Laboratory Data Management

September, 2010 at SMMU, Shanghai

Agenda

1. Overview
2. Types of Laboratories
3. Lab Data Management
4. Standards
5. Key Messages

1. Overview

- **The vast majority of clinical studies use laboratory to provide:**
  - Safety Data
    - Hematology, Chemistry, Urinalysis
  - Efficacy Data
    - Cholesterol level for Hyperlipidemia patient
    - Plasma Glucose, for diabetes patient
  - Special Data
    - PK / PD data
    - Genomic data
    - Biomarkers

1. Overview

There are many challenges in the data management task, but the ultimate challenge is managing the Lab Data.

- Laboratory data derived from tests of both drug safety and efficacy have been estimated to be as much as 60-80% of the data generated during the conduct of clinical trials
- Lab results tell the physician significant information about the body that are directly related to the safety and well-being of the subject
- Lab data are more difficult to interpret if one does not have clinical training, and what is appropriate to query is less clear
- If involving central lab, data transfer and loading, it requires data manager having more understanding in database structure and data transfer technical and processes

Agenda

1. Overview
2. Types of Laboratories
3. Lab Data Management
4. Standards
5. Key Messages

2.1 Types of Laboratories

Central Lab	**A central lab** processes samples from multiple clinical sites or studies at one central location. It often supports multicenter and international studies.
Local Lab	**A local lab** is in close proximity to individual clinical study sites or patients and are most often used when timely results are needed.
Virtual Central Lab	**A virtual central lab** is typically a group of labs located throughout the world that are under the umbrella of one company. It is based on a central calibration that runs in parallel with samples from all labs in a clinical study.

2.1 Types of Laboratories

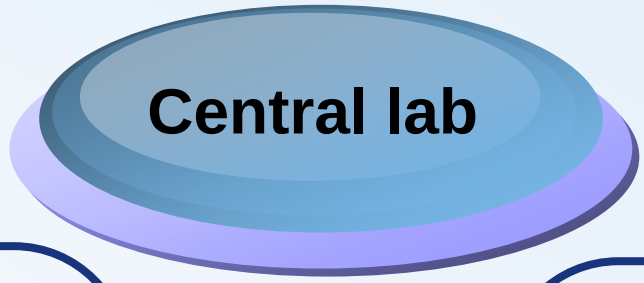
Core Labs

A core lab specialize in a particular therapeutic area or body system. Examples include stem cell core lab, ECG core lab, imaging core lab, cardiovascular core lab, etc.

Specialty Lab

A specialty lab is used to analyze samples or run assays for non-traditional tests, which typically take a considerable amount of time and effort to produce. Examples include biomarkers, genetic testing, isolation of cancer genes, etc.

2.2 Pros and Cons of Different Labs



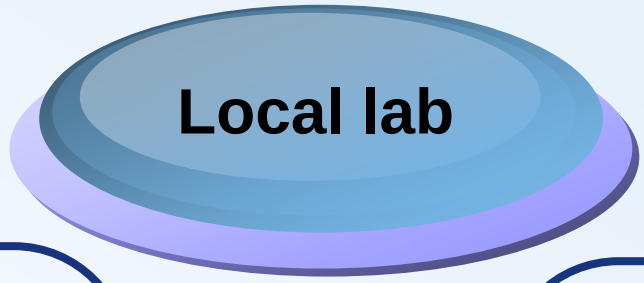
Pros

1. Uses one set of analytical equipment, methodologies, kits and reagents.
2. Provides training and instructions for collection and shipping of samples.
3. Standardized results from one set of reference ranges and units.
4. Typically transfer data electronically from lab to the sponsor.

Cons

1. Very expensive due to logistical support and sample shipping.
2. The turnaround time needed to receive central lab test data may be too long when immediate results are needed.

2.2 Pros and Cons of Different Labs



Pros

1. Lower costs and shorter turnaround time due to not having to ship samples.
2. Local lab experience with processing samples from their subject population.
3. Quick availability of test data, especially where the results could be the deciding factor on screening, dosing, etc.

Cons

1. Greater potential for errors due to paper-based data transcription.
2. Differences between reference ranges from one lab to another.
3. Variability in the methods used to perform tests.
4. Reference ranges may be difficult to obtain.
5. Time-consuming in verifying reference ranges.

2.2 Pros and Cons of Different Labs

Pros

1. Reduced shipping costs.
2. Simpler data processing due to having a central calibrator.
3. Standardized results from one set of reference ranges and units.

Cons

1. Requires detailed process and quality control measures to ensure lab results are reproducible with minimal variance from site to site.

