Quality Data Management in Clinical Trials

Wenle Zhao, PhD

Department of Pubic Health Sciences

Medical University of South Carolina, Charleston, SC

St. Jude Children's Research Hospital, Memphis, TN,

October 17, 2016

MUSC
MEDICAL UNIVERSITY of SOUTH CAROLINA

DCU
Data Coordination Unit

Data Coordination Unit at MUSC

Neurological Emergencies Treatment Trials Network

Statistics & Data Management Center

Established 2007

Sponsored by: The National Institute of Neurological Disorders and Stroke (NINDS)

DCU

Acute Liver Failure Study Group

Statistics & Data Management Center

Established 2010

Sponsored by: The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)

DCU

NIH StrokeNet

Prevention | Treatment | Recovery

National Data Management Center

Established 2014

Sponsored by: The National Institute of Neurological Disorders and Stroke (NINDS)

DCU

Contents

1. Clinical trial data quality risk factors
2. Prevention of data quality problems
3. Detection of data quality problems

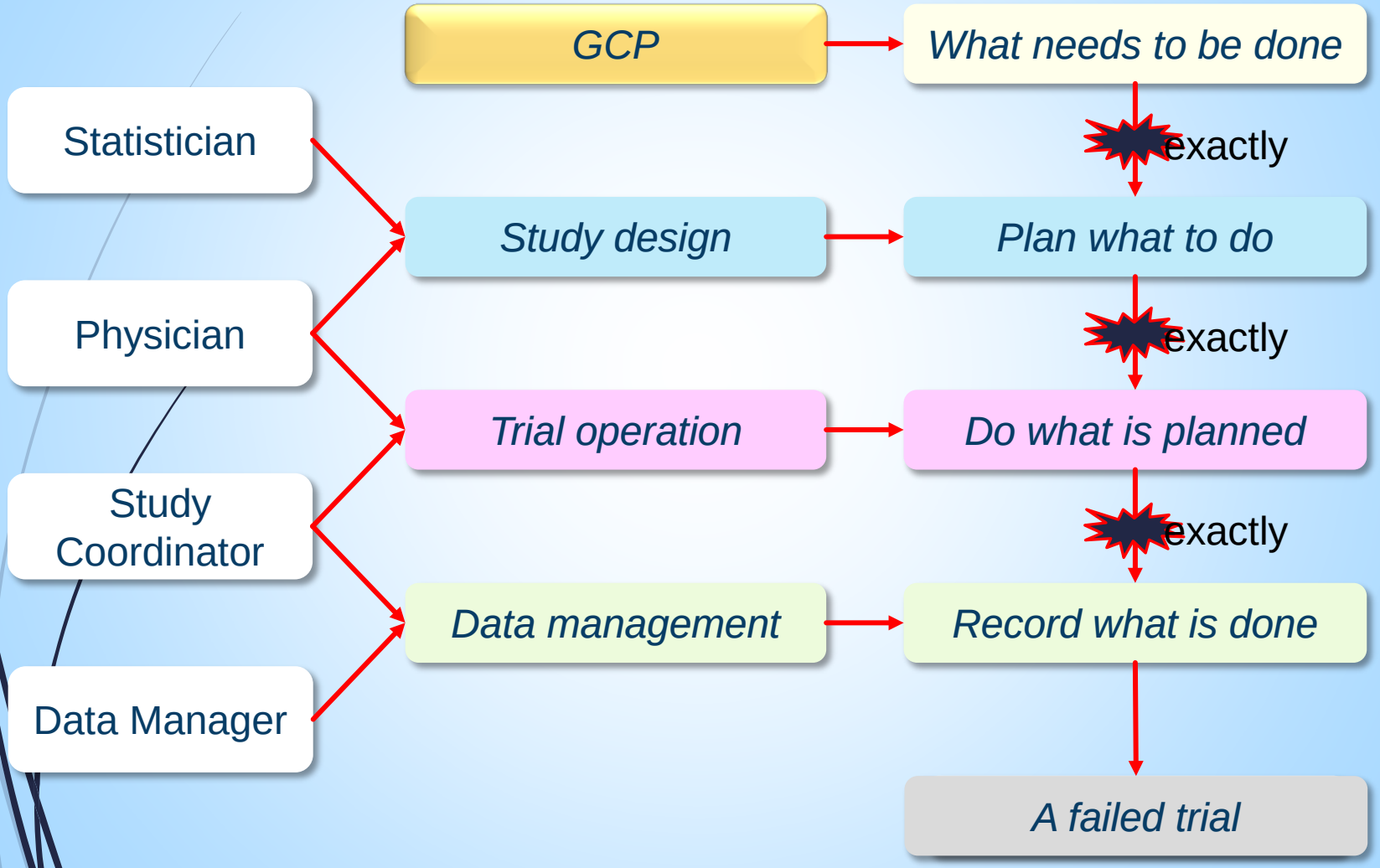
1.

Clinical trial data quality risk factors

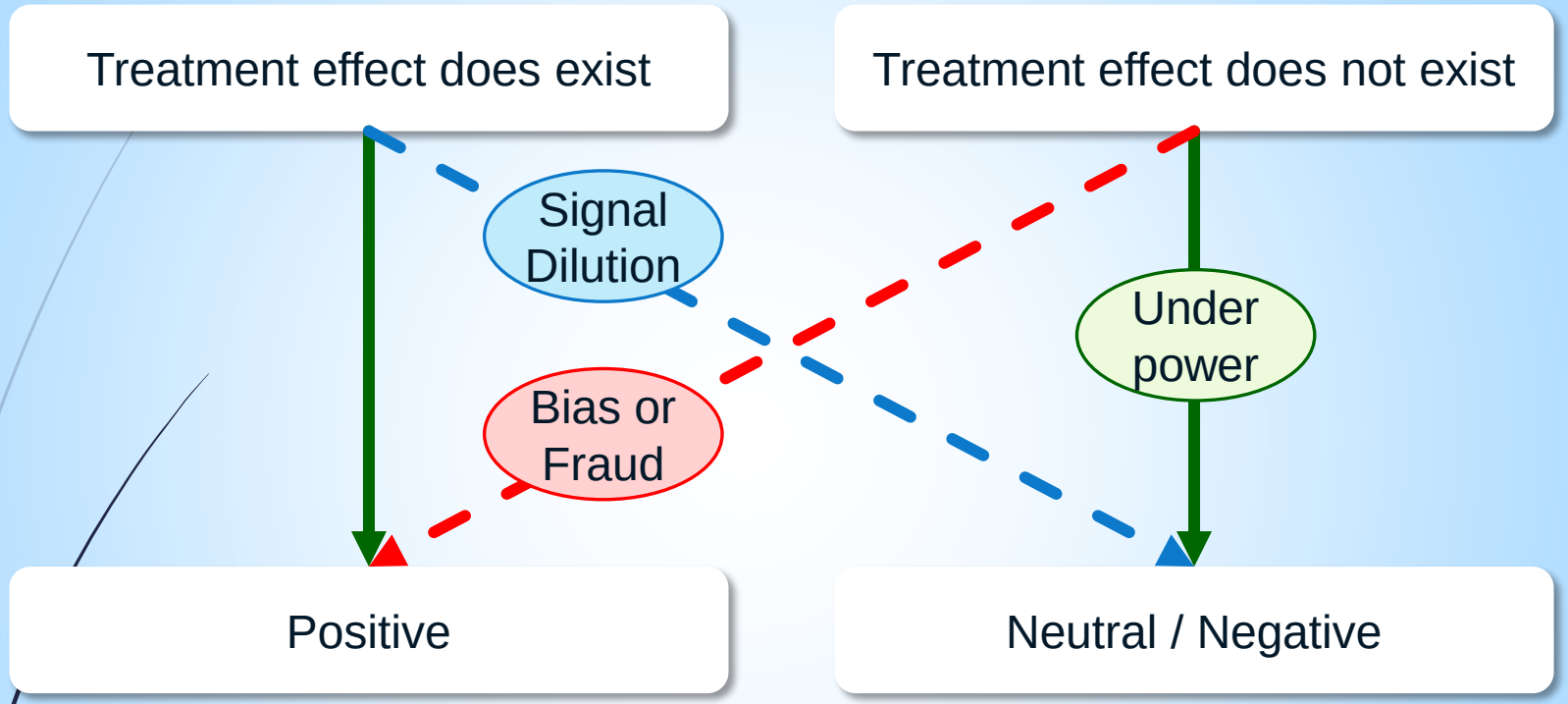
There is only one way of doing things correctly.

