

GD007

Clinical Data Management Study Setup

In SOP077 - Data Management Overview, it describes that before the study can be set up, the protocol must be nearing completion and a Data Management Plan (DMP) should be drafted. This is because most of the decisions for setup will be completed in that process.

1. Precursors

- a. The protocol should be reviewed by the Data Management Lead (DML) to make sure that the primary and secondary outcomes are achievable. In particular, the inclusion criteria should be reviewed to make sure they are clear to use and worded correctly. For example:
 - i. An inclusion criteria worded as “Patient must NOT be pregnant” would be better as the exclusion criteria “Is the patient pregnant”.
 - ii. Age must be greater than 18, implies that the youngest person eligible is 19.
 - iii. Questions that are not clear cut, “Women of childbearing potential” as this could mean different things to different people.
- b. A Data Management Plan (DMP) should be started, with the basic framework and the initial data decisions; the DMP is a living document which will be updated during the study and finalised at the end of study. SOP078 – Data Management Plan covers this in detail.

2. Design

- a. The first step of the design process is to convert the protocol into variables to be collected. The variable list (FRM047) is the best method, as it helps to gather all the items that will be collected and their metadata. This is the when, what, why and other details of the variable. It is completed by the clinical data lead (CDL), with support from the Principle Investigator (PI) and data team.
- b. Once the variable list has been completed, the data manager should review it before meeting the CDL to clarify any questions.
- c. Once the Variable list is finalised the CDL/PI, the data manager and the study statistician should review both to confirm that the variables meet the requirements of the protocol. Once all have been approved the build can start.
- d. To assist in the timely review of Serious Adverse Events (SAEs) all studies using OpenClinica should enter their SAEs directly into an eSAE form in OpenClinica.
 - i. Templates for these are available within OpenClinica test site, but as each study has their own SAE requirements, FRM055 eSAE Case Report Form (CRF) Checklist, list all the options for the eSAE form.
 - ii. The DML should liaise with the Clinical Project Manager and the CDL to finalise the design for the eSAE form using FRM055.

3. Build

- a. See appendix A for design considerations for eCRFs.
- b. Building in OpenClinica
 - i. The data manager starts building the CRFs for OpenClinica, basing them on the variable sheet and the DVP.

- ii. They are then imported into OpenClinica and the study is built with a schedule of events.
 - iii. Once the data manager has built the first draft of the CRFs, these can then be shown to the CDL to ensure that they are meeting expectations.
 - iv. The data manager then proceeds with programming the CRFs with rules: first the rules that provide flow, then the validations.
 - v. Throughout the build process the data manager test the CRFs to make sure they function as expected.
 - vi. As it has been found that PIs and CDLs sometimes find eCRFs unclear. The design/build circle might change between eCRFs and pCRFs.
 - c. Building in Excel
 - i. Once the paper CRFs have been developed the design of the spreadsheet can start.
 - ii. It is advised to use multiple tabs for different CRFs within an event/visit and a spreadsheet for each event.
 - iii. The validations should be programmed based on the DVP.
 - d. Building in other data repositories.
 - i. This is not recommended, but should follow as much as possible, the OpenClinica or Excel steps, depending on the software.
- 4. Test
 - a. The first testing is informal test by the data manager, confirming that the whole set of CRFs work correctly.
 - b. The next step is the formal user acceptance testing (UAT). In the DMP, a decision will have made if the UAT is to be fast tracked or not. Clinical Trials of Investigational Medicinal Products (CTIMP) studies may not be fast tracked. (See GD008 for further details on UAT)
- 5. Validation
 - a. Depending on the complexity of the study, the Data Validation Specification (DVS) (FRM048) can be used to record all of the validation rules within the database.
- 6. Production
 - a. The first step of the move to production is getting final approvals (FRM018) from the PI/CDL, the statistician and the data manager. Depending on the complexity of the study this could include Clinical Project Managers (CPMs) and other staff. For direct entry studies this must be completed before any patient is enrolled; for others it is strongly recommended.
 - b. Moving to production.
 - i. OpenClinica
 - 1. Before moving to production, paper CRFs, if applicable, need to be made.
 - 2. Once approved, a registration form (FRM051) is required.
 - 3. Then the study can be set up on the production site.
 - 4. The CRFs are copied into a new directory then renamed and renumbered following the Version control SOP (SOP60). They are then imported into the production site of OpenClinica.

