# Clinical Trials Data Management Systems: What a Project Manager Needs to Know

## Katherine L. Monti, Ph.D.
Director, Massachusetts Office, Rho, Inc.
Adjunct, Biostatistics, University of North Carolina

## Nancy Yovetich, Ph.D.
Director, Project Operations, Rho, Inc.

# 1. Introduction

Goals for this session

- Present an overview of clinical data management, including those aspects most germane to the project manager.
    - Learn what happens to data after data collection
    - Become familiar with the key parameters for data management planning and reporting
    - Gain insight into data management challenges to aid in troubleshooting
    - Identify efficiencies that lead to $$$ savings

# 1. Introduction

1. Introduction – Goals and outline

2. DM activities that occur prior to data capture

3. Details of a data management system: paper data

4. Details of a data management system: eData

5. From clean data to database lock

# 1. Introduction

6. DM planning

7. DM reporting

8. Troubleshooting DM challenges

9. Costs and cost savings

10. Recap

# 2. Before Data Capture: General Issues

- Prior to actual data collection, decisions are made that affect data management.
    - Data-capture method and data management team
    - CRF design
    - Data collection training
    - Regulatory issues
    - Database development

# 2.1 Getting Started

- Selection of data capture method:
  Paper or electronic?

- Selection of the data management team:
  - Quality
  - Timeliness
  - Professionalism
    - *"We'll just hire some high school students to key in the data."*

# 2.2 CRF Design

- CRF design can affect the quality of the data and even the interpretability of the results.

- Forms should comply with the following conventions:
  - Adequate space to write
  - Clear instructions
  - Clearly-specified, unambiguous questions
  - Forced choice responses (unless part of a skip pattern)
  - Categorical data should be all inclusive

# 2.2 CRF Design

# 2.2 CRF Design

- Poorly designed AE form:
  - Attempt to squeeze too much information on one page.
  - Form is too complicated:

    If AE is ongoing, "Write 'Ongoing' in

    Stop Date field."

# 2.2 CRF Design

- In this study, the AE stop date field alone resulted in over 12% of the queries, an average of about 1 query per patient.

# 2.2 CRF Design

- Dates in general are a problem
  - Specify the format, such as ddMMMyyyy (09APR2003).
  - Allow adequate space to write the numbers legibly.
  - Allow adequate space for corrections to be entered on the page in the event of a recording error.

# 2.2 CRF Design



**Concomitant Treatments** *ctfi*

**Concomitant medications**

CNON

MEDNONE $1 = 1 if "None" or "No" recorded    PROPHY    ONGOING MEDSPOT

| Medication | Route/formul. | Dosing scheme | Indication | Prophyl. check, if | MEDSTOT Start date | check if ongoing | MEDSP or Stop date |
|---|---|---|---|---|---|---|---|
| MED $80 | ROUTE | DOSE | IND | O | MEDSTMO/DA/YR | O | MEDSP mo/DA/YR |
| | $20 | $20 | $20 | O | | O | |
| | | | | O | | O | |

# 2.2 CRF Design

- In this study, dates have accounted for 3970 queries so far, 40% of the total queries.

# 2.2 CRF Design

- Avoid Type III errors

  - Use check boxes

  - Mind the wording

# 2.2 CRF Design

- Check boxes assure that the question you intended to ask is the same as the question actually answered.

  EXAMPLE:
  - PAIN RIGHT NOW: _____
  - PAIN RIGHT NOW: ___ Yes ___No
  - PAIN RIGHT NOW:          ___ None
                             ___ Mild
                             ___ Moderate
                             ___ Severe

# 2.2 CRF Design

- EXAMPLE (from the demography page):

SEX:        ___ Male
            ___ Female

SEX: _____

# 2.2 CRF Design

- If you ask an ambiguous question, you may get the wrong answer!

# 2.2 CRF Design

- Question to a woman in labor:

  Is this your first baby?

  *Possible meanings:*

  Is this your first pregnancy?

  Is this your first live birth?

# 2.2 CRF Design

- Consider the following study design for a drug for chronic pain:
    - Study 1 (placebo or active drug)
    - Between-study period of 2-50 weeks; rescue meds allowed
    - Follow-up Study 2 (open-label active drug)
- In Study 2, patients are asked "Compare this drug to your previous drug."
- Which "previous drug"? The drug in Study 1 or the between-study drug?

# 2.2 CRF Design

- Experienced project managers often play a role in the design or review of the CRF. Careful review of the form can help those who later fill out the form, reduce the number of queries that will come back, and save time ($$$) for the clinical staff and data management.

# 2.3 Data Collection Training

- Well-trained study coordinators generate better quality data, resulting in fewer queries.

- Problems that frequently arise in training:
  - Person filling out the forms did not attend training session.
  - Personnel turn-over occurs, so the coordinator never received adequate training.
  - Lag time from training to first patient recruited may be long.
  - Lag time between patient enrollments may be long.
  - The training session may be adequate for highly trained coordinator but inadequate for less experienced personnel.

# 2.3 Data Collection Training

- Monitors have the opportunity to detect training issues and address them.
  - Look for common problems on the completed forms (at site visits).
  - Look for recurring types of queries.