2.2 Pros and Cons of Different Labs

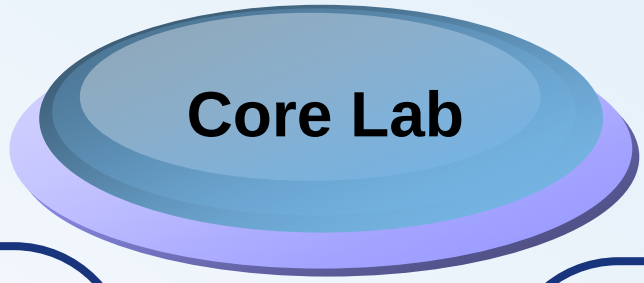
Pros

1. Highly experienced and qualified for performing specialty tests.

Cons

1. Many specialty tests requires more time to generate test results.

2.2 Pros and Cons of Different Labs



Pros

1. More focused quality control, more accurate results and a higher degree of standardization and specialization within a designated area.

Cons

1. Additional time may be incurred for centralized processing.

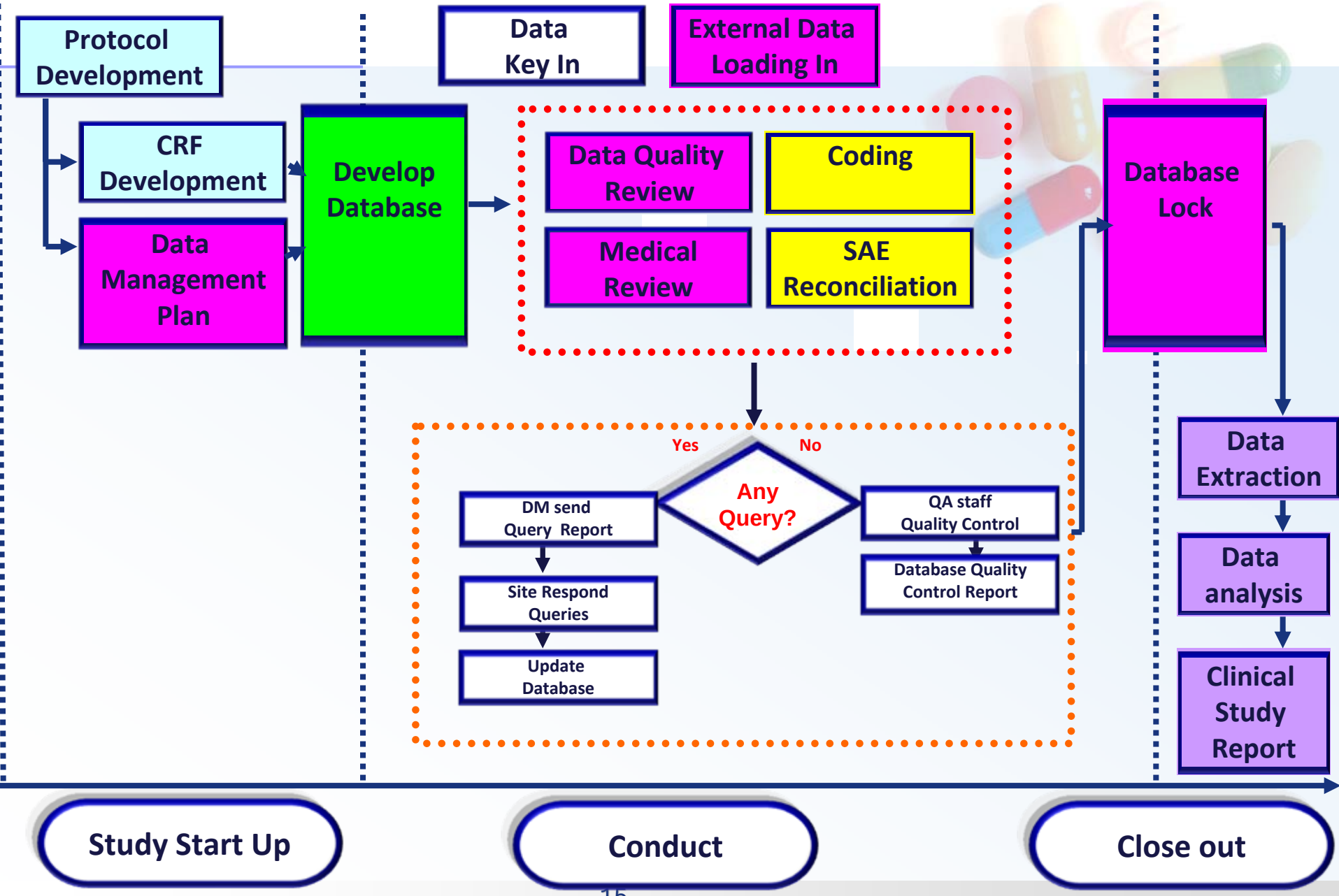
2.3 Central Lab Vs. Local Lab

	Central Lab	Local Lab
Sample Handling	All samples shipped to the central lab for analysis and reporting	Samples are analyzed at the local lab. No shipping
Cost	Expensive	Relatively inexpensive
Results availability	Need extra time due to sample shipping	Quick
Reference Range and Units	One set of reference range and units	Multiple lab-specific reference ranges and units
Data handling	Usually data can be electronically transferred and uploaded into the database	Data needs to be manually entered into the database
Data Quality	High quality of data	Prone for transcription error
Data Analysis	Units conversion not needed and results are comparable	Need units conversion

Agenda

1. Overview
2. Types of Laboratories
3. Lab Data Management
4. Standards
5. Key Messages

DM Flow


Protocol Development
CRF Development
Data Management Plan
Develop Database
Data Key In
External Data Loading In
Data Quality Review
Coding
Medical Review
SAE Reconciliation
Database Lock
Yes
No
Any Query?
DM send Query Report
Site Respond Queries
Update Database
QA staff Quality Control
Database Quality Control Report
Data Extraction
Data analysis
Clinical Study Report
Study Start Up
Conduct
Close out

3.1 Central Lab Data Management

Start Up Phase

- Protocol Development
  - What tests is required and when (Schedule of Assessment
  - Central lab or local lab
- Case Report Design
  - Only a few questions are required
    - Requisition Number (Accession No., Sample ID)
    - Sample Collection Date
    - Sample Collection Time
    - Results are NOT needed to collect

CRF Example _ Central Lab

Blood sampling record (for pharmacokinetics)

Scheduled sampling time	Sample number	Sample collection date dd/mm/yy	Sample collection time[1] (24 hr clock) h:min	Remarks[2]
