Quality Data Management in Clinical Trials

Catherine Dillon, MS
Wenle Zhao, PhD
Department of Public Health Sciences
Medical University of South Carolina, Charleston, SC

July 20, 2017

The presenter has no commercial or financial interests, relationships, activities, or other conflicts of interest to disclose.

This presentation will not include information on unlabeled use of any commercial products or investigational use that is not yet approved for any purpose.
Data Coordination Unit at MUSC

Global expertise in biomedical research...

Neurological Emergencies Treatment Trials Network
Statistics & Data Management Center
Established 2007
Sponsored by: The National Institute of Neurological Disorders and Stroke (NINDS)

ACUTE LIVER FAILURE STUDY GROUP
Statistics & Data Management Center
Established 2010
Sponsored by: The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)

NIH StrokeNet
Prevention | Treatment | Recovery
National Data Management Center
Established 2014
Sponsored by: The National Institute of Neurological Disorders and Stroke (NINDS)
1. Clinical trial data quality risk factors
2. Prevention of data quality problems
3. Detection of data quality problems
1. Clinical trial data quality risk factors
Trial quality and data quality

A successful trial

Plan what to do
Do what is planned
Record what is done
A failed trial

What needs to be done

GCP

Study design

Study Coordinator

Data management

Data Manager

Statistician

Physician

Trial operation
Causes of failed trials

- Treatment effect does exist
  - Signal
  - Dilution
  - Positive
- Treatment effect does not exist
  - Under power
  - Bias or Fraud
  - Neutral / Negative

Positive result ≠ Successful trial
Neutral / Negative result ≠ Failed trial

The goal of a trial is to find out the truth; positive, negative, or neutral.
<table>
<thead>
<tr>
<th>Problem</th>
<th>Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot recruit enough patients</td>
<td>Narrow eligibility criteria for patients, or complex study procedures for sites.</td>
</tr>
<tr>
<td>Science is not there</td>
<td>Insufficient or unreliable data for the hypothesis to be tested.</td>
</tr>
<tr>
<td>Too many research aims</td>
<td>Excessive clinical assessment and data collection demands on limited resource.</td>
</tr>
<tr>
<td>Unstable study design</td>
<td>Lack of detailed study protocol. Frequent protocol amendments.</td>
</tr>
<tr>
<td>Potential bias and fraud</td>
<td>Randomization method is vulnerable to selection bias and assessment bias.</td>
</tr>
</tbody>
</table>
Excessive trial operation procedures

\[ \text{Quality} = \frac{\text{Resource}}{\text{Quantity}} \]

With limited resources, more work to be done suggests lower quality to be expected.
Example 1: RAMPART, a large simple trial

<table>
<thead>
<tr>
<th>CRF</th>
<th>Baseline Jan-2011</th>
<th>End of Study Jan-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Enrollment</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Protocol Violations/Deviations</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>ED Arrival Form</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Data Logger (Central Reader)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Informed Consent Log</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Adverse Events</td>
<td>✓</td>
<td>❌</td>
</tr>
<tr>
<td>Affirmation of Adverse Event Assessment</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>End of Study Form</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Unique CRFs = 8
Unique Data Items = 178
Study Visits = 2
Visit-CRF Posting = 9

Trial of the Year award, Society for Clinical Trials 2013
Example 2: ProTECT, a complex trial

<table>
<thead>
<tr>
<th>CRF</th>
<th>Data Item 1</th>
<th>Data Item 2</th>
<th>Data Item 3</th>
<th>Data Item 4</th>
<th>Data Item 5</th>
<th>Data Item 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Item 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Item 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Item 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Item 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Item 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Item 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unique CRFs = 44
Unique Data Items = 1278
Study Visits = 10~39
Visit-CRF Posting = 158~535

After **882** subjects enrolled from **38** sites in **4** years, this study was terminated in 2014 due to futility.
Example 3: Investigator assessment bias

Hamilton Rating Scale for Depression (HRSD) assessment

- HRSD at Baseline
- HRSD after Treatment

Mean difference in HRSD between CA and IBA

Example 3: Investigator assessment bias

Hamilton Rating Scale for Depression (HRSD) assessment

- HRSD at Baseline
- HRSD after Treatment

Mean difference in HRSD between CA and IBA
Example 4: Suspicious selection bias

The NINDS rt-PA Stroke Study

31 patients had randomization glitches, 26 tPA, 5 placebo. \( p = 0.000096 \)

1 placebo \(\rightarrow\) tPA \quad 21 tPA \(\rightarrow\) placebo \quad \( p = 0.0000055 \)

Selection bias?