- $ The sooner a data capture problem is detected and corrected…
  - the less money is wasted generating queries
  - the higher the final quality of the data

# 2.4 Regulatory Issues

- Protection of subject confidentiality
- Error correction rules for CRFs should be set up using GCP
- Computer systems must comply with 21 CFR 11
- Computer system must have an audit trail
- Signature authority for original data and all changes or corrections
- SOPs have to be in place and followed

# 2.5 Database Development

*Data are the information of a clinical trial.*

…. On to database development

# 3. The Paper Trail: Processing Data from Paper CRFs

- Objective:  Introduce CDM procedures by tracing the flow of data through a system.

- We start with data captured on traditional paper CRFs.

  - People tend to be more familiar with these steps than with EDC steps.

  - EDC steps are compressed: multiple "steps" occur at essentially the same time.  It's easier to understand the steps if first seen in the context of traditional paper CRFs.

# 3.1 Source Data Capture: Record Information in Medical Records

- "Data capture" means different things to different people.

- "Source data capture" refers to the event when data are initially recorded, "at the source."

  - Most clinical trials data are initially captured into a site's medical records that exist separately from the clinical trial data forms.

27

# 3.2 (CRF) Data Capture: Transcribe data to the CRF

- The study coordinator takes the CRF to "the source"
  - Patient
  - Patient's chart or other medical records, etc.

  and *transcribes* data onto the paper CRF.
  - This is just the first of many transcription processes in clinical data management.

# Important Data Management Principle

- Every transcription process introduces errors into the data.

- The question is not whether a transcription process introduces errors, but what is the transcription process's error rate.

# 3.2 (CRF) Data Capture: Transcribe data to the CRF

- CRF data capture is usually an unverified transcription: only the person doing the work checks the work for errors.

- Clinical monitors (CRAs) may perform partial verification later.

# Important Data Management Principle

- Unverified transcription processes generally have higher error rates than verified transcription processes.

  – Simple reason:  verification catches and corrects some of the transcription errors.

# 3.3 Monitor Review/Correction: Compare CRF to Source Docs…

- The CRA reviews the CRF, checking for legibility, medical/scientific sensibility.

- The CRA compares CRF data to data on source documents and resolves problems as they are detected.

- The CRA, in this step, plays a crucial role in the capture of accurate, scientifically valid data.

# Important Data Management Principle

- Earlier error detection typically results in:
    - Lower final error rates, and
    - Lower error detection/error correction costs.

# 3.4 Obtain Signatures Required by Regulations

- FDA regulations and Good Clinical Practice guidelines require that the site's investigator or authorized representative sign the CRFs.

# 3.5 CRF Harvest, Transmission to the Data Management Center

- Paper CRFs must be "transmitted" from the clinical site to the Data Management Center (DMC).

- Typical method:  CRA "harvests" CRFs, carries or sends them to DMC.
  - Sometimes CRFs make intermediate stops, slowing the process.

- Other options:
  - Site express mails CRFs directly to DMC.
  - Site faxes CRFs to DMC (before or after monitoring).

# 3.6 Login:
# Inventory CRF Pages at DMC

- The first thing the DMC does with newly-arrived forms:  log each page in to the DMS.

  – Each page is inventoried.

  – The DM system now "knows," for the first time, that these CRF pages exist and are at the DMC.

  – Problems are detected:

    - Missing pages
    - Duplicate pages (same identifying fields)

    and Manual Queries are generated.

  – Henceforth, the DM system can track the progress of data through the system.

# 3.7 Utilize or Create CRF Images

- When faxes or scanned images are received (rather than paper), some DMCs store the images in a CRF image database.

- When paper CRFs are received, some DMCs scan CRF pages and store the images.

- Subsequent processing steps may utilize images rather than pieces of paper.

# 3.8 Pre-Entry Review

- Pre-entry Review: an intelligent, well-trained human examines the data for "anomalies."
  - Legibility
  - Failure to follow instructions
  - Inappropriately missing data
  - Scientifically invalid data (systolic BP = 20)
  - Inconsistent data
- Issue: Why not let the computer find the errors?
- This step generates Manual Queries.

# 3.9 Data Entry and Verification

- Data entry: independent double entry
  - One person keys data into computer.
  - Second person keys data into computer, independently.
  - Computer compares results, reports discrepancies.
  - Third person (the "referee") resolves discrepancies.
- Variations:
  - Single entry
  - Key and (dependent) verify

# 3.10 "Committing" Data: Regulatory Stature

- The goal of data entry is to create digital data that match entries on the CRF.

- During the data entry process, data may be modified as necessary in order to correct data entry errors.

- Once the data entry process is complete, including corrections of data entry errors, the data are "committed" to the database.

- After this point data modifications must be made in compliance with federal regulations.

# 3.11 Error Detection (1): Univariate

- "Univariate": Computer checks one data field at a a time.

  - *Valid value test*: Check to see if a data value is one of the valid values for a field.

    - E.g.: "SEX" must take one of the values: F, M.
    - Really bad idea: code SEX numerically.

  - *Valid range test*: Check to see if a data value is within a valid range for a field.

    - E.g.: 17 < AGE (yrs) < 61.

# 3.11 Error Detection (1): Soft and Hard Tests

- A "Hard" test is used for impossible data values.

  - E.g., HEIGHT must be greater than 0.

- A "Soft" test is used for improbable data values.

  - E.g., for males, HEIGHT < 78" (=6'6") .

