Clinical Data Management
(Process and practical guide)

Nguyen Thi My Huong, MD. PhD
WHO/RHR/SIS

Training Course in Sexual and Reproductive Health Research
Geneva 2013

World Health Organization
Reproductive Health and Research
UNDP • UNFPA • WHO • World Bank
Special Programme of Research, Development
and Research Training in Human Reproduction

OUTLINE

- Overview of Clinical Data Management (CDM)
- CDM: processes, practical guide and challenges
- WHO online data management (DM) system (OpenClinica)

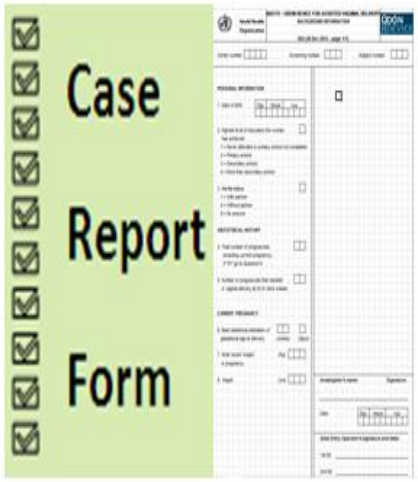
Clinical Data Management (CDM)

- Data - important products of the scientific research
- CDM - a critical task in clinical research that involves in all aspects from data collection to data extraction for analysis
- End result for CDM:
  - Provide a study database that is accurate, secure, reliable and ready for analysis.
  - Accelerate timeline from data collection to analysis

Clinical Data Management (CDM)

- Good CDM - foundation for good clinical trial (CT) that ensures the delivery of the quality data on-time and within the trial budget
- Good DM practices will enable you to effectively create, organize, manage and store data that makes your data easier to use, analyze, share and REUSE in the future

Clinical Data Management process



- Protocol, CRFs development
- DM plan, Database setup
- Training

→ Data collection → Data processing → Clean Database → Data Analysis

CDM process

- DM plan development
- Study setup
- Training
- Data collection
- Data processing
- Quality Assurance & Quality Control
- Audit trail
- Monitoring data quality and data safety
- Security and confidentiality
- Database Closure, Data storage and archive

WHY - CDM?

- WHY do we need to process the research data?
- WHY do we implement edit checks and query inconsistencies?
- WHY do we need GCP-compliant CDM?

Data Management Plan (DMP) development

- DMP describes all the components of the DM process to ensure consistent and effective DM practices
- Good DMP - successful DM implementing
- Each component in the DM process should specify:
  - Work to be performed
  - Responsible staff for the work
  - Guidelines and/or SOPs will be complied with
  - Output will be produced

DMP development (cont'd)

- DMP should be developed for each study and early during the setup of the study
- Provide budget information for DM
- Responsible staff should review and agree with the DMP to make sure a consistent approach to the process and guidelines
- DMP - a living document throughout the study life cycle, to address any updates/changes made during the conduct of the study

Study setup

Includes:

- Case report form (CRF) design
- CRF completion guidelines
- Trial database (DB) setup
- Validation checks

Case report form (CRF) design

- Quality of the data relies on the quality of data collection instruments (CRFs)
- CRFs design:
  - during the protocol development
  - cover all the data specified by the protocol
- Collection of extraneous data adversely affects data quality

CRF design (cont'd)

- Avoid redundant data
  - Unnecessary work for the site staff
  - Unnecessary need for checking data consistencies
- Data based on the same measurement should not be collected more than one.
- Raw data are generally preferable to the calculation based on raw data *(example: DOB is preferable to the age)*

CRF design (cont'd)

- Flow of data from perspective of the person completing the CRF
- Flow of study procedures and organization of data in medical records define the flow of CRFs
- Logically related data should be grouped together
- Separate CRF for each visit: SCR, ADM, FUP…
- Questions and instructions - clear and concise
- Use consistent codes, appropriate date and time formats and units of measurements

CRF design (cont'd)

- Data in coded form:
  - Minimize errors
  - Reduce processing time
  - *Coded formats: drop-down list*
    - *1=yes*
    - *2=no*
    - *3=not sure*
- Consistency in the order of similar response options
  1=yes, 2=no throughout the CRF
- Minimize free text
- Pilot-testing

CRF completion guidelines

- Full, accurate completion of CRFs is critical to:
  - Quality of data captured
  - Fewer queries
  - Quicker validation of data
- Complete, concise and logical guidelines for CRF completion ensure:
  - All required fields are completed
  - Data recorded in the CRFs are logical
  - Free text entries are clinically appropriate

