TESA Data Management Workshop

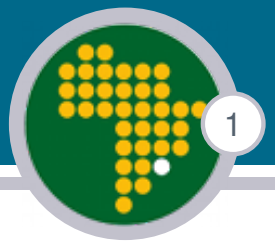
Data Management in Clinical Research

Management of clinical trial data
Data Analysis

Arsénio Nhacolo
Joe Brew

Blantyre, Malawi, July 09-12, 2019

Outline I



Management of clinical trial data

Day 1 (Arsénio Nhacolo)

- Introduction
- Data management plan
- Data entry
- Data traceability
- *Practical session*

Day 2 (Joe Brew)

- Managing data quality and validation
- Data coding
- Data sharing
- *Practical session*

Outline II

Data analysis

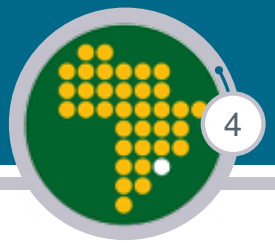
Day 3 – Part 1 (Joe Brew)

- Study registration
- Information in the protocol
- Statistics and statistical plan (SAP)
- Interim analysis
- Publication
- Data processing
- Data cleaning

Day 3 – Part 2 (Arsénio Nhacolo)

- Statistical analysis
  - Descriptive statistics
  - Exploratory data analysis
  - Confirmatory data analysis
- Communication
- *Practical session*

MANAGEMENT OF CLINICAL TRIAL DATA



Drug development process

Clinical data are data generated by, and related to, trials testing new drugs or devices in humans.

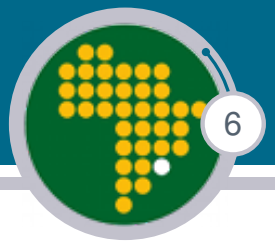
Drug development is done in following phases (also applicable to medical devices):

Pre-clinical phase (*in vitro* and animal studies)

- Candidate compounds to address a particular disease cause, progression mechanism, or symptom are identified in the laboratory or on a computer.
- Promising candidates are moved to preclinical testing (experiments in test tubes and in animals), with the goal of identifying a candidate that is safe and practical for testing in humans.

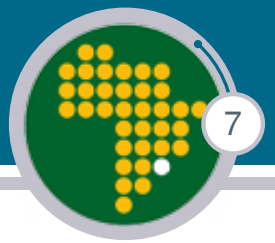
Clinical phase (clinical trials)

- *Phase I:* small and short first in-human testing studies, commonly conducted in healthy volunteers, with focus on safety and initial identification of appropriate dosing.
- *Phase II:* relatively larger and longer studies conducted in the target population, with the main goals being to show effectiveness of the treatment, gather further safety information, and determine an appropriate dose.
- *Phase III:* trials conducted in target population, involving more subjects and with longer duration, aiming to show the effectiveness of the treatment and to assess the benefit–risk profile of the treatment with respect to side effects. Positive results of Phase III trials are used to submit application to the regulatory bodies (e.g. FDA, EMA) to gain approval to market the drug.



Steps in a clinical trial

Initiation	Protocol written, statistical analysis plan drafted, CRF or eCRF designed, study database built and released
Visits	Data recorded on source documents, data transcribed to CRF or eCRF, source document verification/monitoring
Data	Data entry (paper studies only), receipt of electronic non-CRF data, data cleaning
Lock	Complete and accurate data, extraction for analysis and study report



The importance of clinical data management

Clinical data management (CDM) is the work performed on data from a clinical trial from the preparation to collect that data through the time it is extracted for final analysis.

Data management is responsible for delivering complete datasets that are of a quality (accurate, clean) to reliably support a conclusion.

If the data is not accurate, reliable, and analysable, all the resources invested in conducting the study have gone to waste, therefore CDM is of utmost importance.

Clinical data management activities

CDM activities can be divided into the following categories:

- *Study start-up* activities: designing paper/electronic CRFs, specifying cleaning rules (edit checks), building and testing the database, and releasing the study database to collect data. The data management plan is also created during study start up.
- *Study conduct* activities: collecting the data on CRFs, cleaning the data, managing adverse event (AE) and serious adverse event (SAE) collection, and producing reports.
- *Study close-out* activities: focus on ensuring the data is complete and of a quality to support the final analysis.

Study start-up

Data management plan
Concept

Data management plans (DMPs) are created by CDM to document how data management for a given study will be/was carried out.

At the beginning of a study, the DMP provides a focus for identifying the data management work to be performed, who will perform that work, and what is to be produced as documentation of the work.

During the study, the DMP is updated as key elements of the data management process change so that at the end of the study, the DMP provides an accurate record of how the study was carried out.

Data management plan
Concept

DMPs are not required by any law or regulation but are so common across biopharmaceutical companies that they are considered an auditable document.

After looking at standard operating procedures (SOPs) and training records, an auditor investigating clinical data management practices will typically ask for the data management plan for a study being reviewed.

Data management plan
Contents

What should a DMP contain?