Sources: Marc Walton, Clinical Review for PLA 96-0350, June 12, 1996.
Operation issues of a failed trial

- **Missing primary outcome** ➔ Under worst scenario: \( \Delta^* = \Delta(1 - p_{\text{missing}}) \)

- **Enroll patients too healthy** ➔ Alway success: \( \Delta^* = \Delta(1 - p_{\text{too healthy}}) \)

- **Enroll patients too sick** ➔ Alway fail: \( \Delta^* = \Delta(1 - p_{\text{too sick}}) \)

- **Treatment cross-over** ➔ Under ITT: \( \Delta^* = \Delta(1 - 2p_{\text{cross-over}}) \)

Having a subject with such error is worse than not enrolling the subject!

To recover the power loss for 1 cross-over, 2 subjects are needed.
Trial operation errors

- Ineligible patient
- Treatment cross-over
- Treatment non-adherence
- Outcome assessment errors
- Missing primary outcome

Unbiased trial operation errors dilute the treatment effect, if it exists.
Power reduction due to signal dilution

\[ p_1 = 0.32, \quad p_2 = 0.25, \quad n = 1400, \quad \alpha = 0.05, \quad \text{Power} = 82.74\% \]

- \( r_a = 3\%: \) Not sick enough \((p=1)\),
- \( r_b = 3\%: \) Too sick \((p=0)\),
- \( r_c = 3\%: \) cross-over,
- \( r_d = 10\%: \) non-adherence \((\delta = 50\%)\),
- \( r_e = 3\%: \) success \(\rightarrow\) failure,
- \( r_f = 3\%: \) failure \(\rightarrow\) success,
- \( r_g = 3\%: \) missing outcome,

\[ \text{Actual Power} \approx 60\% \]

To get the same power under this trial operation quality, we may need a sample size of \(2400\).
1. The quality of a clinical trial starts from the design.

2. Operation errors dilute the treatment effect.

3. Data management cannot fix problems caused by trial design issues and operation errors.

4. Data managers can help to prevent trial design mistakes and operation errors.
2. Prevention of data quality problems
“An ounce of prevention is worth a pound of cure.”

— Benjamin Franklin
### Three types of data problems in clinical trials

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Operation Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not done</td>
</tr>
<tr>
<td></td>
<td>Done incorrectly</td>
</tr>
<tr>
<td></td>
<td>Done correctly</td>
</tr>
<tr>
<td><strong>Not done</strong></td>
<td>Missing data not confirmed</td>
</tr>
<tr>
<td><strong>Done incorrectly</strong></td>
<td>Unintentional data error</td>
</tr>
<tr>
<td><strong>Done correctly</strong></td>
<td>Confirmed missing data</td>
</tr>
<tr>
<td></td>
<td>Confirmed protocol violation</td>
</tr>
<tr>
<td></td>
<td>Well done!</td>
</tr>
</tbody>
</table>
Action #1: prevent missing data

- **Study visit**: Ensure that required study visits cannot be skipped in the study visit transition matrix.

- **CRF**: Ensure that required CRFs are completed in data collection schedule for each study visit.

- **Data item**: Ensure that required data questions are answered before CRF submission.

- Provide the options for confirmation of missing data.
Example 5: Prevent missing study visits

- Each subject starts from [Baseline] visit and ends at [End of study] visit.

- Transition condition from one visit to another visit must be clearly defined.
### Example 6: Prevent missing study visits

Transition logic must cover all possible scenarios, including protocol permitted skipped visits.
## Digitalization of study visit transition logic