  - Yes, there are males taller than 6'6", but they are more rare than data errors, i.e., a height of 79" (=6'7") for a male is more likely a data error than a correct value.

# 3.11 Error Detection (2): Multivariate Tests

- "Multivariate": Computer checks two or more data fields for inconsistency.
  - E.g.: SYSBP < DIASBP is inconsistent.
  - E.g.: MON=FEB, DAY=29, YR=2002 inconsistent
- First, it is advantageous to require that values have passed univariate tests prior to subjecting them to multivariate tests.
  - E.g., if SYSBP = 15 and DIASBP=75, why query the inconsistency, SYSBP < DIASBP, until the value of SYSBP has been corrected?
  - This prevents creating a "query cascade."

# 3.11 Error Detection (2): Multivariate Tests, cont'd

- Typically, much more effort is required to install a multivariate test into a DMS than a univariate test.

- However, if multivariate tests are not performed, statisticians tend to stumble on the data inconsistencies at some later time.

# 3.11 Error Detection

- When a data field passes a test, the system sets a data flag ("data status flag," or "data quality flag") that documents this fact.

- A data field that fails an error detection test may actually be correct.

  - Goal: challenge (query) data values that are more likely errors than correct values.

- After running the error detection procedure, a procedure is run to generate queries for data values that failed one or more error detection tests.

# 3.11 Error Detection: When is it done?

- Some error detection tests are performed at data entry time.
  - If a value of HEIGHT > 77" (=6'5") is keyed, the operator may be notified and given a opportunity to change the value.
    - This is not "an event with regulatory stature," such as changing a data value on a paper CRF. This is viewed as correcting an erroneous keystroke in order to make the data in the computer match the data on the CRF.
- Some error detection tests are performed later, especially multivariate tests.
  - Multivariate tests usually require that all fields have previously passed univariate tests.

46

# 3.12  Error Correction: Query Generation and Management

Basic procedure for error correction:

- Report each putative data error to the site.
- The "report" is called:
  - A Query
  - A DCF (Data Clarification Form)
  - Other names (some unprintable)
- The site "resolves" each query.
  - Change incorrect data values.
  - Confirm correctness of correct, but improbable, data values.
- The Query is returned to the DMC.

# 3.12  Error Correction: Query Generation and Management

Additional Information

- A query has regulatory stature.

  - Each generated query must be tracked and ultimately accounted for.

  - Like a CRF, a query must bear an authorized signature.

  - This applies to both manually- and computer-generated queries.

# 3.12  Error Correction: Query Generation and Management

Additional Information

- CRAs often play an important role in transmitting queries:

  DMC ⟵⟶ clinical sites.

- CRAs sometimes perform query resolution.

# 3.12  Error Correction: Query Generation and Management

An interesting paradox:

- A query can introduce errors into the database:

  – A correct (but improbable) data value can be changed to an incorrect value.

  – One can replace the wrong data value.

  – An incorrect data value can be confirmed as being correct.

- Query processing is error prone:  the query processing error rate is greater than routine data processing error rate.

# 3.13 Wrap-Up

- There are other topics, such as quality control, quality assurance, other data sources (e.g., lab data), and database lock.

- These topics are essentially similar for paper CRF and EDC systems:  we'll cover them later.

# 4. The Electronic Trail: Electronic Data Capture (EDC)

- EDC is currently gaining in popularity, but like everything else, it has pros and cons.

# 4. The Electronic Trail: Electronic Data Capture (EDC)

- EDC isn't really new – we did our first EDC project for the Kidney Transplant Histocompatibility Study in 1972-3.
  - Major finding:  MDs, nurses, clinical site study coordinators did not like being data entry operators.
  - We began using EDC routinely in 1983, with the advent of IBM PC computers.
    - Primarily for NIH-sponsored clinical trials.
- Let's discuss EDC operations, and compare to processing traditional paper CRFs.

# Important Data Management Principle

- Unverified transcription processes generally have higher error rates than verified transcription processes.

  – Simple reason: verification catches some of the transcription errors.

# 4.1 Electronic Source Data Capture

- "Source data capture" refers to the initial recording of data, "at the source."

- Typically with EDC, some source data are captured electronically:

  – While interviewing the patient, the study coordinator keys data directly into a computer using EDC software.

  – A computerized instrument captures data.

  – A study coordinator keys data directly from an instrument (e.g., thermometer) into a computer – this is unusual.

# 4.1 Electronic Source Data Capture: Data Entry

- The study coordinator keys "Key [or ID] Fields" that identify the data to be keyed.
  - E.g.: Study ID, Site ID, Patient ID, type of form (e.g., Demographics), visit number.
- The EDC system presents a screen that resembles a CRF page.
- The study coordinator keys the data.
- The system performs error detection steps as the data are keyed.
  - We will discuss error detection/correction later.

# 4.2 EDC Directly from Clinic Records

- Typically, some source data are captured from clinic records directly into the EDC system:

  - The study coordinator may take a laptop to the records or bring the records to a computer to key in the data.

  - This may be logistically difficult if the EDC system requires an internet connection during keying.

# 4.3 Source Data Capture via a Workbook

- Most data are initially captured into a clinic's (or site's) medical records that exist separately from the EDC system.

- The study coordinator takes a "Workbook" – essentially a paper CRF, to "the source"

  – Patient's chart or other medical records, etc.

  and *transcribes* data into the workbook.