There are many ways of doing things wrong.

Trial quality and data quality



Causes of failed trials



Positive result ≠ Successful trial

Neutral / Negative result ≠ Failed trial

The goal of a trial is to find out the truth; positive, negative, or neutral.

Design issues of a failed trial

Cannot recruit enough patients ➔	Narrow eligibility criteria for patients, or complex study procedures for sites.
Science is not there ➔	Insufficient or unreliable data for the hypothesis to be tested.
Too many research aims ➔	Excessive clinical assessment and data collection demands on limited resource.
Unstable study design ➔	Lack of detailed study protocol. Frequent protocol amendments.
Potential bias and fraud ➔	Randomization method is vulnerable to selection bias and assessment bias.

Excessive trial operation procedures

$$Quality = \frac{Resource}{Quantity}$$

With limited resources, more work to be done suggests lower quality to be expected.

Example 1: RAMPART, a large simple trial

CRF	Baseline [redacted]-Jan-2011	End of Study [redacted]-Jan-2011
Subject Enrollment	☑	
Protocol Violations/Deviations	☑	
ED Arrival Form	☑	
Data Logger (Central Reader)	☑	
Informed Consent Log	☑	
Adverse Events	☑	☐
Affirmation of Adverse Event Assessment	☑	
End of Study Form		☑

Unique CRFs = 8

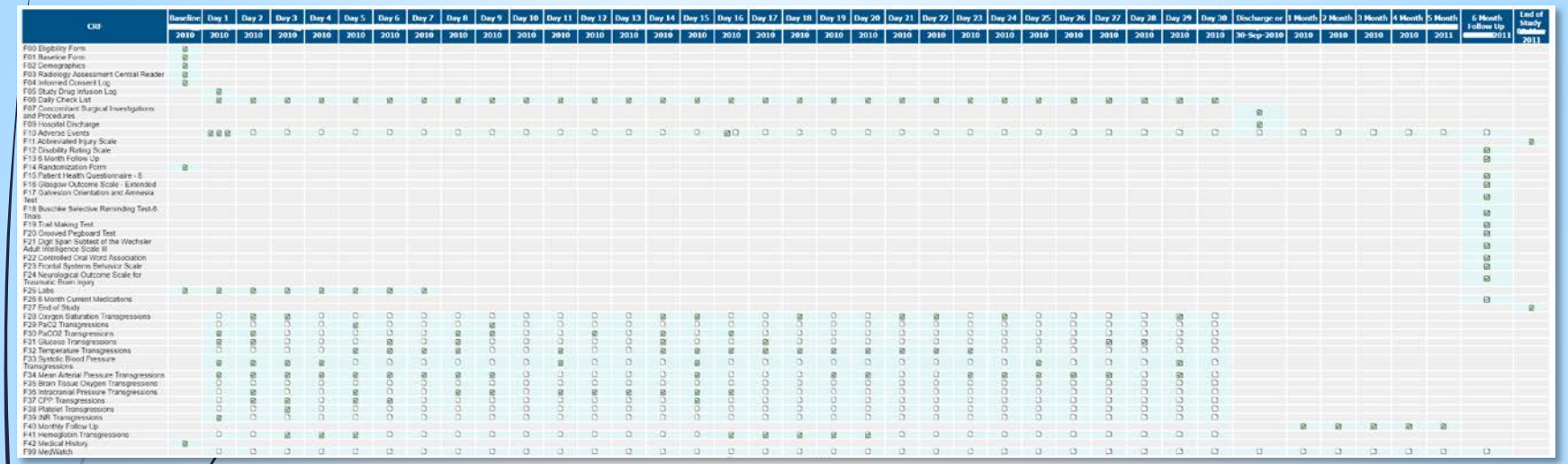
Unique Data Items = 178

Study Visits = 2

Visit-CRF Posting = 9

Trial of the Year award, Society for Clinical Trials 2013

Example 2: ProTECT, a complex trial



Unique CRFs = 44

Unique Data Items = 1278

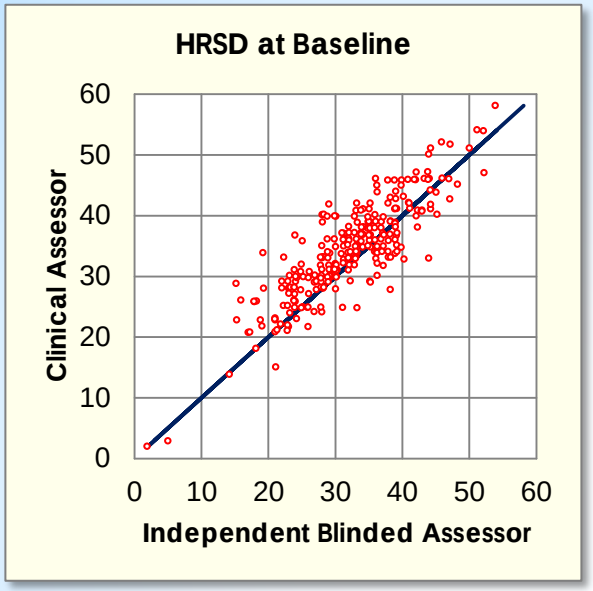
Study Visits = 10~39

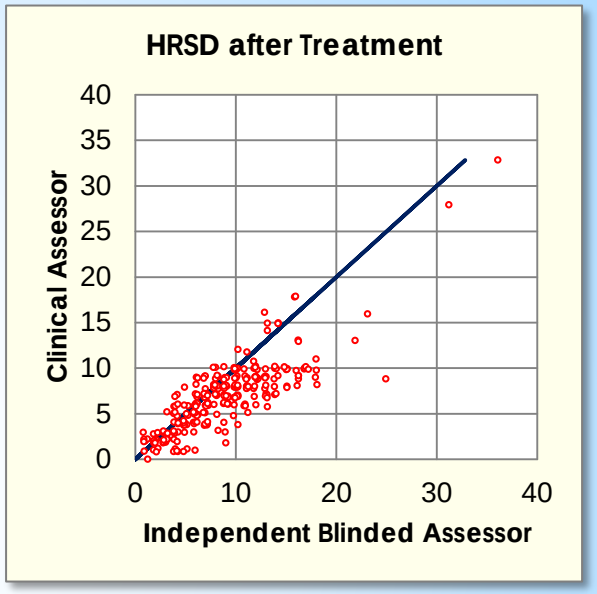
Visit-CRF Posting = 158~535

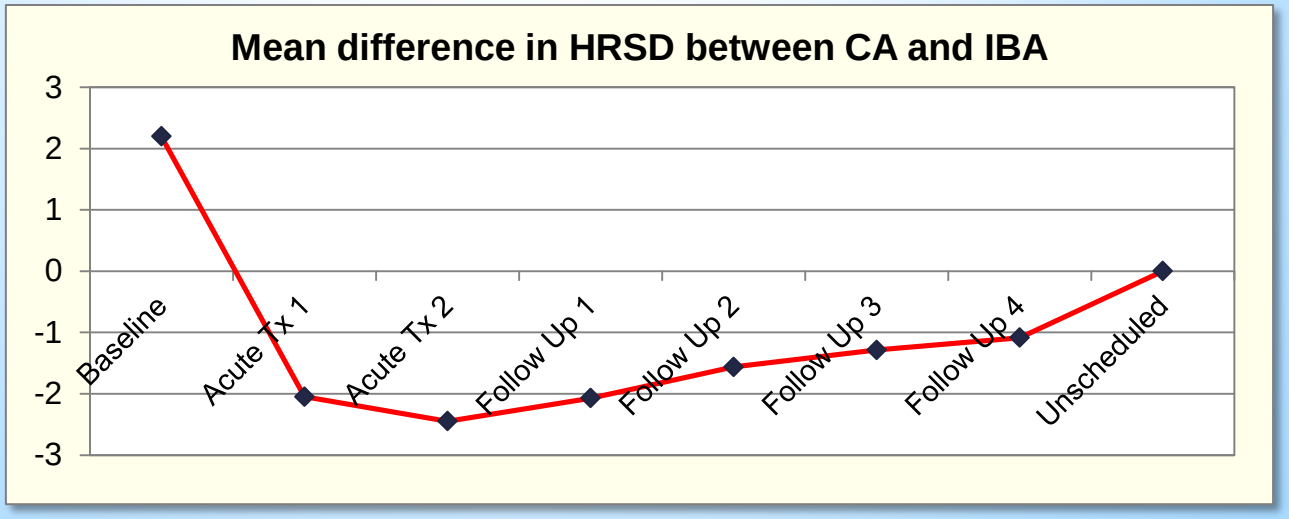
After **882** subjects enrolled from **38** sites in **4** years, this study was terminated in 2014 due to futility.