<table>
<thead>
<tr>
<th>Current Visit</th>
<th>Subject Study Progress</th>
<th>Next Visit</th>
<th>Condition</th>
<th>Computer Logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Subject has been randomized</td>
<td>24 Hour</td>
<td></td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=3) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>24 Hour</td>
<td>Subject has completed the 24 hour visit</td>
<td>Hospital discharge</td>
<td>Discharge to Day 30 only if Day 30 is not previously posted</td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=4) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Hospital discharge</td>
<td>Subject has completed discharged visit</td>
<td>Day 30</td>
<td></td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Day 30</td>
<td>Subject has completed the Day 30 visit</td>
<td>Day 90</td>
<td></td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Day 90</td>
<td>Subject has completed the Day 90 visit</td>
<td>End of Study</td>
<td>Day 90 to End of Study only if Discharge is previously posted</td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>24 Hour</td>
<td>Subject has completed the 24 hour visit and was not discharged</td>
<td>Day 30</td>
<td></td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Hospital discharge</td>
<td>Subject has terminated early or completed the study</td>
<td>End of Study</td>
<td></td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Hospital discharge</td>
<td>Subject has completed discharged visit</td>
<td>Day 90</td>
<td>Discharge to Day 90 only if Day 90 is not previously posted</td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Day 30</td>
<td>Subject has terminated early</td>
<td>End of Study</td>
<td>Day 30 to End of Study only if Discharge is previously posted</td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Day 30</td>
<td>Subject has terminated early or was discharged</td>
<td>Hospital discharge</td>
<td>Day 30 to Discharge only if Discharge is not previously posted</td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
<tr>
<td>Day 90</td>
<td>Subject has terminated early or was discharged</td>
<td>Hospital discharge</td>
<td>Day 90 to Discharge only if Discharge is not previously posted</td>
<td>select zSubject.ID, case when exists(select ID from dbo.zSubjectVisit where zSubject.ID = zSubjectVisit.zSubjectID and zVisitID=5) then 0 else 1 end as ConditionValue from dbo.zSubject where zSubject.ID = $theSubjectID$</td>
</tr>
</tbody>
</table>

Computerized logic enforces protocol compliance and data completeness.
Enforcement of study visit transition logic

Only visits that meet the transition logic are available for selection.
Prevent missed study visit.
Promote immediate data entry.
Prevent missing CRFs

To protect the data integrity:

- All subject related study data are collected on Case Report Forms.
- CRF collection schedule is defined by study visits.
- CRFs are posted for the subject only when the subject visit is registered.
- Define CRF requirements:
  - Unconditional required
  - Conditional required
  - Optional
- Prevents duplicate, missing, and mismatched CRFs.
Example 7: CRF collection schedule

Well defined data collection schedule protects data integrity.
Example 8: Flag missing CRF

All missing CRFs are clearly flagged on the data collection schedule for each subject.

<table>
<thead>
<tr>
<th>CRF</th>
<th>Screening 2016</th>
<th>Baseline 2016</th>
<th>End of Week 1 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>F101 Inclusion and Exclusion Criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F102 Randomization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F501 Medical and Social History</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F502 Arm Motor Fugl-Meyer Assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F503 Box and Block Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F120 Modified Ashworth Scale of Spasticity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F504 Nottingham Sensory Assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F117 Vital Signs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F144 Modified Rankin Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F143 NIH Stroke Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F505 Stroke Knowledge Exam</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F506 Prior and Concomitant Medications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F507 Pre-Stroke Handedness Inventory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F169 Stroke Impact Scale - Hand Subsection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F508 Physical Activity Enjoyment Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F509 Optimization in Primary and Secondary Control Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F510 Patient Satisfaction Questionnaire</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F104 Adverse Event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F210 Study Therapy Compliance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F511 Non-Study Rehabilitation Therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 9: Prevent missing data items

- **Yes**: Data is required for CRF record saving.
- **Rejection, if missing**: Data is required for CRF submission.
- **Warning, if missing**: CRF can be submitted with explanation for missing data.
- **No**: Data is optional.
Example 10: Flag missing data items

- Missing data item is flagged.
Action #2: prevent data with unintentional error

- Make data questions clear and easy to answer.
- Create multiple choice questions with sound logic.
- Use skip pattern control to reduce data error.
- Use real-time rule-based data validation to prevent typos.
- Use standardized CRFs to avoid design errors and revisions.
- Control protocol amendments and CRF changes.
- Maximize computer potential and minimize site work load.
**Ask simple questions**

Q01: Baseline blood glucose < 50mg/dL or > 400mg/dL? Yes / No

Q01: Baseline blood glucose is ________ (mg/dl)

Q02: How long since your last dentist visit? ______ (days)

Q02: Date of your last dentist visit: _ _/ _ _/ _ _ _ _ (mm/dd/yyyy)

Q03: Did the patient have chickenpox or measles in the past 12 months and ear infection in the past 6 months? Yes / No

Q03: Did the patient had chickenpox in the past 12 months? Yes/No
Q04: Did the patient had measles in the past 12 months? Yes/No
Q05: Did the patient had ear infection in the past 6 months? Yes/No
Q07: What color is it?