  – The process is identical to using paper CRFs.

# 4.3 Source Data Capture via a Workbook: Data Entry

- Workbook data must be keyed (another transcription) into the EDC system by site personnel.

  – This is also called "Remote Data Entry."

- The procedure is identical to data entry described above, except the study coordinator is keying from a workbook.

# 4.4 Transmit Data to the Central Database/ Data Management Center ("DMC")

- Local data entry: data are entered directly into the site's database, at the site, and subsequently uploaded to the central database.

  – Upload is typically via the internet.

- Web data entry: data are entered directly into the central database.

  – During data entry the computer is connected via the internet to the "DMC" and central database.

- Some systems support both modes.

# 4.5 Error Detection & Correction (a.k.a. "Data Validation")

- For discussion, we categorize error detection/ correction procedures:

  – Computer procedures,

  – Manual (non-computer) procedures.

61

# 4.5 Error Detection & Correction: Manual Procedures

- EDC has fewer manual procedures (and fewer processing steps) than paper CRF processing.

- Pre-entry procedures:

  - Minimal: primarily the study coordinator reviews her/his own work.

  - Upside: this person is intimately familiar with the patient and data, has the greatest likelihood of spotting errors.

  - Downside:

    - Extreme dependent (vs. independent) verification
    - Proofreading one's own work

# 4.5 Error Detection & Correction: Manual Procedures

- Post-entry procedures:
  - The study coordinator reviews her/his own work.
    - Same upside and downside as for pre-entry procedures.

# 4.5 Error Detection & Correction: Computerized Procedures

- Most error detection is performed during the initial (and only) data keying.

- "Univariate tests:" the computer checks each keyed data value vs.:
  - A list of valid values (e.g., "F" or "M") for SEX)
  - A valid range (e.g., 60-110 for DIASTOLIC_BP)

- "Multivariate tests:" the computer checks a keyed data value for consistency with other data.
  - E.g., Dates must be in proper sequence.

# 4.5 Error Detection & Correction: Computerized Procedures

- When a data value fails a test the data entry operator (study coordinator) is alerted (Beep! Flash!) and an error message is displayed.

- The operator typically has two options:
  - Deal with the problem immediately, or
  - Postpone action until later.
    - The system marks the data value(s) as requiring action.

# Important Data Management Principle

- Earlier error detection typically results in:
    - Lower final error rates, and
    - Lower error detection/error correction costs.

# 4.5 Error Detection & Correction: "Query Management"

- Most systems do not have DCRs (paper query forms, "Data Clarification Requests").

- On request the system provides the operator with a list of data values requiring action.

- From time to time the study coordinator logs in and resolves data issues.

- Data records are not available for investigator signature until all issues have been resolved.

  - Some errors or missing data may be "resolved" as "not correctable" or "permanently unavailable."

# 4.6 "Commit Data Records": Obtain Signatures Required by Regulations

- FDA regs/GCP guidelines require that the site's investigator or authorized representative sign the CRFs.

- EDC systems capture the "electronic signature" of the data entry person at the time of data entry or data modification.

- Later, the investigator logs onto the system and electronically "signs" the data records.

- Any subsequent data modifications "un-do" signatures.

# 4.7 "Monitor" the Data: a CRA Verifies the Data in the Database

- The CRA monitors data, comparing data in the database to "source" data.

  - With electronic source capture the database contains the source data: no need to "compare."

  - With electronic capture from clinic records, the CRA's work is very similar to paper CRF procedures, except data are in the database.

  - With workbooks, procedures vary.

- Note that data have already been through error detection/correction procedures.

# 4.7 "Monitor" the Data:
# a CRA Verifies the Data in the Database

- This is the "verification step" for transcriptions of various types (EDC from clinic records, EDC via workbook).

# 4.8 "CRF Images"

- In EDC there are no CRF images.

- In the "old days," some Remote Data Entry systems printed "CRFs" after all processing was completed, in order to have a paper CRF for a wet investigator signature.

- Electronic signatures have eliminated this step.

# 4.9 Wrap Up

- We have made a quick tour with clinical trial data through a "system" that uses electronic data capture, including remote data entry components.

- Next, we turn to steps that are essentially similar for paper CRF systems and EDC systems.

# 5.  What Happens Next:
## From Clean Data to Database Lock

- At this point, data have been captured –one way or another—and queries have been resolved.

- However, data management is by no means done….