CRF completion guidelines (cont'd)

- Definitions for data items that are not directly measurable *(hypertension)*
- Procedures for making corrections to data

- Handling completed CRFs
- Shipping the CRFs from sites to the DM center

- Update the guidelines

Trial database (DB) setup

- Clinical trial DB contains clinical data and metadata that is structured using CDM software (rows, columns)
- Readable format of the clinical trial DB: SAS, SPSS, Excel spreadsheet…
- Captured clinical data must be entered and stored in a computer system

Trial database setup (cont'd)

- Success of a clinical trial depends on quality and integrity of its DB
- A poor DB design adversely impact DE, data cleaning, extraction and data storage
- Key goal for the DB setup:
  - high quality DB
  - meet both clinical and regulatory requirements
  - store data accurately

Trial database setup (cont'd)

- DB structure considers:
  - ease and speed of DE
  - prevention of errors in data creation and modification
  - efficient creation of data sets for analysis
  - formats of data files requirements
- GCP: "*Ensure that the systems are designed to permit data changes in such a way that the data changes are documented and that there is no deletion of entered data (i.e., maintain an audit trail)".*

METADATA - EJBGI V2.1 (15 Aug 2012)

Item Name	Description	Units	Item Data Type	Response Label	Response Options	Response Values
bgicen	Center number		integer	text	text	text
bgiscr	Screening number		integer	text	text	text
bgisub	Subject number		integer	text	text	text
bgi01	Facility name		integer	facility	, 1- MS Durban, 2- MS Isipingo, 3- MS Cato Manor, 4- MS Umlazi, 5- Mosaic	, 1, 2, 3, 4, 5
bgi02	Assessment types		integer	assessment	, 1- Eligibility, 2- Follow-up	, 1, 2
bgi03	Date of birth	dd-mmm-yyyy	date	text	text	text
bgi04	Language		integer	language	, 1- IsiXhosa, 2- IsiZulu, 3- Afrikaans, 4- English, 5- Other	, 1, 2, 3, 4, 5
bgi04a	Other		character string	text	text	text
bgi05	Education		integer	text	text	text
bgi05a	Post matric education		integer	N/Y	, 1- No, 2- Yes	, 1, 2
bgi06	Work		integer	N/Y	, 1- No, 2- Yes	, 1, 2
bgi06a	What work		character string	text	text	text

Validation checks

- Validation checks:
  - Crucial tool for each study DB
  - Created for all study endpoints and safety data
  - Identify data inconsistencies and potential errors
  - Increases data quality
  - Provide greater efficiency for data cleaning
  - Validation check document - a living document throughout the study life cycle, is updated to CRF changes or errors need correcting

Validation checks (cont'd)

- Missing values
- Valid range
- Logically inconsistent checks across fields or across CRFs
- Protocol violations
- Checks for duplicates

Training

- Effective training ensures:
  - Regulatory compliance
  - Performance effectiveness
  - Job satisfaction of CDM staff
- Staff involve in the DM process must be trained
- GCP: "Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his/her respective task(s).”

Training (cont'd)

- Training documentation (SOPs, guidelines)
- Training content:
  - Consistent across all training materials
  - Consistently conveyed by instructors or mentors
- Types of training is defined by the roles

Data collection

- Clinical data capture can be done using:
  - Paper CRFs (pCRFs)
  - EDC system (online, offline, combination of both)
- GCP requirements:
  - *All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification*
  - *Data reported on the CRF, that are derived from source documents, should be consistent with the source documents or the discrepancies should be explained.*

Data processing

Data processing workflow at the DM center:

- Data receipt
- CRFs tracking
- Data review and coding
- Data entry (DE)
- Data validation
- Query management

Data receipt

- Data receipt vary across the clinical research that may be received through:
  - Fax transmissions
  - Regular mail
  - Web entry or transferred through other electronic means
  - …
- The processes by which data are received, confirmed as received, and made available for DE should be documented

CRF tracking

- All CRFs should be tracked
- CRF logging can be:
  - Manual Subject Form Register (SFR)
  - e-SFR
- Missing CRF should be specified

Subject Form Register

Project: A65779 "Assessment of eligibility and follow-up care for early medical abortion"

CENTER No. 2065 **ASSESSMENT OF ELIGIBILITY** **FACILITY No. AND NAME: 1 – MS Durban**

Screening No.	Subject No.	SCR date	BGI	ASE_CHW	ASE_Clinician	EXE	Screening No.	Subject No.	SCR date	BGI	ASE_CHW	ASE_Clinician	EXE
001							011						
002							012						
003							013						
004							014						
005							015						
006							016						
007							017						
008							018						
009							019						
010							020						