Example 3: Investigator assessment bias

Hamilton Rating Scale for Depression (HRSD) assessment



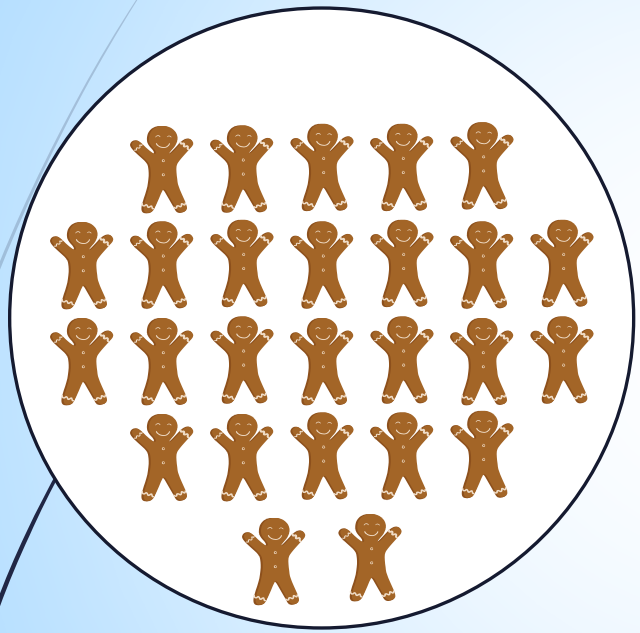




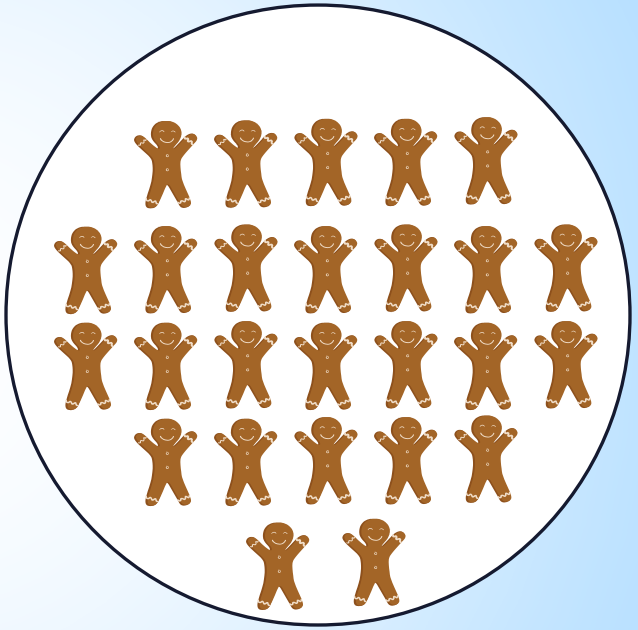
Example 4: Suspicious selection bias

The NINDS rt-PA Stroke Study

tPA arm



Placebo arm



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

1 placebo → tPA 21 tPA → placebo p = 0.0000055

Selection bias?

Sources: Marc Walton, Clinical Review for PLA 96-0350, June 12, 1996.

Vance Berger, Selection Bias and Covariate Imbalances in Randomized Clinical Trials, John Wiley & Sons, Ltd. 2005. Page 74-75.

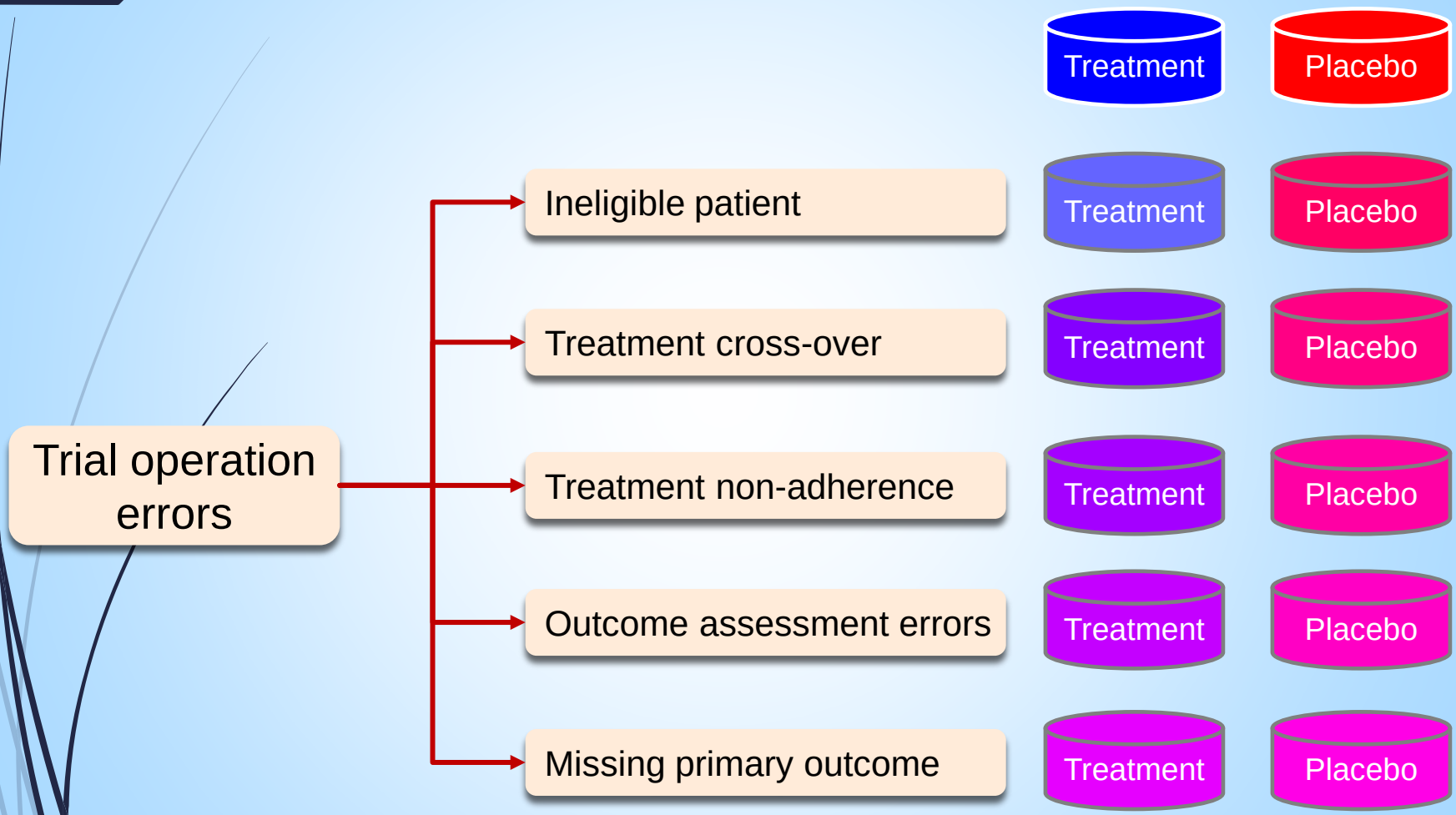
Operation issues of a failed trial

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: $\Delta^* = \Delta(1 - 2\,p_{\text{cross-over}})$

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.

Operation issues of a failed trial



Unbiased trial operation errors dilute the treatment effect, if it exists.