- Red
- Yellow
- Black
- Red
- Salmon
- Fire Brick
- Dark Orange
- Coral
- Tomato
- Orange Red
- Chocolate
- Saddle Brown
- Wheat
- Gold
- Tan
Q01: Primary reason for screen failure (Please choose one only):

0 = Patient is not age 60 years or older
1 = Patient lacks capacity to provide informed consent
2 = Patient has language barriers
3 = Subject has lifetime history of bipolar affective disorder
4 = Subject has lifetime history of schizophrenia
5 = Subject has lifetime history of schizoaffective disorder
6 = Subject has lifetime history of intellectual disability
7 = Subject has a current diagnosis of delirium
8 = Subject has a current diagnosis of dementia
9 = Subject has substance dependence in past 6 months
10 = Patient has a medical condition contraindicating Li or VLF

Replace long list of radio button options with free text.
Have a central investigator do the grouping afterwards.
Example 11: Consider all possible responses

- Including “Other” in Q02 to cover all not listed.
- Including “None of the above” in Q05 to positively confirm the response.
Example 12: Use skip pattern controls

- Define skip pattern controls to avoid basic logic conflicts.
- Enforce skip patterns in EDC user interfaces.
Example 13: Define data validation rules

Real-time alert for users to prevent data entry typo errors.

For this example, the End of Study CRF has 23 data validation rules, checking missing data, data value range, and date/time sequences.
Levels of data validation rules

1. **Reject record saving**: prevent basic logic violations.
   - data type mismatch
   - missing index field data

2. **Reject CRF submission**: prevent data logic conflict, and ensure CRF completion.
   - Checked “Other”, without “Other specify”
   - Incorrect date/time sequence

3. **Protocol violation**: entered value indicates protocol violation. Allow CRF submission after confirmation of protocol violation.
   - Eligibility criteria violation.
   - Study treatment (dose, timing, duration, etc.)

A data validation rule may include multiple data items on the same CRF or across different CRFs.
Use validated assessments

- Hamilton Rating Scale for Depression (HRS-D)
- NIH Stroke Scale (NIHSS)
- The Short Form (36) Health Survey (SF-36)
- Modified Rankin Scale (mRS)
- Clinical Global Impression (CGI)
- Glasgow Outcome Scale (GOS)
- Montgomery-Asberg Depression Rating Scale (MADRS)
- Quality of Life in Neurological Disorders (Neuro-QOL)
- Pediatric Stroke Outcome Measure Short Neuro Exam (PSOM-SNE)
- King's Outcome Scale for Childhood Head Injury (KOSCHI)

Use validated assessments with version and source information to ensure data validity and intellectual property protection.
When using validated assessments:

1. Respect the integrity of the validated assessment. Do not make changes to the question or the answer options.

2. When *total score* is needed, allow users to enter manually calculated total score and include the computer derived total score for cross check.

3. For patient completed assessments, use original paper form to collect data and enter data into EDC afterwards. Use of electronic version requires additional validation.
Use of standards

Clinical Data Interchange Standards Consortium (CDISC)

Health Level Seven International (HL7)

NCI Enterprise Vocabulary Services (EVS)

Federal Interagency Traumatic Brain Injury Research (FITBIR)

NINDS Common Data Elements (CDE)
Changes on CRF may be requested after study start:

1. Add or remove CRFs or modify CRF collection schedule.
2. Add or remove data items.
3. Modify question text.
4. Add, modify, or remove response options.
5. Add, modify, or remove data validation rules.

General strategy for CRF changes:

1. Do not do it unless absolutely necessary, and benefit > cost.
2. Use a new data item for modified questions. Keep the old data item with data in the database.
3. When response options change, send collected CRFs back for reassessment.
Consider the workload distribution

Maximize work for Mr. SQL Server

Reduce work for Dr. P-value

Minimize burden for Ms. CDM

Employee of the year
2015
Mr. SQL Server

Employee of the year
2016
Mr. SQL Server
You can't wake a person who is pretending to be asleep.

Fake data is created intentionally.

Fake data destroys the validity of the clinical trial.
Prevention of fake data

Remove the motivation:
- Data manager’s responsibility is to get true data, good or bad.
- Well designed CRFs prevent forced lying.
- Do not ask the question if the answer is likely not available.
- Do not offer limited options that are not exhaustive.

Remove the capacity:
- Minimize treatment allocation prediction.
- Maintain treatment blind protection.
- Avoid PI’s micro-management of trial operations.
1. Prevent missing data by better design of study visit transition matrix, data collect schedule, and CRF.

2. Prevent data errors by better EDC user interface design.

3. Prevent fake data by removing the motivation and the capacity.
3. Detection of data quality problems
“Your best teacher is your last mistake.”

— Ralph Nader
1. Monitor subject study progress. Ensure compliance with study treatments and assessments specified in the protocol.