predose	101			
3 h post-dose	102			
6 h post-dose	103			
8 h post-dose	104			
10 h post-dose	105			
12 h post-dose	106			

3.1 Central Lab Data Management

Start Up Phase

- Determine if data should be loaded into the clinical database
- Identify and involve lab vendor as early in the process as possible
- Identify key individuals for communication and follow through
- Collect the reference range and units from the central lab

3.1 Central Lab Data Management

Start Up Phase

- Establish the procedures for collecting, transferring, loading, validating and editing external data and document.
  - Data Transfer Agreement (DTA)
  - File Format Specifications (FFS)
  - Data Cleaning Plan
- Perform test transfer successfully before production

3.1.1 Data Transfer Agreement (DTA)

- **DTA defines:**
  - The format of files (Excel spreadsheet, ASCII, SAS dataset, text file, etc)
  - Frequency of data transfer
  - File naming conventions
  - Encryption and method of transfer (password encrypted email attachment, CD, secure FTP, etc)
  - Type of transfer (accumulative vs. incremental)
  - Primary and secondary contacts of senders
  - Primary and secondary contacts of recipients
  - Transfer confirmation agreement

3.1.2 File Format Specifications (FFS)

- **FFS defines:**
  - File naming conventions
  - File format
    - Excel spreadsheet, ASCII, SAS dataset, text file, etc
    - The delimiter ('|', ';', '||' or blank)
    - List of the variables and their order
    - Type and length of each variable (char, numeric, date)
    - Column position and field justification (for ASCII files)
    - Key variables, which uniquely describe each sample record (study id, site/inv id, subjid, visit, sample id)

3.1.2 File Format Specifications (FFS)

- **FFS defines:**
  - Contents
    - Test name, long and short (BUN vs. Urea)
    - Units (conventional vs. SI units)
    - Date and time format (YYYYMMMDD, DDMONYYYY, HH:MM 24hr)
    - Handling of >, <, signs
    - Special Codes (NEG, POS, NOS, etc)
    - Comments regarding the condition of the sample

3.1.2 File Format Specifications (FFS)

- **FFS defines:**
  - Unexpected/Unscheduled lab data
  - Procedures for Database Updates
    - New data
    - Updates to already loaded data
  - Procedures for ensuring blinding