Power reduction due to signal dilution

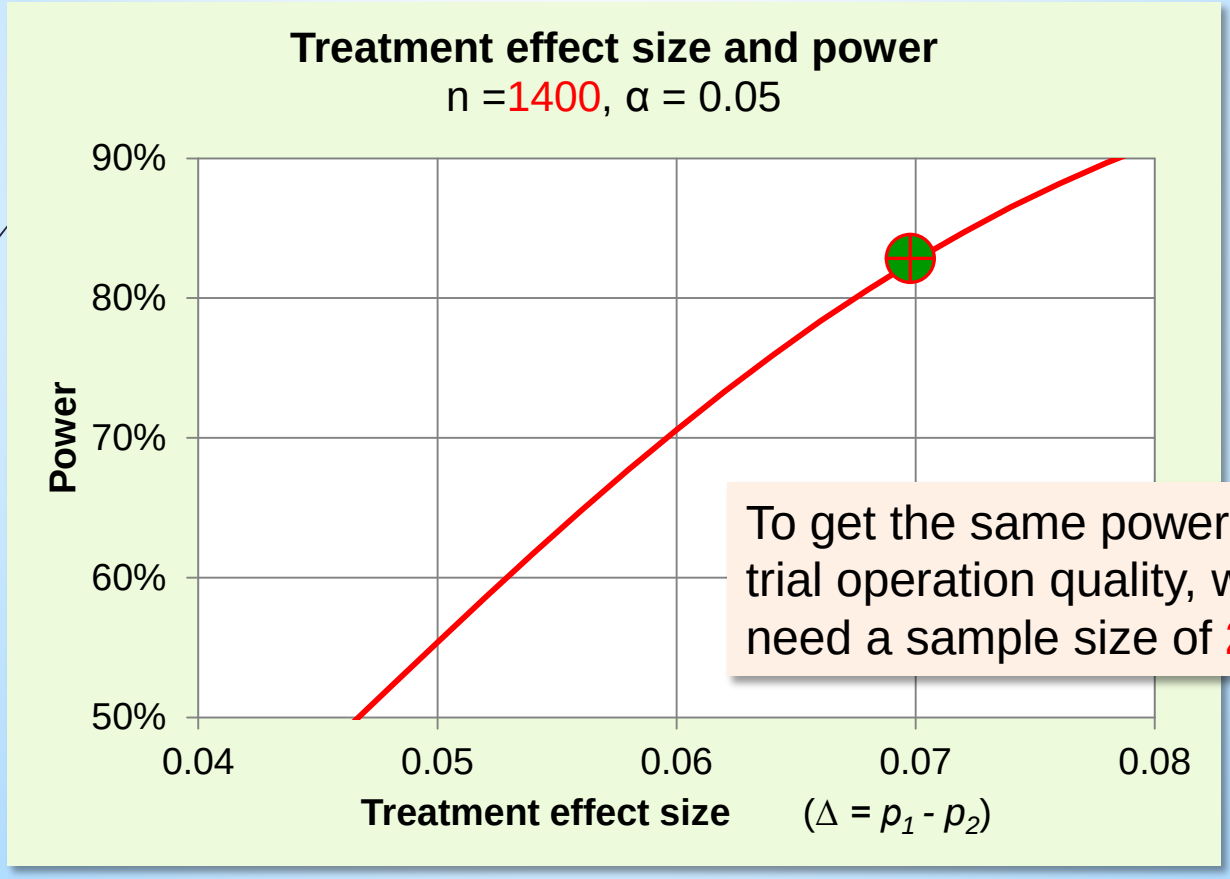
$p_1 = 0.32, \quad p_2 = 0.25, \quad n = 1400, \quad \alpha = 0.05, \quad Power = 82.74\%$

$r_a = 3\%$: 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, $Actual\ Power \approx 60\%$



To get the same power under this trial operation quality, we may need a sample size of **2400.**

Summary – Section 1

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.

5. A well designed and conducted trial can be ruined by poor data management.

2.

Prevention of data quality problems

“An ounce of prevention is worth a pound of cure.”

— Benjamin Franklin

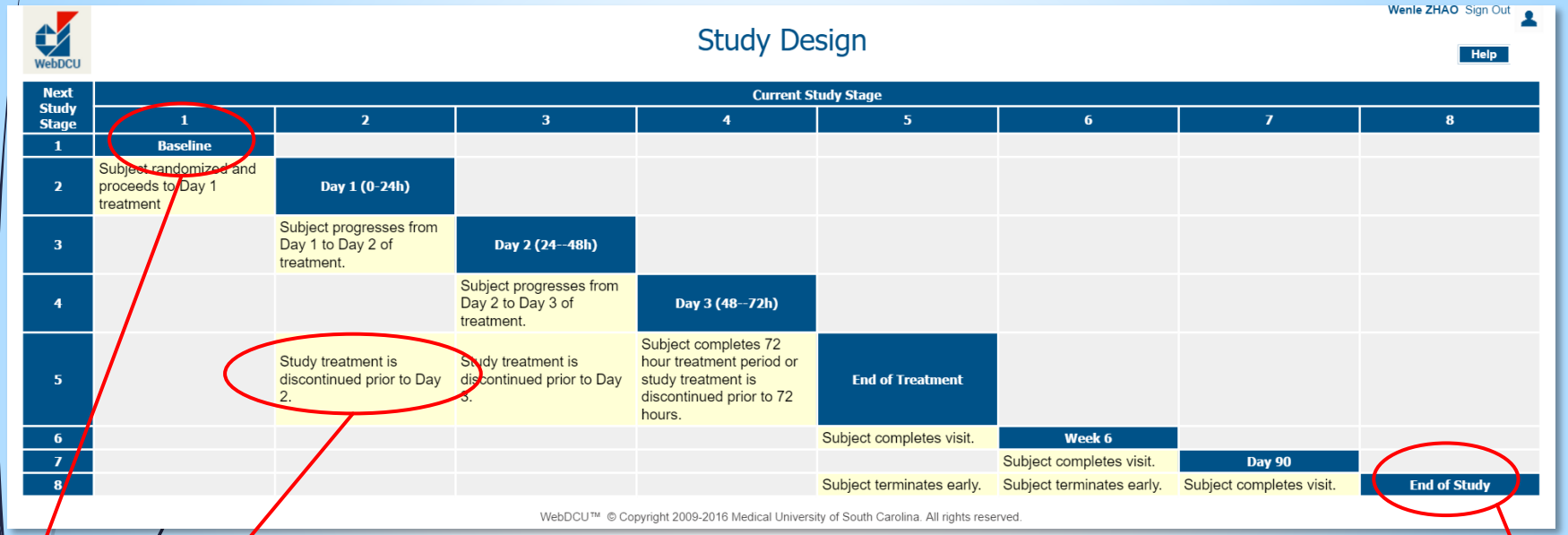
Three types of data problems in clinical trials

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="3">Operation Procedure</th></tr>
<tr><th>Not done</th><th>Done incorrectly</th><th>Done correctly</th></tr>
<tr><td rowspan="3">Data Collection</td><td>Not done</td><td>Missing data not confirmed</td><td colspan="2">Missing data</td></tr>
<tr><td>Done incorrectly</td><td rowspan="2">Fake data</td><td colspan="2">Unintentional data error</td></tr>
<tr><td>Done correctly</td><td colspan="2">Fake data</td></tr>
<tr><td colspan="2"></td><td>Confirmed missing data</td><td>Confirmed protocol violation</td><td>Well done!</td></tr>
</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