73

# 5.1 Medical Coding Review: ConMeds and AEs

**Auto and Manual Medical Coding**

**Send Listings to Sponsor for Medical Review**

**Medical Reviewer Returns List to Data Management**

**Apply Changes to Coded Data**
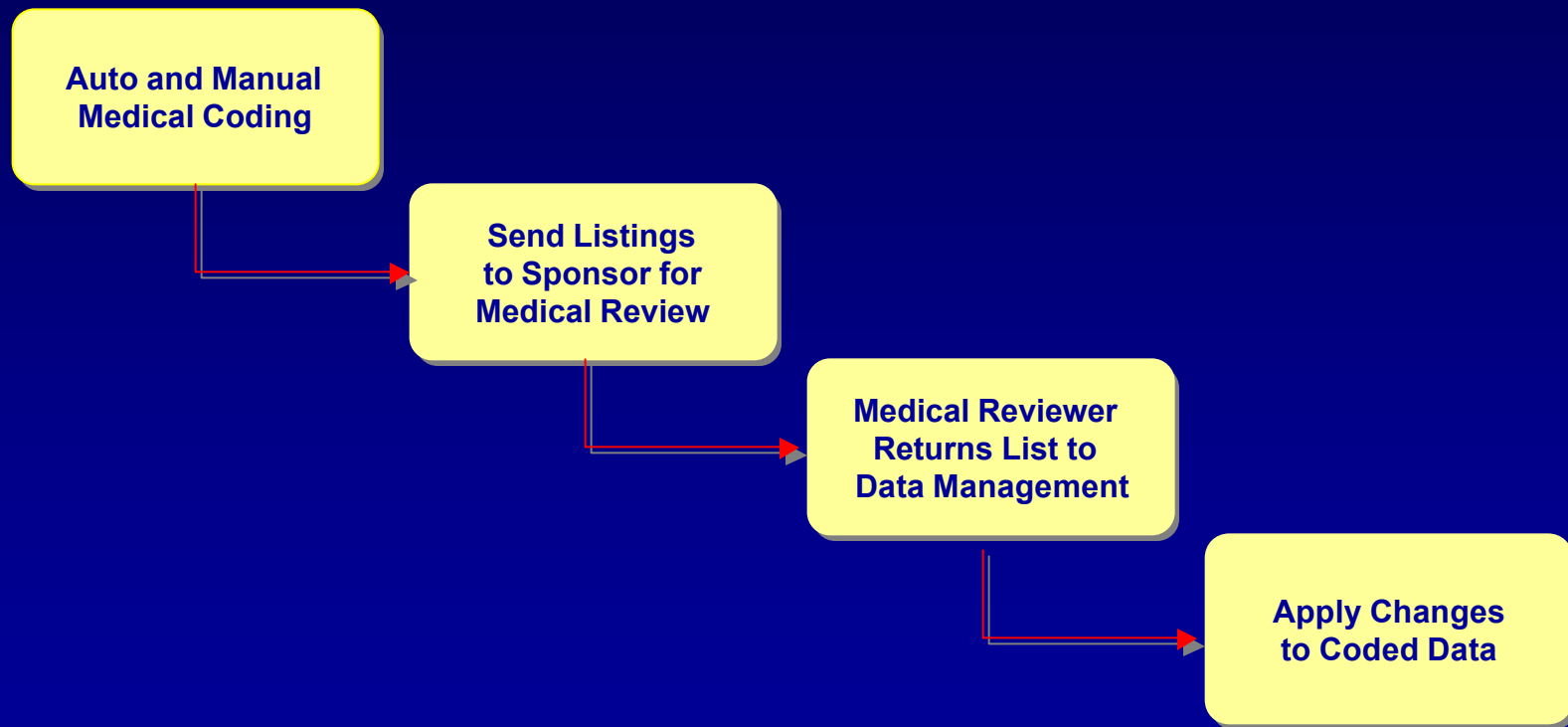
# 5.2  External Data

- Lab data
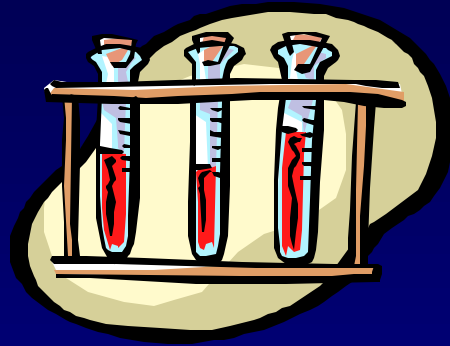
- Other instrument-collected data
  - Is the instrument "right"?

    The case of a heart rate ~ 800 bmp.

75

# 5.3 Activities That Occur Throughout the DM Process

- We are almost ready to lock the data, but the DM process often includes some other activities that may occur at various points along the way:
    - QC
    - Interim database snapshots

# 5.4 Quality Control

- *Quality Control* (QC) is an set of procedures used to manage the quality of ongoing data processing activities.  These are procedures implemented by the same group of individuals performing the work.

    - The goal is to keep error rates below specified limits.

77

# 5.4 Quality Control

- The quality control at the source to CRF level is the review by the clinical monitor of a sample (up to 100%) of the fields on the CRF against the source documentation.

- Univariate and multivariate error checks are a second set of indirect quality control checks of the source to CRF accuracy.

# 5.5 Interim Database Snapshots

- In some cases, snapshots of the data may be required prior to the end of the study.
    - For an interim analysis or DSMB report
    - To allow programmers to start initial programming for tables and listings

