 minute walking distance (6MWD), measured in meters. If a subject was present but did not perform the test at a visit because the subject was physically incapable of performing the test the result of the 6MWD will be set to zero meters. Here the physical incapability of performing the test is considered the clinical outcome of the test rather than as missing data.

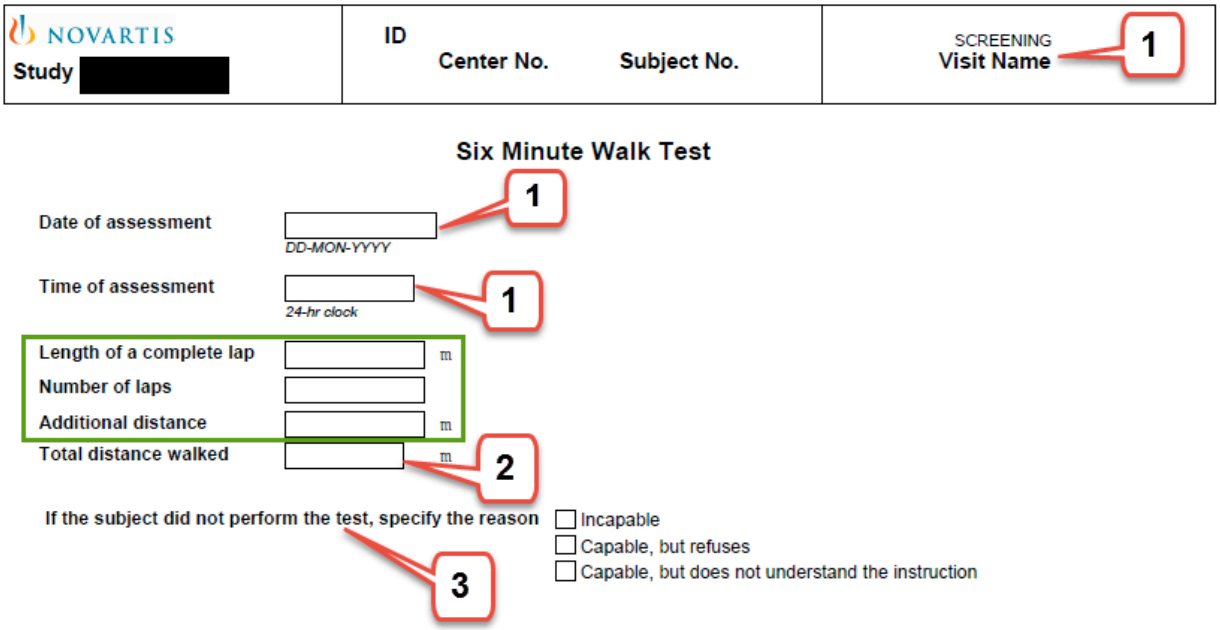
In the protocol, the objective is explained and detailed: increase of distance walked by performing a particular test such as the 6 minute walking test. Then this is translated in an analysis in the SAP with some rules for specific cases: e.g. if patient is incapable then the result is set to zero.

With this information, I know what I need to derive the primary endpoint as described in the SAP. In order to derive this, I need to:

1. Determine baseline and week 52 assessments
2. Retrieve total distance walked
3. Determine reason not done so the rule about incapable patients can be applied

In Figure 2, the data to support the above requirements is annotated with the appropriate number. As discussed previously in the paper, highlighting redundant data is an expected responsibility from CTT members. In Figure 2, a green box highlights what looks like redundant data based on the analysis requirements outlined above. However, no action was taken as this lower-level granular information was deemed useful for clinical review.

- Figure 2:


NOVARTIS
Study [redacted]
ID
Center No.
Subject No.
SCREENING
Visit Name
1
Six Minute Walk Test
Date of assessment
DD-MON-YYYY
1
Time of assessment
24-hr clock
1
Length of a complete lap
m
Number of laps
Additional distance
m
Total distance walked
m
2
If the subject did not perform the test, specify the reason
Incapable
Capable, but refuses
Capable, but does not understand the instruction
3
ZW12I01
Epoch
Verified
Approved
Locked
Frozen
Document Number
Page Name
6 MIN WALK TEST-BASELINE
Subevent#
Status
Page Number
53.1
Version