WebDCU

Wenle ZHAO Sign Out

Help

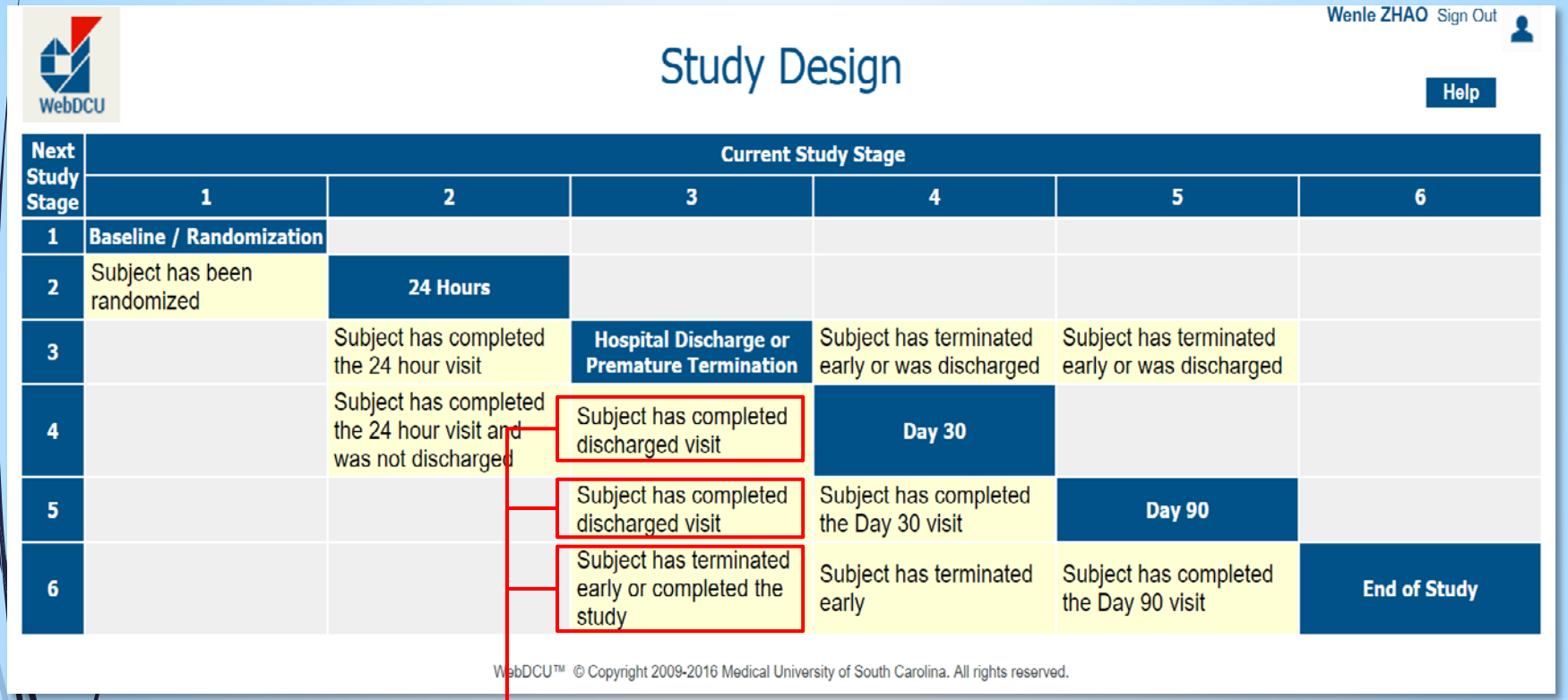
Study Design

Next Study Stage	Current Study Stage							
	1	2	3	4	5	6	7	8
1	Baseline							
2	Subject randomized and proceeds to Day 1 treatment	Day 1 (0-24h)						
3		Subject progresses from Day 1 to Day 2 of treatment.	Day 2 (24--48h)					
4			Subject progresses from Day 2 to Day 3 of treatment.	Day 3 (48--72h)				
5		Study treatment is discontinued prior to Day 2.	Study treatment is discontinued prior to Day 3.	Subject completes 72 hour treatment period or study treatment is discontinued prior to 72 hours.	End of Treatment			
6					Subject completes visit.	Week 6		
7						Subject completes visit.	Day 90	
8					Subject terminates early.	Subject terminates early.	Subject completes visit.	End of Study

- 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



WebDCU

Wenle ZHAO Sign Out

Study Design

Help

Next Study Stage	Current Study Stage					
	1	2	3	4	5	6
1	Baseline / Randomization					
2	Subject has been randomized	24 Hours				
3		Subject has completed the 24 hour visit	Hospital Discharge or Premature Termination	Subject has terminated early or was discharged	Subject has terminated early or was discharged	
4		Subject has completed the 24 hour visit and was not discharged	Subject has completed discharged visit	Day 30		
5			Subject has completed discharged visit	Subject has completed the Day 30 visit	Day 90	
6			Subject has terminated early or completed the study	Subject has terminated early	Subject has completed the Day 90 visit	End of Study

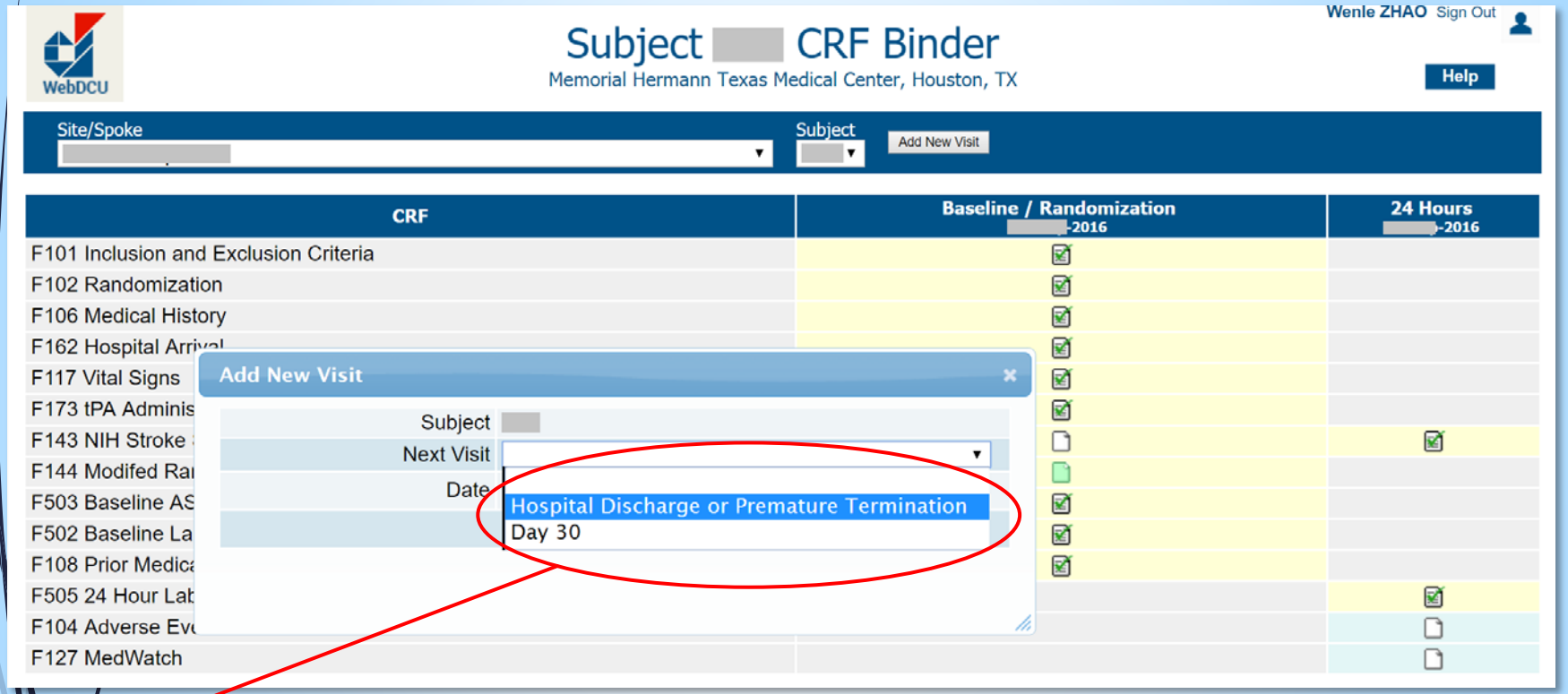
Transition logic must cover all possible scenarios, including protocol permitted skipped visits.

Digitalization of study visit transition logic

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

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.