3.1.3 Data Cleaning Plan

- **Data checks that can be applied include:**
  - Reconciliation between CRF data and loaded lab data
    - Demographic (subj. ID, subj. initials, date of birth, sex)
    - Sample collection (visit name, collection date and time, sample id)
  - Missing individual test result within a panel
    - Missing RBC result in Hematology test at screening visit
  - Missing / invalid reference range and units
  - Out-of-range values (against medical history or adverse event)
  - Inclusion/exclusion criteria involving lab data
  - Duplication test results

3.1 Central Lab Data Management

Conduct Phase

- Perform production transfers as per DTA
- Resolve loading problems with the lab vendor, if any
- Review and clean the external lab data on an ongoing basis according to data cleaning plan
- Resolve data discrepancies with the site and the lab vendor
- Maintain the reference range and units and perform change control

3.1 Central Lab Data Management

Close-down Phase

- Perform the last production transfer which contains all results of all lab samples
- Resolve loading problems with the lab vendor, if any
- Review and clean all the external lab data
- Resolve all data discrepancies
- Release blinded data, if appropriate
- File storage and archiving

3.2 Local Lab Data Management

Start Up Phase

- Protocol Development
  - What tests is required and when (Schedule of Assessment
  - Central lab or local lab
- Case Report Design
  - Requisition Number (Accession No., Sample ID)
  - Sample Collection Date
  - Sample Collection Time
  - Results are collected on CRF

CRF Example _ Local Lab

Laboratory analysis

Sample collection date	dd mm yy
Hematology	Result[2]
Haemoglobin	
Hematocrit	
Erythrocytes	
Leukocytes (total)	
Neutrophils	
Basophils	
Eosinophils	
Lymphocytes	
Monocytes	

3.2 Local Lab Data Management

Start Up Phase

- Collect the reference range and units from the individual local labs
- Verify the reference ranges and units
- DTA and FFS are not needed but data cleaning plan is still required

Data checks that can be applied include:

- Reconciliation between CRF data and loaded lab data NOT needed
- Missing individual test result within a panel
  - Missing RBC result in Hematology test at screening visit
- Missing / invalid reference range and units
- Out-of-range values (against medical history or adverse event)
- Inclusion/exclusion criteria involving lab data
- Duplication test results

3.2 Local Lab Data Management

Conduct Phase

- Resolve data discrepancies with the site
- Maintain the reference range and units and perform change control

Close-down Phase

- Review and clean all the lab data
- Resolve all data discrepancies

Agenda

1. Overview
2. Types of Laboratories
3. Lab Data Management
4. Standards
5. Key Messages

4. Standards in Lab Data Collection & Interchange

4.1 Test Name and Units

Test Name

- WBC vs. White Blood Cell; BUN vs. Urea; AST vs. Aspartate transaminase vs. SGOT
- Standards
  - CDISC standard - Terminology (临床数据交换标准协会 – 术语)
  - LOINC Codes (观测指标标识符逻辑命名与编码系统)
  - SNOMED CT (医学系统命名法—临床术语)

CDISC – Clinical Data Interchange Standards Consortium

LOINC – Logical Observation Identifies Names and Codes

SNOMED CT – Systematized Nomenclature of Medicine – Clinical Terms

CDISC Terminology (Laboratory Data) - example

CDISC PT (LBTESTCD)	**Long Name /Description** (LBTEST)	**Definition**	**Synonyms**
BILI	Bilirubin	A measurement of the total bilirubin in a biological specimen.	Total Bilirubin
BILIND	Indirect Bilirubin	A measurement of the unconjugated or non-water-soluble bilirubin in a biological specimen.	
BUN	Blood Urea Nitrogen	A measurement of the urea nitrogen in a blood specimen.	

LOINC Codes - example

LOINC Code	Component	Property	Time Aspect	System	Scale	Method	LOINC Short name
14631-6	BILIRUBIN	SCNC	PT	SER/PLAS	QN		Bilirub SerPl-sCnc
1974-5	BILIRUBIN	MCNC	PT	FLU	QN		Bilirub Fld-mCnc
1977-8	BILIRUBIN	ACNC	PT	UR	ORD		Bilirub Ur Ql

Component – analyte being measured;
Time Aspect – time of measurement;
Scale – Quantitative, qualitative, ordinal;
Property observed –
System – specimen
Method – where applicable

4.1 Test Name and Units

Test Units

- Reported Units
- Conventional Units (typically based on US measuring methods)
- SI Units (le Systeme International d'Unites)

For example, reference and units for Potassium (血钾）

- 3.5 - 5.0 mmol/L
- 3.5 - 5.0 mEq/L
- 13.7 - 19.5 mg/dL

4.2 CRF Design Standards (CDISC CDASH)

Central processing

- Lab status (whether or not lab sample was collected)
- Date of collection
- Time of collection
- Panel name (e.g. Chemistry, Hematology, Urinalysis)
- Planned time point
- Protocol-defined testing conditions met (e.g Fasting)
- Accession number