# 5.6 Database Lock

```
All CRFs Entered          All DCFs Resolved          Closure Activities
and Verified                                          Performed
     |                          |                          |
     v                          v                          v
All Electronic Data       Required Data              QA Audit of
Integrated into           Changes Made               Database Completed
Database                  to Database
     |                          |                          |
     v                          v                          v
Electronic and Manual     Ongoing QC and Review
Edit Checks               of Medical Coding                    Database
Performed                 Performed                            Lock
```
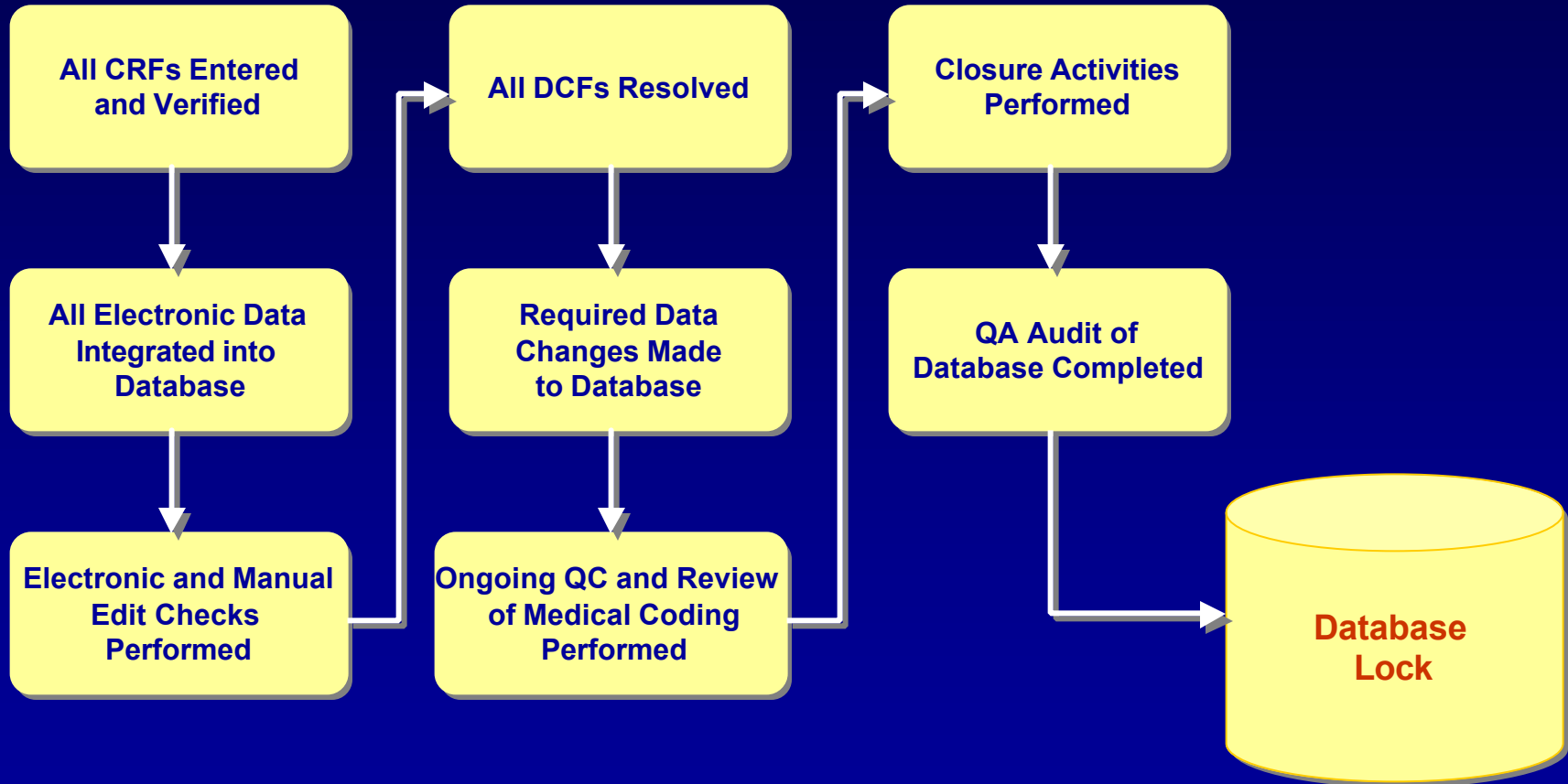
# 5.6 Database Lock

- Document database lock
- Retain the data and CRFs for period of years
- Note: Changing a locked database is a regulatory event.

# 5.6 Database Lock

- Now there is locked data.

- Are the data clean?

  - If these processes are rigorously followed, the data should be as clean as feasible, given what is available to data management.

# 5.6 Database Lock

- Why clean data?
  - *Data are the information of a clinical trial.*
  - It takes more time ($$$) to generate the analyses and a report from a dirty database.   (More later)
- Talking about saving time, it takes more time ($$$) to generate a clean database out of dirty CRF data, so
  - Train coordinators
  - Monitor the data well

# 6. Data Management Planning: Timeline Metrics

*Speaking of time…*
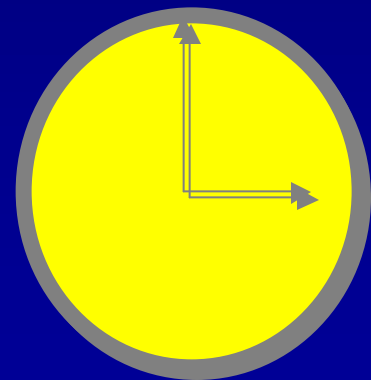
# 6. Data Management Planning: Timeline Metrics

- Time from finalized protocol to approved CRF
- Time from approved CRF to database ready for data entry
- Time from receipt of CRF to issuance of queries on that CRF
- Time from last patient, last visit to DB lock
  - Time from last query received to DB lock
- Time from DB lock to draft Tables, Listings, Figures (TLF) for key variables
  - Time from draft TLF to draft report

# 6.1 The Data Management Plan (DMP): Purpose

- To provide a guide to data managers, data processors, data entry staff on the study-specific DM handling.