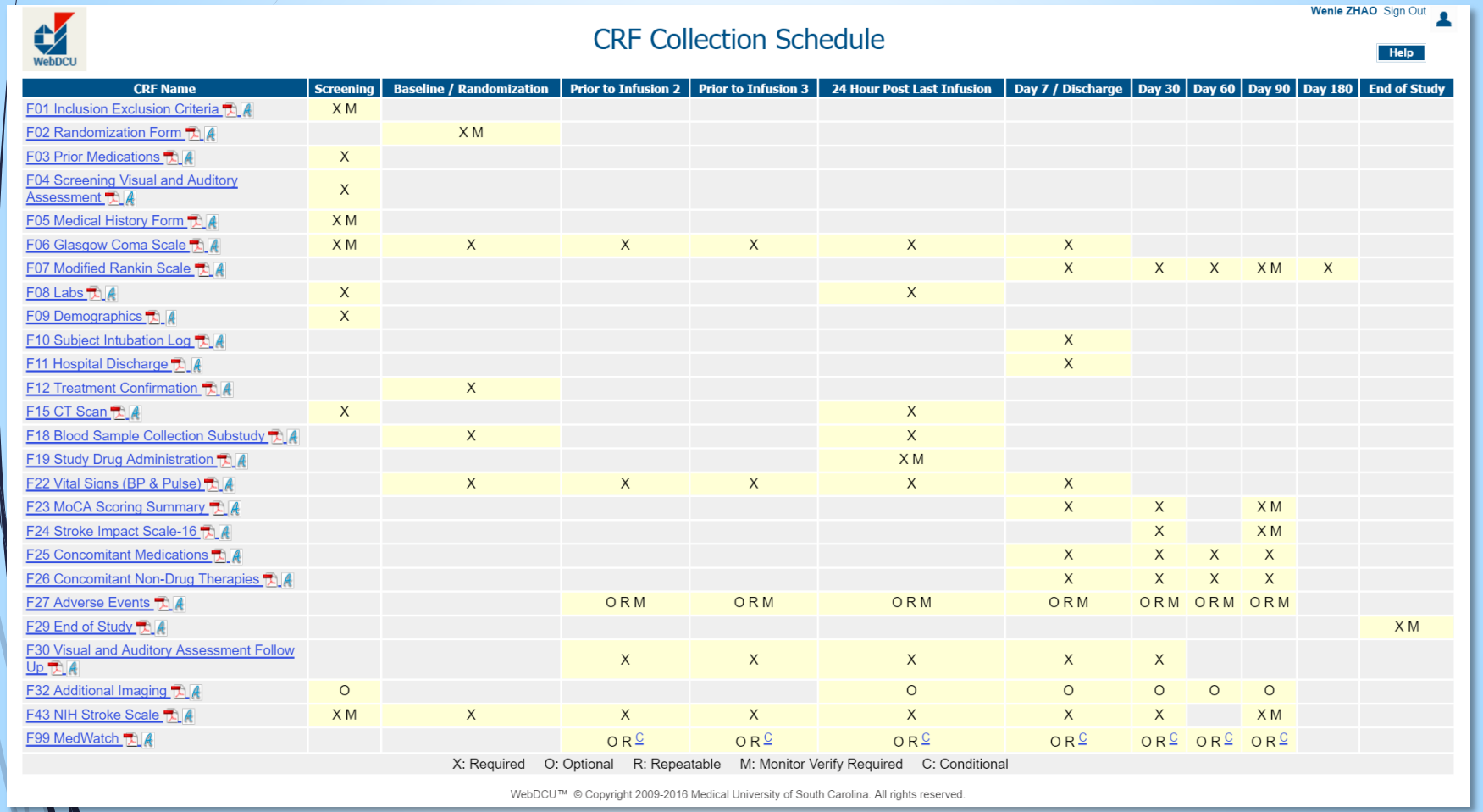
Promote immediate data entry.

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



WebDCU

Wenle ZHAO Sign Out

Help

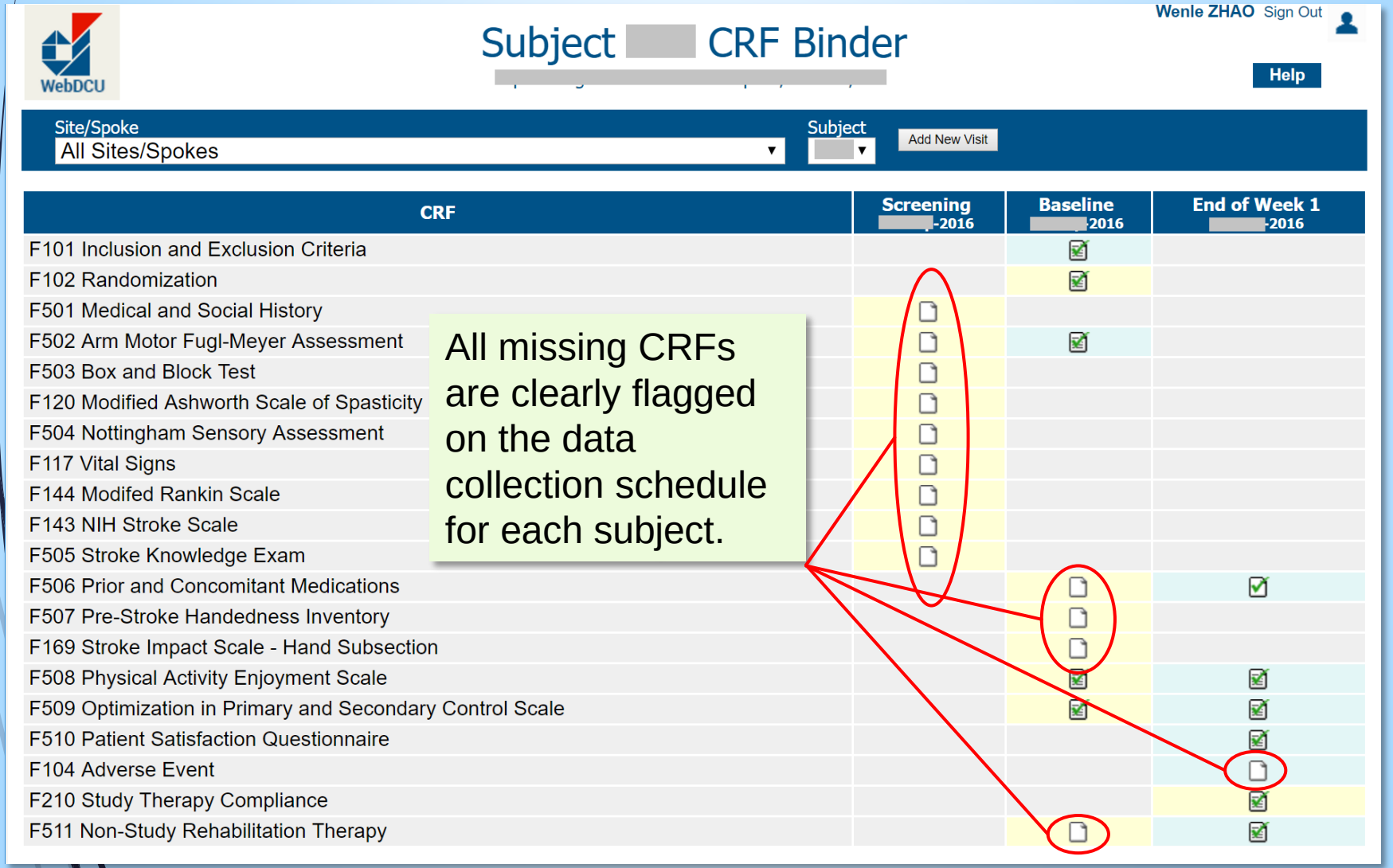
CRF Collection Schedule

CRF Name	Screening	Baseline / Randomization	Prior to Infusion 2	Prior to Infusion 3	24 Hour Post Last Infusion	Day 7 / Discharge	Day 30	Day 60	Day 90	Day 180	End of Study
F01 Inclusion Exclusion Criteria	X M										
F02 Randomization Form		X M									
F03 Prior Medications	X										
F04 Screening Visual and Auditory Assessment	X										
F05 Medical History Form	X M										
F06 Glasgow Coma Scale	X M	X	X	X	X	X					
F07 Modified Rankin Scale						X	X	X	X M	X	
F08 Labs	X				X						
F09 Demographics	X										
F10 Subject Intubation Log						X					
F11 Hospital Discharge						X					
F12 Treatment Confirmation		X									
F15 CT Scan	X				X						
F18 Blood Sample Collection Substudy		X			X						
F19 Study Drug Administration					X M						
F22 Vital Signs (BP & Pulse)		X	X	X	X	X					
F23 MoCA Scoring Summary						X	X		X M		
F24 Stroke Impact Scale-16							X		X M		
F25 Concomitant Medications						X	X	X	X		
F26 Concomitant Non-Drug Therapies						X	X	X	X		
F27 Adverse Events			O R M	O R M	O R M	O R M	O R M	O R M	O R M		
F29 End of Study											X M
F30 Visual and Auditory Assessment Follow Up			X	X	X	X	X				
F32 Additional Imaging	O				O	O	O	O	O		
F43 NIH Stroke Scale	X M	X	X	X	X	X	X		X M		
F99 MedWatch			O R C	O R C	O R C	O R C	O R C	O R C	O R C		