Only the highly-recommended variables are listed

CDISC – Clinical Data Interchange Standards Consortium

CDASH – Clinical Data Acquisition Standards Harmonisation

4.2 CRF Design Standards (CDISC CDASH)

Local processing

- Lab status (whether or not lab sample was collected)
- Date of collection
- Time of collection
- Panel name (e.g. Chemistry, Hematology, Urinalysis)
- Planned time point
- Protocol-defined testing conditions met (e.g Fasting)
- Test name
- Test result
- Units

Only the highly-recommended variables are listed

4.3 Data Structure Standards (CDISC SDTM)

- **LB domain**
  - **One record per lab test per time point per visit per subject**
    - **See the example in the next two slides**
  - **For more details, please visit CDISC website:**

    http://www.cdisc.org/sdtm

SDTM – Study Data Tabulation Model

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req	SDTMIG 2.4.4
DOMAIN	Domain Abbreviation	Char	LB	Identifier	Two-character abbreviation for the domain.	Req	SDTMIG 2.4.4, SDTMIG 4.1.2.2, SDTMIG Appendix C2
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req	SDTMIG 2.4.4, SDTMIG 4.1.2.3

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core	References
LBSEQ	Sequence Number	Num		Identifier	Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.	Req	SDTMIG 2.4.4
LBGRPID	Group ID	Char		Identifier	Used to tie together a block of related records in a single domain for a subject.	Perm	SDTMIG 2.4.4, SDTMIG 4.1.2.6
LBORRES	Result or Finding in Original Units	Char		Result Qualifier	Result of the measurement or finding as originally received or collected.	Exp	SDTMIG 2.4.3, SDTMIG 4.1.5.1

4.4 Data Interchange Standards

- It is estimated that the cost to the industry per year simply for laboratory data interchange itself is at least $150m and that between approximately 30% and 60% of that cost could be saved from the use of a standard
- The existing standard models for the interchange of laboratory data includes:
  - ASTM International
  - Health Level 7 (HL7)
  - ACDM (the Associate for Clinical Data Management)
  - X12

4.4 Data Interchange Standards

CDISC standard – Laboratory Data Model

- The default implementation of the LAB Model is bar delimited ASCII
- The superset of data fields are in 12 levels;
  - Good Transmission Practice
  - Study
  - Site
  - Investigator
  - Subject
  - Visit
  - Accession
  - Record Extension Type
  - Base Specimen
  - Base Battery
  - Base Test
  - Base Result

Example of lab data file

4.4 Data Interchange Standards

- CDISC standard – ODM
  - A vendor neutral, platform independent format for interchange and archive of clinical study data
  - The model includes the clinical data along with its associated metadata, administrative data, reference data and audit information

ODM – Operational Data Model

Take-home Key Messages

- Local lab data and central lab data are handled in different ways (CRF design, data collection and data cleaning, etc)
- The procedures for collecting, transferring, loading, validating and editing central lab data apply for other external data as long as they need to be loaded into clinical data base. The following documents should be developed before data are received:
  - Data Transfer Agreement
  - File Format Specifications
  - Data Cleaning Plan
- Understanding and use of standards in data collection, processing and transferring will greatly improve the quality, integrity and plausibility of the external data.

The 2nd Clinical Data Management Training

Thank You !

Good Transmission Practice

Variable	Description
Model Version	The version of the CDISC Laboratory Data Interchange Standard model indicating by definition which fields are contained within it.
File Creation Date and Time	The local date and time the data file was created at the central laboratory. This includes a Universal Time Offset plus/minus hours and minutes.
Transmission Source ID	The ID of the organization that is the source of the data transmission.
Transmission Source Name	The Name of the organization that is the source of the data transmission.

Subject

Variable	Description
Screen ID or Number	The ID of the subject before randomization
Subject ID or Number	The ID of the subject after randomization
Spare subject level ID or Number	Spare subject level identifier. (For use with original screen IDs in cases where re-screening with new numbers is allowed, for example).
Subject Initials	The initials of the subject.
Subject Sex	The sex of the subject.
Subject Sex Code List ID	If utilized, the code list identifier and version number for the Subject Sex Code.
Subject Date Of Birth	The date of birth of the subject
Subject Race	The biological race of the subject.
Subject Race Code List ID	If utilized, the code list identifier and version number for the Subject Race code.