- To provide information to other study team members about how the data were managed (reference document).

- May act as an archive for DM process modifications during the study.

# 6.2 DMP:  Scope

- Standard data management processes
- Specialized processes introduced to handle study specific requirements
- Processes from CRF design through DB lock and DB transfer

# 6.3 DMP:  General Contents

- Often written from a company specific template, based on DM SOPs
- Hardware/Software
- Database Structure
- Edit Checks
- Data Processing
- Data Conventions

# 6.3 DMP:  General Contents

- Data Flow Logistics
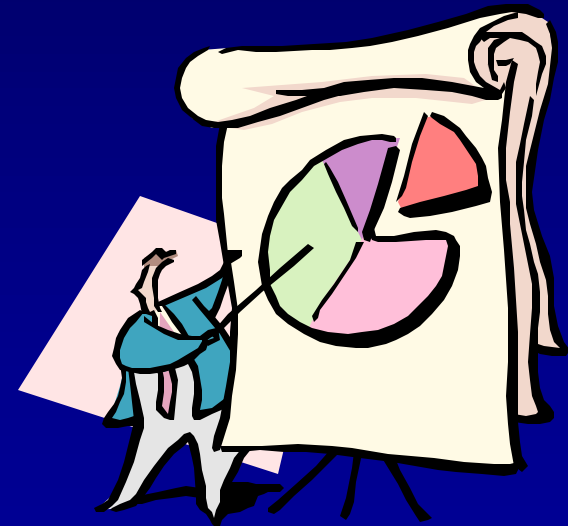- Query Logistics
- Quality Control Procedures
- Medical Coding and Approval
- Interim DB Transfers
- Handling data from external sources

# 6.3 DMP:  General Contents

- DB Lock
- Quality Assurance
- Data Archiving

# 7.0 DM Reporting

- Common DM Reports
  - CRF Tracking Report
  - Outstanding Queries
  - Query Trends
  - Database Quality
  - Medical Coding Review
  - Database Audit

# 7.1 DM Reporting: CRF Tracking

- There are several reports that provide general CRF tracking info:
    - By subject, by CRF page inventory list
    - By site list of completed CRFs or completed CRF modules
    - Missing pages report
- Purpose
    - To provide a sense of the completeness of the CRF harvesting efforts.
    - To identify trouble-spots for additional data collection.

# 7.2 DM Reporting: Outstanding Queries

- A list of outstanding queries, usually by site and patient ID.  This usually contains the query number and sometimes includes an identification of the number of days outstanding.

- Purpose:  To provide information to the monitors and site staff about the number of outstanding queries.  To provide information about query turnaround metrics.

# 7.3 DM Reporting: Query Trend

- This report provides frequency counts by query type (and possibly by site)

- Purpose
  - To target problematic hot spots on the CRF in order to be proactive (eg, additional site training, protocol amendments, CRF updates)
  - Identifying hotspots early on the in DM process can reduce queries and save $$$

# 7.4 DM Reporting: Database Quality

- This is a report of database quality at interim points during the trial.

- Purpose:  To provide information to statisticians conducting interim analyses or preparing DSMB or Steering Committee reports.

# 7.5 DM Reporting: Medical Coding Review

- This is a list of coded adverse events, concomitant medications, or other codable items.

- Purpose:  The report is provided in a form to evaluate the coding for appropriateness and for consistency across items.

# 7.6 DM Reporting: Database Audit

- This report is prepared immediately following the database audit and prior to database lock. It is a summary of the error rate of the database.

- Purpose:  Provides objective evidence of the degree to which the database matched the CRF and other supporting documentation (eg, queries to clarify a data point)

# 8.0 Troubleshooting DM Challenges

- Evaluate the situation from multiple perspectives
- Consider establishing a written process
- Identify the <u>interest</u> underlying a <u>position</u>
- Think creatively

# 8.1 Troubleshooting: Evaluating from Multiple Perspectives

- Example
  - Study Design
    - Phase 1: Randomized double-blind (45 days)
    - Phase 2: Open Label follow-up (1 year)
  - Data Collection
    - AE data collected at end of double-blind portion
    - Follow-up AE data collected during monthly monitoring
  - Problem: How to handle AEs that are ongoing at the end of Phase 1

# 8.1 Troubleshooting: Evaluating from Multiple Perspectives

- Clinical Perspective – Copy data on to follow-up AE CRF pages

- DM Perspective – Issue queries on AEs noted as on-going at the end of Phase I and let the queries sit at the sites until those events have ended.

- Other Perspectives – Stats, DSMB

# 8.1 Troubleshooting: Evaluating from Multiple Perspectives

- Solution:  Issue a tickler list to prompt CRAs/coordinators to complete the information about the event (ie, stop date and resolution).  Issue site initiated queries to resolve the events in the database.

# 8.2 Troubleshooting: Written Process

- When an issue is complex and it is important to have a clear process…
  - Work it out with the team
  - Establish a written document (eg, part of DMP, working practice document, guideline document)
  - Ensure that everyone is aware of and has access to the document.

# 8.3 Troubleshooting: Interest vs. Positions

- *Getting to Yes, Negotiating an Agreement without Giving in* – Fischer and Ury
- "I want the database locked on Friday" (2 weeks earlier than expected)
- Evaluate the motivating factors behind the position.
- Determine whether you can satisfy the interest <u>underlying</u> the position in another way.