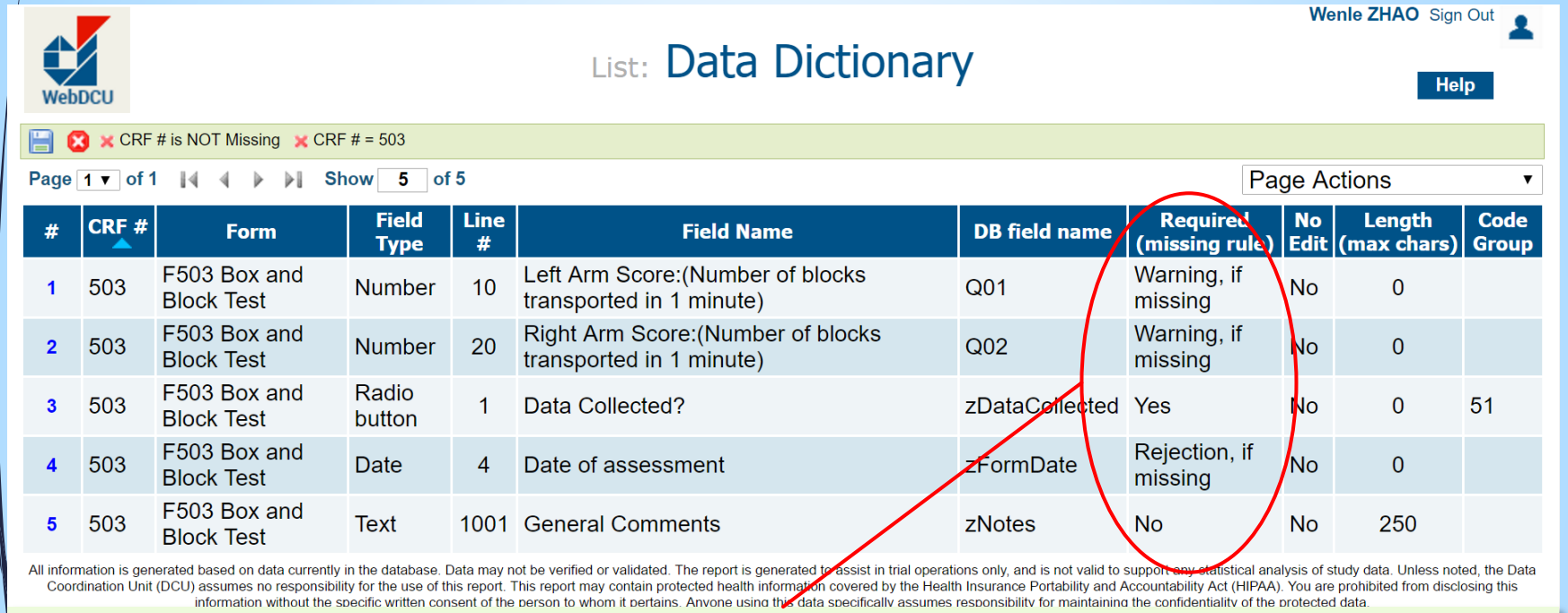
X: Required O: Optional R: Repeatable M: Monitor Verify Required C: Conditional

Well defined data collection schedule protects data integrity.

Example 8: Flag missing CRF



Example 9: Prevent missing data items



WebDCU

Wenle ZHAO Sign Out

List: Data Dictionary

Help

CRF # is NOT Missing | CRF # = 503

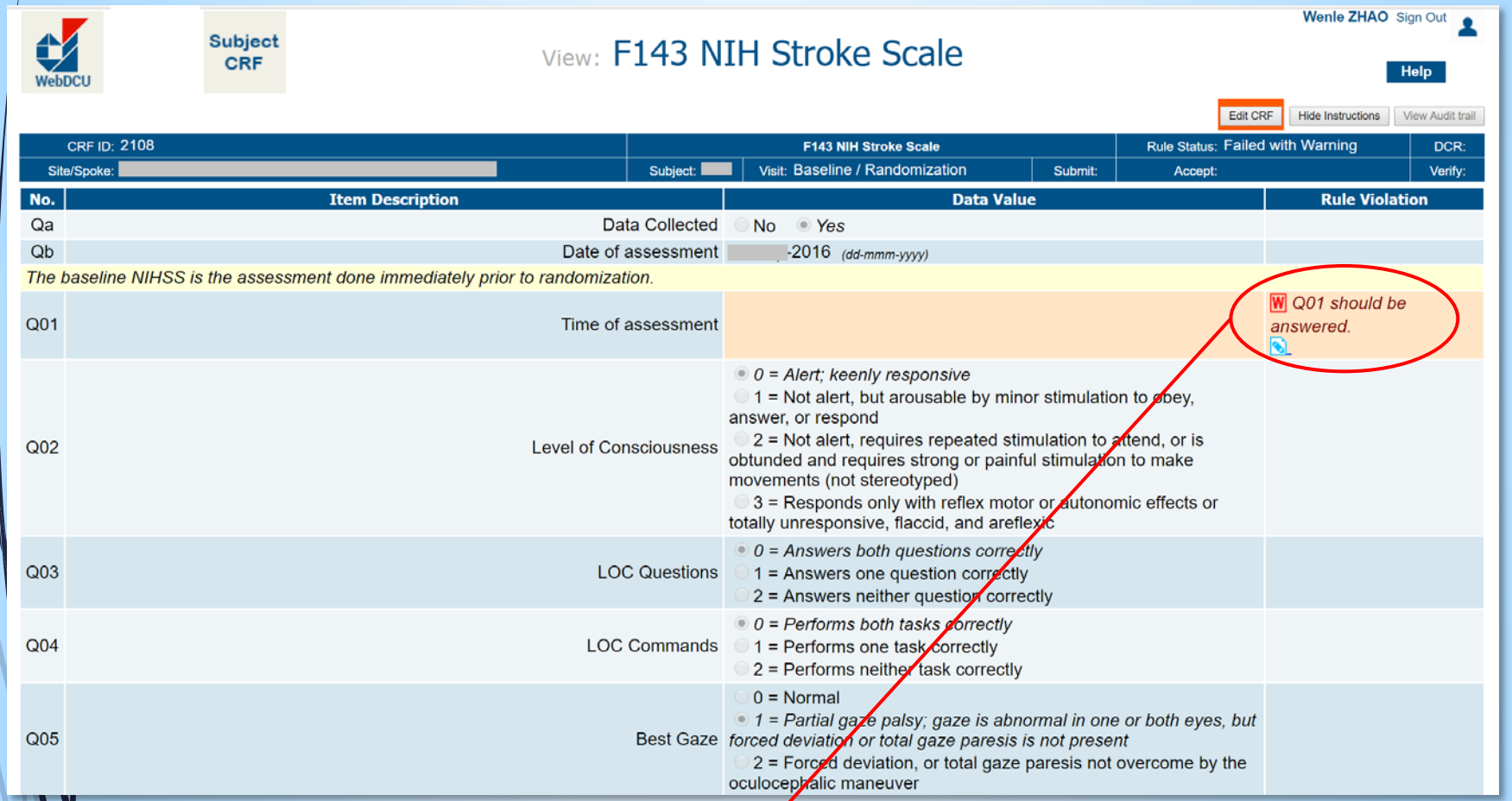
Page 1 of 1 Show 5 of 5 Page Actions

#	CRF #	Form	Field Type	Line #	Field Name	DB field name	Required (missing rule)	No Edit	Length (max chars)	Code Group
1	503	F503 Box and Block Test	Number	10	Left Arm Score:(Number of blocks transported in 1 minute)	Q01	Warning, if missing	No	0	
2	503	F503 Box and Block Test	Number	20	Right Arm Score:(Number of blocks transported in 1 minute)	Q02	Warning, if missing	No	0	
3	503	F503 Box and Block Test	Radio button	1	Data Collected?	zDataCollected	Yes	No	0	51
4	503	F503 Box and Block Test	Date	4	Date of assessment	zFormDate	Rejection, if missing	No	0	
5	503	F503 Box and Block Test	Text	1001	General Comments	zNotes	No	No	250	

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.