The key topics to cover in a DMP are:

- CRF/eCRF creation
- Database design and build
- Edit check specification
- Study database testing and release
- Data or paper workflow
- Reports and metrics
- Query Management
- Managing lab data
- Managing other non-CRF data
- Coding reported terms
- Handling SAEs
- Transferring data
- Study database lock

Data management plan
Contents

For each of those topics, a DMP should specify:

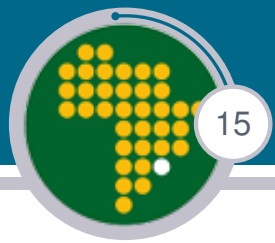
- What is the work to be performed?
- Who is responsible for the work?
- Which SOPs or guidelines will apply?
- What documentation or output will be collected or produced?

In addition to documenting standard CDM activities, the DMP also provides details on the computer systems used to collect clinical trial data as is recommended by the Food and Drug Administration (FDA)'s guidance document "Computerized Systems Used in Clinical Investigations", Section IV.F.

Signing off on the DMP

At some companies the DMP is a document internal to CDM. In this case, the lead or senior data manager for a study creates the document and signs it to show that it is accurate as of a given date.

At other companies, the DMP also serves as an agreement between data management and other groups, such as clinical operations and biostatistics, as to how the study will be run. In that case, it would be reviewed and approved by representatives of those groups in addition to the lead data manager.



Revision of the DMP

Whenever there is a significant change in the data management process or computer software during the course of the study, the DMP must be revised and updated to document how to conduct the study from that point forward.

DMPs and the study files

The output documents specified in the DMP and in the SOPs for data management activities must be filed in the *study file* or *data management study file*[1]

The study file may be a folder in a cabinet, a binder in a data manager's office, or an electronic folder on a shared drive.

[1] This is not the same thing as the trial master file that is managed by clinical operations and contains key study documentation required by GCP.

Quality assurance and DMPs

Quality assurance (QA) is the prevention, detection, and correction of errors or problems.

Since good practice is closely tied to following regulations, QA is closely tied to regulatory compliance.

A key requirement of most quality methods is the creation of a plan, and a key requirement of GCP is the documentation of what has happened during a study.

The DMP helps fulfill both of these requirements by creating the plan and detailing what documents will record the conduct of the study. It can be used as the starting point when conducting internal QA audits of the data management process.

SOPs for DMPs

A data management group or department should an SOP for creating and maintaining a DMP. The SOP should define:

- define a point at which the DMP for a given study must be in place;
- state the circumstances under which the DMP must be revised;
- state what signatures are required.

Along with the SOP, there should be a blank template document or an outline for the DMP to assure consistency across studies.

Each section in the template should have instructions on what kind of information and what level of detail is expected.

Using DMPs

Although it might be tempting to avoid spending time planning and documenting everything when there is “real” work to be done, DMPs have benefits that worth the effort:

- Everyone knows what is expected because the work to be done and responsibilities are clearly stated at the start of the study.
- The expected documents are listed at the start of the study so they can be produced during the course of, rather than after, the conduct of the study.
- The document helps everyone fulfil regulatory requirements.
- Data management tasks become more visible to other groups when the DMP is made available to the project team.

Data management plan
Use

- The DMP provides continuity of process and a history of a project. This is particularly useful for long-term studies and growing data management groups.

To avoid overwhelming staff with documentation requirements, managers of data management groups should encourage the use of templates and the use of previous plans as examples.

Other study start-up topics
(Out of the scope of this training)

Apart from the DMP, there are other relevant topics relate to study start-up that are not covered in this training:

- CRF design (goals: collecting required, analysable data, reducing queries, avoiding duplicate, ambiguous data).
- Database design (design decisions, clinical database concepts, standards, QA, responsibilities, *Title 21 CFR Part 11*).
- Edit checks (missing values, range/pattern checks, logical inconsistencies, cross-form checks, protocol violations, QA, connection to queries).
- Preparing to receive data (testing and validating the study database, moving to production, study database change control, QA).

Study conduct

Data entry
Concept

Despite the trending use of electronic data capture (EDC), there is a considerable number of clinical trials still capturing data on paper CRFs.

Data entry is the process of transferring data from paper into electronic storage (database).

Data entry may be entirely manual or it may be partly computerized using optical character recognition (OCR).

The main data entry that have to be addressed are:

- Selecting a method to transcribe the data
- Determining how closely data must match the CRF
- Creating processes to deal with troublesome data
- Making edits and changing data without jeopardizing quality
- Implementing QA of the entire process

Data transcription methods

Errors in transcription are usually due to typographical errors (typos) or illegibility of the values on the CRF.

