

Clinical Data Management

(Process and practical guide)

Dr Nguyen Thi My Huong
WHO/RHR/RCP/SIS

Training Course in Sexual and Reproductive Health Research
Geneva 2012



World Health Organization



Reproductive Health and Research



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

OUTLINE

- Clinical Data Management
- CDM – progress and practical guide
- WHO online DM system (OpenClinica)



World Health Organization



Reproductive Health and Research



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

Clinical Data Management (CDM)

- CDM - all aspects of processing the clinical data
- End result for the CDM:
 - A study database - accurate, secure, reliable and ready for analysis.
 - Timeline from data collection to analysis
- Good CDM - delivery of the quality data on-time and within the trial budget



World Health Organization



Reproductive Health and Research



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

CDM process

- CDM process - from DM group was first formed until quality DB can be delivered
- CDM process - activities:
 - Data management plan (DMP) development
 - Study setup
 - Training
 - Data collection



CDM process (cont'd)

CDM process – activities (cont'd):

- Data processing
- Monitoring data quality and data safety
- Audit trail
- Database (DB) Closure
- Data storage
- Security and confidentiality
- Data archive



Data Management Plan (DMP) development

- DMP should be developed for each study and early during the setup of the study
- Describe all the components of the DM process
- 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)

- 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 life cycle of a study, to address any updates/changes made during conduct of the study.



Study setup

Study setup includes:

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



World Health Organization



Reproductive Health and Research



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

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)

- Questions and instructions - clear and concise
- 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
- Separate CRF for each visit: SCR, ADM, FUP...
- Logically related data should be grouped together



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
- *(DOB is preferable to the age)*



CRF design (cont'd)

- Data in coded form:
 - Minimize errors
 - Reduce processing time
 - *Coded formats - multiple or single choice 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:
 - 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 spelled correctly and clinically appropriate



CRF completion guidelines (cont'd)

- Definitions for 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 CRF completion guidelines



Trial database setup

- All clinical data must be entered and stored in a computer system
- A DB - a structured set of data (rows, columns)
- (Excel spreadsheet, a Microsoft Access application, SAS tables, or a set of tables built in one of the applications such as Oracle, OpenClinica)



Trial database setup (cont'd)

- Success of a CT 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 requirement: "*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)*".



Trial database setup (cont'd)

- Output from DB design is a specification of the DB (information for each variable):
 - Name and label
 - Type (e.g., numeric, character, integer, date...)
 - Length (number of characters)
 - Definitions for all coded values
- e-CRFs (DE screens) - identical to the paper CRFs



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
 - Increases data quality
 - Greater efficiency for data cleaning
 - Identify data inconsistencies and potential errors
 - Validation check document - a living document throughout the life cycle of the study, is updated to CRF changes or errors need correcting



Validation checks (cont'd)

- **Missing values** checks: apply to critical variables (*center number, subject numbers, primary safety and efficacy variables*)
- **Range checks** – common checks, identify errors outside of the expected range
- (*expected weight for the study subject was between 40 – 80 kg*)



Validation checks (cont'd)

- Logically inconsistent checks across fields or across CRFs
- *(One question indicates that a subject is pregnant first time but the other question indicates that she already had previous abortion or live birth...)*
- Protocol violations checks: identify specific data that may be indicative of protocol violations
- *(subject's age "50" is out of the valid range "18-49" years old for eligibility criteria)*
- Checks for duplicates: detect single form being entered twice



Training

- Effective training ensures:
 - Regulatory compliance
 - Performance effectiveness
 - Job satisfaction of CDM staff
- GCP guideline states: "Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his/her respective task(s)."
- Training documentation (SOPs, guidelines)



Training (cont'd)

- Training content requirements:
 - Consistent across all training materials
 - Consistently conveyed by instructors or mentors
- Types of training vs Roles:
 - DE: handling the completed CRFs, CRF workflow and DE module in the DM system
 - Discrepancy resolution: managing data discrepancies and Discrepancy module in the DM system



Data collection

- Clinical data capture at study sites:
 - Paper CRFs (pCRFs)
 - EDC system
- 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 collection (cont'd)

Primary modes of capturing data for a CT:

- **Offline:**

- Traditional paper-based method
 - Collects clinical data at the sites
 - Sends pCRFs to DM center
- EDC system that works without Internet connection

- **Online**

- EDC method
 - Records clinical data online (eCRFs)
 - Stores data at a central server
- **Combination of offline - online** methods: that involves the use of both offline and online EDC methods



Data collection (cont'd)

CRF dispatch:

- Completed CRFs should be prepared in batches and sent to the DM center
- Ensure blinding of subject identifying information (e.g., name, address, or subject initials) submitted to the DM center.
- List of forms mailed (LOFM) should be enclosed with the batch
- Sending frequency



List of Forms Mailed

Page ____ of ____ pages

Sent on _____

Study number A65779

Study name Assessment of eligibility and follow-up care for early medical abortion

Centre number 2065

Facility number and name: 1 – MS Durban

Sending number _____

Total number of forms sent _____

- 1) Record Screening number in appropriate column when form is ready for mailing. Use a vertical arrow notation, with starting and stopping Screening numbers, to indicate a continuous range of Screening numbers for one form type.
- 2) Prepare this form in duplicate. Include the original with the data forms. Please retain the copy in your files.
- 3) Enter the total number of forms of each type in the bottom box of each column.