Data review and coding

- Manual review of CRFs
  - For all CRFs before DE
  - Detect data errors, frequently encountered problems
- Medical coding - properly classify the medical terminologies
  - Manual coding using common coding dictionaries *(ICD-10, ICPM, WHO Drug dictionary, Medical Dictionary for Regulatory Activities "MedDRA")*
  - Automated coding

Data Entry systems

Local DE system:

- Data entered onsite by local staff
- Quick data resolutions for omissions, errors, inconsistencies

Central DE system:

- Completed CRFs sent to DM center
- Data entered by experienced DE operators
- Forms stored centrally

Web-based DE system:

- Require only web browser and internet connection
- Secure link provided
- Data transmission is not necessary

Data Entry methods

- **Double DE - independent verification**: Two people enter data and a third person resolves discrepancies between both entries
- **Double DE - blind verification**: Two people enter data (*unaware of what values the other entered)* and the 2nd DE operator verifies data, determines the appropriate entry and saves data (overwrite the prior value)

Data Entry methods (cont'd)

- **Double DE - interactive verification**: Two people enter data and the 2nd *DE operator* resolves discrepancies between 1st and 2nd entry while *being aware of the previous values*
- **Single data entry – review**: One person enters data and 2nd person reviews the entered data against the source data
- **Optical character recognition (OCR)**: Software is used to recognize characters from pCRFs or faxed images then these data are placed directly into the database. Data obtained through OCR should always be reviewed for accuracy

Data Entry guidelines

- Standard conventions for DE ensures consistency in the entry of data throughout the study
- DE timelines: timing expectations between data collection and DE
- Instructions for handling error messages triggered from edit checks

Data validation

- DB automatically checks data against the pre-defined validation rules to detect:
  - Missing values
  - Outliers
  - Inconsistencies
  - Protocol violations
- Validation checks:
  - At the time of DE
  - Run on batches of data

Data validation (cont'd)

- Data validation using descriptive statistics
- Manual review for data validation vs Programmatic validation
- Data validation focus:
  - Primary and other endpoints
  - Key safety fields

Query management

Ensure rapidity of query generation and problem resolutions

- Review validation outputs
- Confirm queries and create query sheets
- Resolve returned queries
- Update pCRFs and DB

View Discrepancy Notes

A65779 (08 Jun 2012)

Study Subject ID	Date Created	CRF	Entity Name	Entity Value	Description	Resolution status	Created By	Proposed answer
T065	08-Jun-12	EJASE	ase02a	21	EJ_ASE02A Q2a: Assessor ID should be 1-20 if assessor is CHW or 21-40 if assessor is clinician	New	fersurellal	
T065	08-Jun-12	EJASE	ase03	11-Jun-2011	EJ_ASE03_2 Q3: Date of assessment should be the same date of SCR	New	fersurellal	
T065	08-Jun-12	EJASE	ase05	2	EJ_ASE05 Q5: should be blank if Q4=negative, otherwise it should be answered	New	fersurellal	
T065	08-Jun-12	EJASE	ase06	01-Feb-2012	EJ_ASE06_1 Q6: LMP should indicate pregnancy of first trimester	New	fersurellal	
T065	08-Jun-12	EJASE	ase06	10-Mar-2012	EJ_ASE06 Q6: should be blank if Q4=negative, otherwise it should be answered	New	fersurellal	
T065	08-Jun-12	EJASE	ase07	2	EJ_ASE07 Q7: should be blank if Q4=negative, otherwise it should be answered	New	fersurellal	

Quality Assurance & Quality Control

- QA: Strategies used **before and during data collection** to ensures the best possible data will be collected (CRF design, training staff, testing data collection tool and supporting technologies "DE system, validation tool)
- QC: processes applied **after data collection** to evaluate the quality of the collected data (data cleaning, SDV, making decisions for data issues)

Quality Assurance & Quality Control

- Implementing QA and QC procedures enhances the quality of data, minimizes errors and identifies potential problems and techniques to address them.
- QA-QC processes vary substantially in their cost and effectiveness (QC - more difficult and resource-intensive than QA. It's easier to prevent than repair problems and much cheaper in the long run).