The following methods are commonly used to reduce transcription errors:

- Double data entry with third-party reconciliation of differences *(blind double entry)*
- Double data entry with a second person resolving differences (*heads-up second entry*)
- OCR as first entry with one or more subsequent entry or review passes
- Single entry with extensive data checking

Match to the CRF

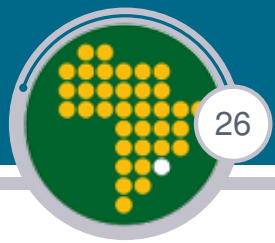
Some questions regarding accurate data transcription need to be addressed:

- Do the data in the database have to exactly match that in the CRF?
  - *Datum must be typed as it is; if it cannot be stored in de database, a discrepancy must be issued.*
- Are data entry operators some flexibility to substitute texts, make assumptions, or change the data? E.g.:
  - Substitute only symbols, e.g., ↑ by increasing.
  - Use standard notations for units (e.g., *g* for *grams*).
  - Abbreviate texts to fit in fields (e.g., *subj* to replace *subject*).
  - Correct some misspellings.
  - More flexibility in the transcription of text fields, and/or changes to values found in numeric or date fields.

Data entry
Match to the CRF

The current common practice:

- except for symbols, data should be entered as seen or left blank;
- any necessary changes are made after entry during the cleaning process so that there is a record of the change in the audit trail of the database along with the reason for the change.



Dealing with troublesome data

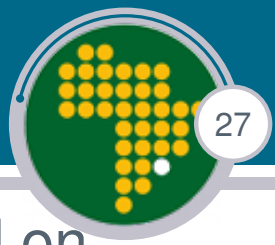
The most common problems with CRF data are illegibility and notations or comments in the margins.

The data management group must specify the procedures for handling each kind of problem in data entry guidelines.

Illegibility fields

Questions to consider when planning an approach to illegible fields:

- Can entry operators discuss the value with each other?
- How do entry operators indicate illegibility?
  - Leave the field blank?
  - Guess and flag the field?
  - Type special flagging characters (e.g.,***)?



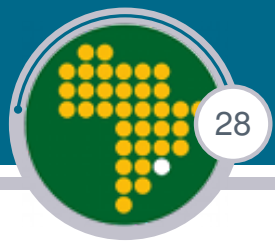
- Should data managers make educated guesses based on a review of other pages?
- Can the clinical research associate (CRA) make a decision based on medical information or experience?

Notations in margins

Sometimes investigators provide unrequested data, frequently as comments in the margins of a CRF page or repeated measurements written in between fields.

Data management together with the clinical team must decide what is to be done:

- Store the data in the database as a comment or annotation (but then it may not be easily analysed).
- Ignore the data.
- Ask the investigators to remove the data or transcribe it somewhere appropriate.



Changing data after entry

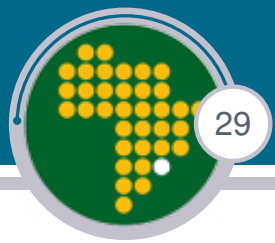
Following initial entry, there are often corrections to the data.

These corrections are identified internally in data management, or by the CRA, or through an external query.

There should be a well-defined process for making these corrections.

Any changes after initial entry, made by any person, must be recorded in an *audit trail*[2].

[2]The Food and Drug Administration (FDA) requires audit trails to record changes made to clinical data (see 21 CFR [Code of Federal Regulations] Part 11), and it should be possible to view this audit trail at any time.



Quality assurance and quality control

Quality assurance (QA) is a process, and quality control (QC) is a check of the process.

QA for data entry builds on good standards and procedures and appropriately configured data entry applications.

The approach that assures quality data entry is documented in the data management plan and is supported by standard operating procedures (SOPs) and entry guidelines.

QC for data entry is usually a check of the accuracy of the entry performed by auditing the data stored in the central database against the CRF (*database audit*).

Data entry
Quality assurance and quality control

Database audits are carried out by people who did not participate in data entry for that study (from data management staff or external QA groups)

Generally the auditing goes as follows:

- Identify the CRFs to be used
- Pull the appropriate copies and associated query forms
- Compare those values against the ones stored in the central database
- Produce audit report listing the number of errors, the number of fields checked, the final error rate, and any action taken

Data entry
Quality assurance and quality control

There must an audit plan beforehand that includes:

- What data at which proportion will be sampled

  Most frequently 10% of the subjects, full CRFs, pages, or data, often supplemented by a 100% of audit of safety fields (e.g., AE) and/or selection of key efficacy fields and/or primary and secondary endpoints.

- A definition of an acceptable error rate

  Generally 10–50 errors per 10 000 fields (0.1%–0.5%).

- What to do if the error rate is unacceptable

The audit plan can be defined in:

- an SOP – if it is consistent across studies
- the data management plan or in a separate audit plan document – if it is study specific

SOPs for data entry

If the entry process is consistent, the process itself may be laid out in an SOP.

> E.g., a data entry SOP may always require blind double entry with third-party arbitration of discrepancies.

If there are variations in data entry across studies, the SOP may only state a commitment to accuracy and indicate that study-specific guidelines are to be followed.

> The data management plan is a good place to identify any study-specific exceptions or changes to the standard procedures.

Data traceability

Data traceability

Data traceability

Data traceability

Bibliography I

[McF07] Eleanor McFadden, *Management of data in clinical trials*, vol. 23, John Wiley & Sons, 2007.

[Pro11] Susanne Prokscha, *Practical guide to clinical data management*, CRC Press, 2011.

[RVW00] Richard K Rondel, Sheila A Varley, and Colin F Webb, *Clinical data management*, Wiley Online Library, 2000.

Thank you!