3. Review CRF rule violation and protocol violation reports.

4. Review CRF data that cannot be validated by computer rules.
Risk-based monitoring strategy

Risk-based data monitoring target: data that effects the trial results:

1. Eligibility CRF
2. Randomization CRF
3. Study treatment CRF
4. Adverse event CRF
5. End of study CRF
6. CRFs with primary – secondary efficacy outcomes
7. CRFs with confirmed protocol violations
Example 14: Subject study visit monitoring

Take immediate action for overdue study visits.
Example 15: CRF completion status monitoring

Contact site study coordinator in cases of delayed CRF submission.
Example 16: Protocol violation review

Contact site for confirmed protocol violation. Request Corrective and Preventive Action Plan (CAPA) when needed.
Example 17: Plan on-site monitoring visits

Check number of CRFs for risk-based monitoring.
Example 18: Central data quality monitoring

### F505 Demographics

#### Report Definition
- **Data Summarized By:** Wenle ZHAO on 9/18/2016 10:52:30 PM
- **Data Source:** vF505
- **Data Filter:** 1=1
- **Total Number of Records:** 162

#### Field Definition

<table>
<thead>
<tr>
<th>DB Field</th>
<th>Data Type</th>
<th>Field Definition</th>
<th>Unit</th>
<th>With Data</th>
<th>Missing</th>
<th>Mean</th>
<th>Stdev</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q03</td>
<td>Number</td>
<td>Number of years of education:</td>
<td></td>
<td>133</td>
<td>29</td>
<td>7</td>
<td>6.01</td>
<td>9</td>
<td>0</td>
<td>25</td>
<td>943</td>
</tr>
</tbody>
</table>

#### Additional Data

<table>
<thead>
<tr>
<th>DB Field</th>
<th>Data Type</th>
<th>Field Definition</th>
<th>Category</th>
<th>Count</th>
<th>Percent</th>
<th>Cumulative Count</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q01</td>
<td>Selection</td>
<td>Ethnicity</td>
<td>Hispanic or Latino</td>
<td>21</td>
<td>13%</td>
<td>21</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Not Hispanic or Latino</td>
<td>136</td>
<td>84%</td>
<td>157</td>
<td>97%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Unknown</td>
<td>5</td>
<td>3%</td>
<td>162</td>
<td>100%</td>
</tr>
<tr>
<td>Q02</td>
<td>Multiple Selection</td>
<td>Race</td>
<td>American Indian or Alaska Native</td>
<td>0</td>
<td>0%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Asian</td>
<td>10</td>
<td>6%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Black or African American</td>
<td>63</td>
<td>39%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>2</td>
<td>1%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>White</td>
<td>74</td>
<td>46%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Unknown</td>
<td>16</td>
<td>10%</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

All information is generated based on data currently in the database. Data may not be verified or validated. The report is generated to assist in trial operations only, and is not valid to support any statistical analysis of study data. Unless noted, the Data Coordination Unit (DCU) assumes no responsibility for the use of this report. This report may contain protected health information covered by the Health Insurance Portability and Accountability Act (HIPAA). You are prohibited from disclosing this information without the specific written consent of the person to whom it pertains. Anyone using this data specifically assumes responsibility for maintaining the confidentiality of the protected data.
Example 19: Central data quality monitoring

Findings:
1. CRF 5320 has number of years of education = 25.
2. At least 5 CRFs have number of education years = 0. (may be NA?)
**Central data quality monitoring**

### Data value pattern detection
- 1\(^{st}\) and 15\(^{th}\) day in a month; January and July in a year.
- Rounded to 0 or 5.

### Data comparison over time
- Loosened eligibility criteria due to slow enrollment.
- Low site team member retention rate.
- Performance disparity among sites.
- Slow site enrollment $\rightarrow$ less familiar with study protocol.
- Logic error among data on several CRFs.
Share data quality report within the study community with information on:

- Enrollment speed and subject retention rate.
- Data completeness.
- CRF submission and data query response timeliness.
- Number of confirmed protocol violations.
- Number of data error detected, but not corrected.
Trial performance dashboard – CRF submit
1. Data manager data review to detect data error.

2. Central data monitoring program.

3. Plan risk-based site monitoring visits when needed.

4. Share study performance summary data among sites.
Use a proper tool for the job

- White House to Capital Hill
- New York to Boston
- Houston to San Francisco
- Retrospective registry
- Small early phase trial
- Multicenter Phase 3 RCT
Thank You!

Contact:
Catherine Dillon
rileycp@musc.edu