Audit trail

- **GCP requirement:** Any change or correction to a CRF should be dated, initialed, and explained and should not obscure the original entry. That is, an audit trail should be maintained. This applies to both written and electronic changes or corrections
- Audit trail:
  - Documents all modifications to a DB
  - Is stored in a secure system
- All documentation of data changes
  - Essential study documentation
  - Is subject to audit

Monitoring data quality and safety

- Routine progress reports:
  - Recruitment report (actual vs target number of subject recruited)
  - FU report (overdue visits)
  - Data monitoring reports (number of forms received, entered, list of form errors & data errors)
  - Adverse Event Reporting
- Interim analysis: follow predefined frequency and timing
- Site monitoring visits
- Auditing
- Data Safety Monitoring Board (DSMB)

Security and confidentiality

- Keep identifying data (subject name, social security number, medical record number) in a separate place and restrict access to this data
- Make sure only subject ID links to the DB
- Data can be accessed by authorised staff *(Password protect)*
- Qualified personnel for management and modifications
- Copy of data cannot be distributed without investigator's consent

Database closure

- Proper closing a study DB:
  - Preventing inadvertent or unauthorized changes to data
  - Ensuring the integrity for the generation of results, analysis and submissions
- Process for closing the DB and conditions for re-opening the DB must be followed

Data storage

- Secure, efficient and accessible storage of clinical data is very important
- Potential of unauthorized access and data corruption during data storage and transfer are significant and must be prevented to ensure consistency of results and data quality
- Original data collected (e.g., CRFs, lab data, medical notes and e-documents) must be protected and stored in secure areas with controlled access (e.g., locks)

Data storage (cont'd)

- Store clinical data in a way that backup copies can be easily and frequently made
- *(Paper documents should be scanned soon after receiving and archived electronically, whenever possible, as the backup. E-documents are regularly backup)*
- Access permission control, especially important for a the EDC trial that has no paper backups
- Minimize opportunity for data corruption via accidental or intentional manipulation
- Use open formats for archival, storage, and transport of data (e.g., ASCII, SAS Transport, PDF, CDISC ODM Model)

Data archive

- Maintain all documents and electronic records to ensure their raw formats
- Archive clinical data and documents in a secure and stable areas (no flood, fire protected, pest control)

Components must be archived:

- Original study documents: The original and/or scanned images of all CRFs, clinical notes, lab data… DMP, data handling guidelines
- Raw data files: The final raw data preserved in the study DB format and all original data transfers in their raw format

Data archive (cont'd)

- Final data files: Preserved in a standard file format (e.g., ASCII, SAS transport, CDISC Operational Data Model)
- Audit trail
- Discrepancy management logs
- Database design specifications (metadata, validation checks)
- DB closure documentation: of each DB-lock and unlock, describing the time and conditions surrounding those procedures
- Procedural variation documentation: Memos and relevant information about any variations from SOPs or working practices

WHY - CDM?

- WHY do we need to process the research data?
- WHY do we implement edit checks and query inconsistencies?
- WHY do we need GCP-compliant CDM?

CDM - Conclusion

- CDM is really a challenge to many researchers who rarely have formal training in DM
- Implementing DM practices should take the full study life cycle into account.
- Good DM requires adopting best practices

OpenClinica

WHO online DM system

World Health Organization

Reproductive Health and Research

UNDP • UNFPA • WHO • World Bank
Special Programme of Research, Development
and Research Training in Human Reproduction

https://who.eclinicalhosting.com/OpenClinica_test

https://who.eclinicalhosting.com/OpenClinica

OpenClinica® ENTERPRISE

- 100% web-based system that is built on a modern, web-based technology architecture
- The system is always accessible to those who need to access it
- Easy to generate and extract data in numerous formats: SPSS, SAS, Excel
- Users need a simple PC and internet connection to use the system

Introduction

OpenClinica® ENTERPRISE

Open source clinical trials software for capturing and managing clinical trial data

- Software is free and available in source code form
- Free to run studies for any purpose
- It is used for private and non-commercial application
- In addition to private and non-commercial applications, AKAZA Research sells commercial services where customers can get technical supports

OpenClinica® ENTERPRISE

- **Fully complies with GCP and regulatory guidelines**
  - Keep trial data secure
  - Monitor access and data changes
  - Comprehensive auditing to any data changes
  - Control access to study information via different user roles and privileges
  - Prevent unauthorized access to data via user password
  - Ability of electronic signature
  - Daily data back-up

WHO/RHR OC system

- **OC platform is hosted and technically supported by Akaza Research**
- **WHO/RHR/SIS is responsible for:**
  - Design and configuration of the online study as well as guidelines for using the system (SOP)
  - User training (onsite and remote training) on how to employ OC most productively
  - Monitoring and managing database of the project to ensure data is of highest quality

OpenClinica - Key functions

- Submit Data, Notes & Discrepancies
- Extract Data
- Monitor and Manage Data
- Setup and Manage Study
- Administration