# 8.4 Troubleshooting: Think Creatively

- Consider using old tools in new ways.
- Example:
  - Data management rescue following the receipt of a non-approvable letter from the FDA.
  - Cleaning the safety database via clinical review of subject data (eg, adding AEs that were noted elsewhere on the CRF).
  - Sponsor wanted some way to track the origin of each change.

# 8.4 Troubleshooting: Think Creatively

- Solution:

  - Status flags – designed to indicate the cleanliness of the data.

  - Data were not being subjected to electronic edit checks, therefore the status flags were not being used.

  - Use status flags to indicate the origin of the data change (eg, from comment field, from con med page)

# 9. Cost-Savings and Efficiencies

- Clean data sets save time ($$$) because dirty data are costly.

- Project managers can help implement certain efficiencies.

# 9.1 Cost-Savings and Efficiencies: Some Costs of Dirty Data

- Queries are expensive:  reliable estimates of the DMC's cost for creating, tracking, and processing queries range from $25 -$75 per query.

  - This does not include costs of CRAs, site time.

# 9.1 Cost-Savings and Efficiencies: Some Costs of Dirty Data

- Sometimes queries cannot be resolved:
  - The site is not longer present.
  - The clinic staff is no longer available or is unwilling to spend the time.

- This phenomenon is more likely to occur if data errors are found late in the process (often by statisticians or medical writers).

# 9.1 Cost-Savings and Efficiencies: Some Costs of Dirty Data

- When data are missing or are set to missing due to impossible values:
  - We lose data.
  - Data are the information of the study.
  - Lost data means lost money ($$$).
  - Statisticians might have to do analyses twice-- with the "bad" value and with the "bad" value set to missing ($$$) and write-up the results twice ($$$).

    (N.B.  Timelines and budgets are not friendly to the "analyze twice" phenomenon.)

# 9.1 Cost-Savings and Efficiencies: Some Costs of Dirty Data

- When erroneous values are not noticed, there still is a cost.

  - It's best not to have too many errors for an FDA auditor or the FDA reviewers to find.

  - Errors tend to increase the variability--effectively reducing the sample size and lowering the statistical power of the study.

# 9.2 Cost-Savings and Efficiencies: Efficiencies

- Handling certain obvious data errors
  - In some cases, the Sponsor will authorize the data management team to automatically correct certain types of obvious errors.
  - Terminology:
    - Level 1 Correction
    - Self-Evident Correction
    - Allowable Change
    - Sponsor Approved Correction
  - May pursue investigator signature of actual changes or advanced signature on class of changes.

# 9.2 Cost-Savings and Efficiencies: Efficiencies

- Identify data query "hot spots"

  - Target CRF fields for initial review by CRA.

  - Identify CRF sections that have relatively high query rates.

  - Identify sites generating CRFs with relatively high query rates.

  - Retrain coordinators as necessary; redesign a CRF page for this (?) and/or future studies.

# 9.2 Cost-Savings and Efficiencies: Efficiencies

- Before data collection begins, project managers can build in efficiencies if they

  - Ensure that the CRF design is appropriate

  - Encourage adequate training for the coordinators

# 9.3 Cost-Savings and Efficiencies: Comments

- Data management sometimes dislikes comment pages on the CRF because they are so hard to enter.  However, comments provide insight into the conduct of the study and may reduce queries from DM.
  - "Patient missed Visit 4."
  - "Value of 95 (identified by variable and page) was recorded, but instrument was later determined to be out of calibration."

# 9.3 Cost-Savings and Efficiencies: Comments

- Statisticians and project managers like comment pages because comments explain the data and answer questions.

- Because they often do not have access to the entire CRF, it is hard for them to get the "big picture" without comments.

# 10.  Where we have been today

- It is cost-effective to generate data of high quality.
  - No amount of DM can fix really dirty data
  - No statistician can analyze validity into conclusions based on dirty data
  - Dirty data causes analysis and reporting delays
- Project management personnel are in a position to facilitate
  - The development of high quality processes and procedures
  - The implementation of those processes
  - Project specific training for clinical personnel
  - General training on the importance of high quality data

# 10.1 In conclusion…

Clinical trials are designed by one team and conducted by another group, which collects the *data*.

The *data* then are entered by a third team, analyzed by still another group, and sometimes reported by still another team.

Notice that the "*data*" component is in the middle.

# 10.1 In conclusion…

- *Data are the information of a clinical trial.*
  - To be good (clean) information, it must be
    - accurately recorded
    - accurately entered

- Project managers, who function with a unique, relatively global perspective, can help pull all the components together and facilitate communication in order to improve data quality.

# 10.1 In conclusion…

Our contribution to the project is enhanced when we understand the components of the project.

Today we have shared with you some of the not-really-so-mysterious workings of "Data Management" to increase, we hope, your understanding of that component of the process and your ability to facilitate the conduct of clinical trials.

# 10.1 In conclusion…

- *Clean data  → saved time  →*
  *saved $$$  and  high-quality results.*

- *Project mangers can help achieve the goals of well recorded and accurately entered data.*

# Contact Information

- Katherine L. Monti, Ph.D.
  KMonti@RhoWorld.com


- Nancy Yovetich, Ph.D.
  NYovetich@RhoWorld.com


The overheads will be available at

www.rhoworld.com .