<u>SCR</u> Form	<u>BGI</u> Form	<u>ASE_CHW</u> Form	<u>ASE_Clinician</u> Form	<u>EXE</u> Form	<u>ASF_CHW</u> Form	<u>ASF_Clinician</u> Form	<u>ENF</u> Form
Total number of forms:							

FOR ICT USE ONLY

I acknowledge the receipt of this mailing.

Queries: _____
Date: _____
Signed: _____

Data processing

Data processing workflow at the DM center:

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



World Health Organization



Reproductive Health and Research



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

Data receipt

- Data receipt vary across the clinical research that may be received through:
 - Fax transmissions
 - Regular mail
 - Express delivery companies with tracking ability
 - Private couriers
 - Hand delivery by monitors
 - 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, including mandatory and optional CRFs
- CRF logging can be:
 - Manually by DM staff who registers all the CRFs have been received
 - Automatically when the data from the CRFs are entered in the DM system
- Missing CRF should be specified (LOFS, SFR)



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

- Manual review of CRFs
 - For all CRFs before DE
 - Detect data errors, frequently encountered problems
- Query sheet for data errors



Data coding

- Medical coding - properly classify the medical terminologies
- Data coding methods:
 - Manual coding using common coding dictionaries (*WHO Drug dictionary, Medical Dictionary for Regulatory Activities "MedDRA", ICD-10, ICPM*)
 - Automated coding



Data Entry systems

Local DE system:

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

Central DE system:

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

Web-based DE system:

- Software requirements (Internet Explorer)
- No specific hardware requirements
- Require 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	fensurellal	
T065	08-Jun-12	EJASE	ase03	11-Jun-2011	EJ_ASE03_2 Q3: Date of assessment should be the same date of SCR	New	fensurellal	
T065	08-Jun-12	EJASE	ase05	2	EJ_ASE05 Q5: should be blank if Q4=negative, otherwise it should be answered	New	fensurellal	
T065	08-Jun-12	EJASE	ase06	01-Feb-2012	EJ_ASE06_1 Q6: LMP should indicate pregnancy of first trimester	New	fensurellal	
T065	08-Jun-12	EJASE	ase06	10-Mar-2012	EJ_ASE06 Q6: should be blank if Q4=negative, otherwise it should be answered	New	fensurellal	
T065	08-Jun-12	EJASE	ase07	2	EJ_ASE07 Q7: should be blank if Q4=negative, otherwise it should be answered	New	fensurellal	

Monitoring data quality

- **Progress reports:**
 - Frequency of reports
 - Recruitment report (actual vs target number of subject recruited)
 - FU report (overdue visits)
 - Data monitoring reports (number of forms received, list of form errors, list of data errors)
- **Interim analysis:** follow the frequency and timing specified in the protocol
- **Site monitoring visits**
- **Training and re-training staff**
- **Auditing**



Monitoring safety

- Adverse Event Reporting
- Data Safety Monitoring Board (DSMB)



World Health Organization



Reproductive Health and Research



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

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 open format files in a secure system
- All documentation of data changes
 - Essential study documentation
 - Is subject to audit



Audit trail (cont'd)

- **Audit trail:**
 - **Is triggered by the initial DE**
 - **Captures any changes to the entry**
 - **Keeps both previous and current values**
 - **Remains unmodified and intact**



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
- Prior to DB closure, ensure:
 - All data have been processed
 - Quality level has been assessed
 - Relevant study personnel have been notified

All these will decrease the need for unlocking the DB
- At final DB closure, ensure:
 - Edit permissions are removed from all personnel



Database closure (cont'd)

- Process for closing the DB and conditions for re-opening the DB must be followed

Tasks needed for DB closure:

- All tasks defined in the DM plan are completed
- All conditions have been met
- Ensure all site monitoring activities are complete prior to the DB lock
- All data have been received, processed and validated for completeness and consistency



Database closure (cont'd)

- Required coding has been completed
- All outstanding queries have been resolved
- *(It may be acceptable to lock DB with open queries for noncritical data)*
- Ensure quality audit of the data has occurred and all required source document verification has been completed
- All study documentation are updated and stored according to SOPs



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)



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
- Qualified personnel for management and modifications
- Copy of data cannot be distributed without investigator's consent



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



Data archive (cont'd)

- Procedural variation documentation: Memos and relevant information about any variations from SOPs or working practices
- Site copies of data should be read-only datasets delivered on CD-ROM or a similar storage medium