5. The rules need to have all the Unique Object Identifiers (OIDs) changed to match the OIDs in the production site. This can be done manually or using a utility.
 6. The new study must be tested again, following the last UAT script, to confirm the transfer to production did not introduce any errors.
 7. Once confirmed the study can be made available and therefore live.
- ii. Excel
 1. The spreadsheet(s) should be copied to a different location and all test data removed. The access to this new location should be confirmed to be secure.
 2. As there is no audit trail in Excel, it is important that on each day that data is changed, the spreadsheet must be saved and named with that date. The date should be in the format yyyyymmdd, for example 20161225.
7. Paper CRFs (pCRFs)
- a. As Royal Papworth moves to being a Paperlite hospital, pCRFs should be discouraged. Where possible data should be directly entered from source into the eCRFs/data repository.
 - b. The data team will provide current template for pCRFs, and the pCRFs should be created by either the CDL or the data team.
 - c. In general, it is best that the pCRFs are made after the eCRFs, though this can be dependent on which data repository is being used.
 - i. OpenClinica – The paper CRFs, if required, will be made during the build or production step, depending on the complexities. They are based upon the eCRFs. Once completed they should be tested against the eCRFs.
 1. The way in which the eCRF and pCRFs are developed in relation to together may vary between studies. It would be preferable to create eCRF and use the final version to create the pCRF, however in practice this is not always the most efficient method as it can be more practical to develop the design using the pCRFs which allows a better view of the flow. The Data Manager should decide on the best approach to take based on the complexity of the study and the staff member's preference.
 - ii. Excel – It is likely that making the pCRFs before starting the build on the excel sheet(s) is the best way to go.
 - iii. Other – Dependent on the software to be used in the data repository, these should either be before or after the eCRF build.
 - d. pCRFs should be tested before use. The data used in the test should be based on real patient data, but changed in such a way that no patients data in being used without their consent. This is vital as the units, number formats and order of questions might seem correct, until looked at in real environment.
 - e. The pCRF must be finalised prior to the start of the trial and before any patient is enrolled into the trial.
 - f. See appendix A for design considerations.

8. Changes to Design after Moving to production
 - a. Following these steps should reduce the need for changes in the pCRFs/eCRFs after the designs have been finalised, but sometimes protocols change during a study, or errors are discovered.
 - b. The CRF Design change form (FRM037) should be used to document the changes required.
 - c. When a design change is requested, part of the process for review is whether the change should be initiated for previous patients. One consideration for deciding this is the Source Data Verification (SDV) status of the previous patients (refer to GD012 Data Cleaning for further guidance).
 - d. It should be noted that in some data repositories any changes to CRF may change the way that data is exported, and therefore extra care needs to be taken when preparing data for analysis.
 - e. The level of testing required is dependent on the design change. Major changes should be implemented in the test environment and a partial UAT should be performed. Minor changes can be done directly on the production instance. The data management lead should determine whether the change is major or minor.

Appendix A – CRF Design Considerations

1. General Principles

- a. The CRF must be designed to ensure compliance with the study or trial protocol.
- b. Patient identifiable data must not be collected in the CRF without appropriate regulatory approvals and patient consent e.g. name, address, hospital number, NHS number.
- c. Ensure that the data to be captured are entered into the CRF in a logical order to minimise transcription error, particularly when data will be transcribed from source documents to CRF – take in to account, for example, the order of trial procedures during a trial visit, order of entry in medical notes and laboratory reports.
- d. Separate the CRF into sections by visit to ease data entry. It may be helpful to include a checklist at the front of each section as reminder of assessments required per visit.
- e. Avoid unnecessary calculations and conversions.
- f. Provide a study schedule to indicate which CRF pages are applicable to particular visits.
- g. Use questions, prompts and provide instructions. These must be clear and concise and must not contradict other CRF pages or the protocol.
- h. All inclusion/exclusion criteria questions must be exact quotations from the protocol and provide a Yes and No tick box for each criteria.
- i. Use Yes/No options or similar rather than single check boxes to minimise ambiguity. As if single boxes are unchecked, this could mean “no” as intended, but also be that the question was skipped.
- j. Avoid free text responses.

2. Format

- a. The CRF should be clearly version controlled (see SOP077 for guidance of CRF version control)
- b. Units should be specified if appropriate, ensure units are the same as specified in the protocol and source documents.
- c. A designated area for sign-off at eligibility review and at the end of the CRF by the investigator should be made available.
- d. Layout across CRFs should be similar, cohesive and not be overly cramped.

3. Design of paper CRFs

- a. Each page of the CRF should have the following information as standard, so that in the event of the page being separated from the rest of the CRF, it will be possible to reunite them: Patient Study Number, Date of Birth, Version and Page Number.
- b. The pages should be numbered in sequential order. In cases where multiple page insertions are required, e.g. Adverse Event Form and Concomitant Medication Form, the use of sub-numbering may be appropriate. (The following numbering pattern can be followed: the initial page as XXX.0, the suffix will increase sequentially with each inserted page as XXX.1, XXX.2, XXX.3).
- c. Make sure to have an appropriate number of boxes for the digits required, for example: date, laboratory results and vital signs readings.
- d. Pre-printed decimal points should be used where applicable.

- e. When laying out the pCRF, be consistent throughout the entire CRF, including alignment, page margins, spacing and fonts.
- f. Make sure to have sufficient space in the page margin to accommodate hole punching/binding.