- ***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



WebDCU

Subject CRF

View: F143 NIH Stroke Scale

Wenle ZHAO Sign Out

Help

Edit CRF | Hide Instructions | View Audit trail

CRF ID: 2108	F143 NIH Stroke Scale			Rule Status: Failed with Warning	DCR:
Site/Spoke:	Subject:	Visit: Baseline / Randomization	Submit:	Accept:	Verify:

No.	Item Description	Data Value	Rule Violation
Qa	Data Collected	No *Yes*	
Qb	Date of assessment	-2016 *(dd-mmm-yyyy)*	
The baseline NIHSS is the assessment done immediately prior to randomization.			
Q01	Time of assessment		*Q01 should be answered.*
Q02	Level of Consciousness	*0 = Alert; keenly responsive* 1 = Not alert, but arousable by minor stimulation to obey, answer, or respond 2 = Not alert, requires repeated stimulation to attend, or is obtunded and requires strong or painful stimulation to make movements (not stereotyped) 3 = Responds only with reflex motor or autonomic effects or totally unresponsive, flaccid, and areflexic	
Q03	LOC Questions	*0 = Answers both questions correctly* 1 = Answers one question correctly 2 = Answers neither question correctly	
Q04	LOC Commands	*0 = Performs both tasks correctly* 1 = Performs one task correctly 2 = Performs neither task correctly	
Q05	Best Gaze	0 = Normal *1 = Partial gaze palsy; gaze is abnormal in one or both eyes, but forced deviation or total gaze paresis is not present* 2 = Forced deviation, or total gaze paresis not overcome by the oculocephalic maneuver	

- 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 month? 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

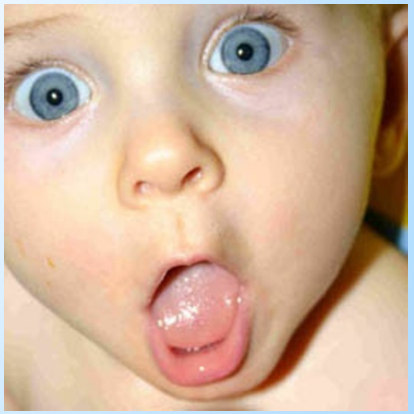
Make multiple choice question with sound logic

34

Q07: What color it is?

O Red
O Yellow
O Black

O Red
O Salmon
O Fire Brick
O Dark Orange
O Coral
O Tomato
O Orange Red
O Chocolate
O Saddle Brown
O Wheat
O Gold
O Tan



Avoid excessive list of options

Q01: Primary reason for screen failure (Please choose one only):

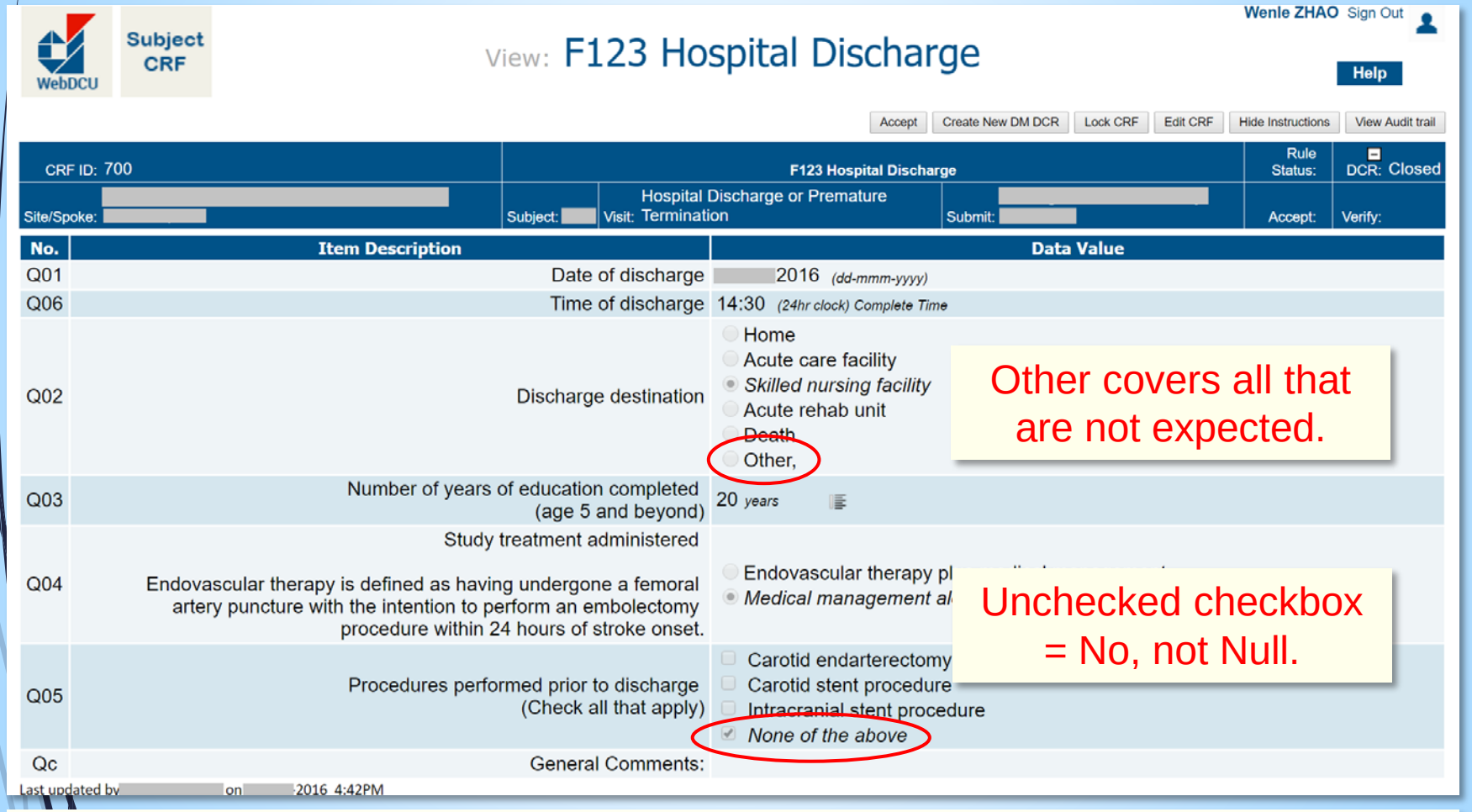
0 = Patient is not age 60 years or older
1 = Patient lacks capacity to provide informed consent
2 = Patient has language barriers
3 = Subject has lifetime history of bipolar affective disorder
4 = Subject has lifetime history of schizophrenia
5 = Subject has lifetime history of schizoaffective disorder
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 VLE

Q01: Primary reason for screen failure is:

Replace long list of radio button options with free text.

Have a central investigator do the grouping afterwards.

Example 11: Consider all possible responses



WebDCU | Subject CRF

Wenle ZHAO Sign Out

View: F123 Hospital Discharge

Help

Accept | Create New DM DCR | Lock CRF | Edit CRF | Hide Instructions | View Audit trail

CRF ID: 700	F123 Hospital Discharge			Rule Status:	DCR: Closed
Site/Spoke: [redacted]	Subject: [redacted]	Visit: Hospital Discharge or Premature Termination	Submit: [redacted]	Accept:	Verify:

No.	Item Description	Data Value
Q01	Date of discharge	[redacted]2016 *(dd-mmm-yyyy)*
Q06	Time of discharge	14:30 *(24hr clock) Complete Time*
Q02	Discharge destination	○ Home ○ Acute care facility ◉ *Skilled nursing facility* ○ Acute rehab unit ○ ~~Death~~ ○ Other,
Q03	Number of years of education completed (age 5 and beyond)	20 *years*
Q04	Study treatment administered Endovascular therapy is defined as having undergone a femoral artery puncture with the intention to perform an embolectomy procedure within 24 hours of stroke onset.	○ Endovascular therapy pl… ◉ *Medical management al…*
Q05	Procedures performed prior to discharge (Check all that apply)	☐ Carotid endarterectomy ☐ Carotid stent procedure ☐ Intracranial stent procedure ☑ *None of the above*
Qc	General Comments:	

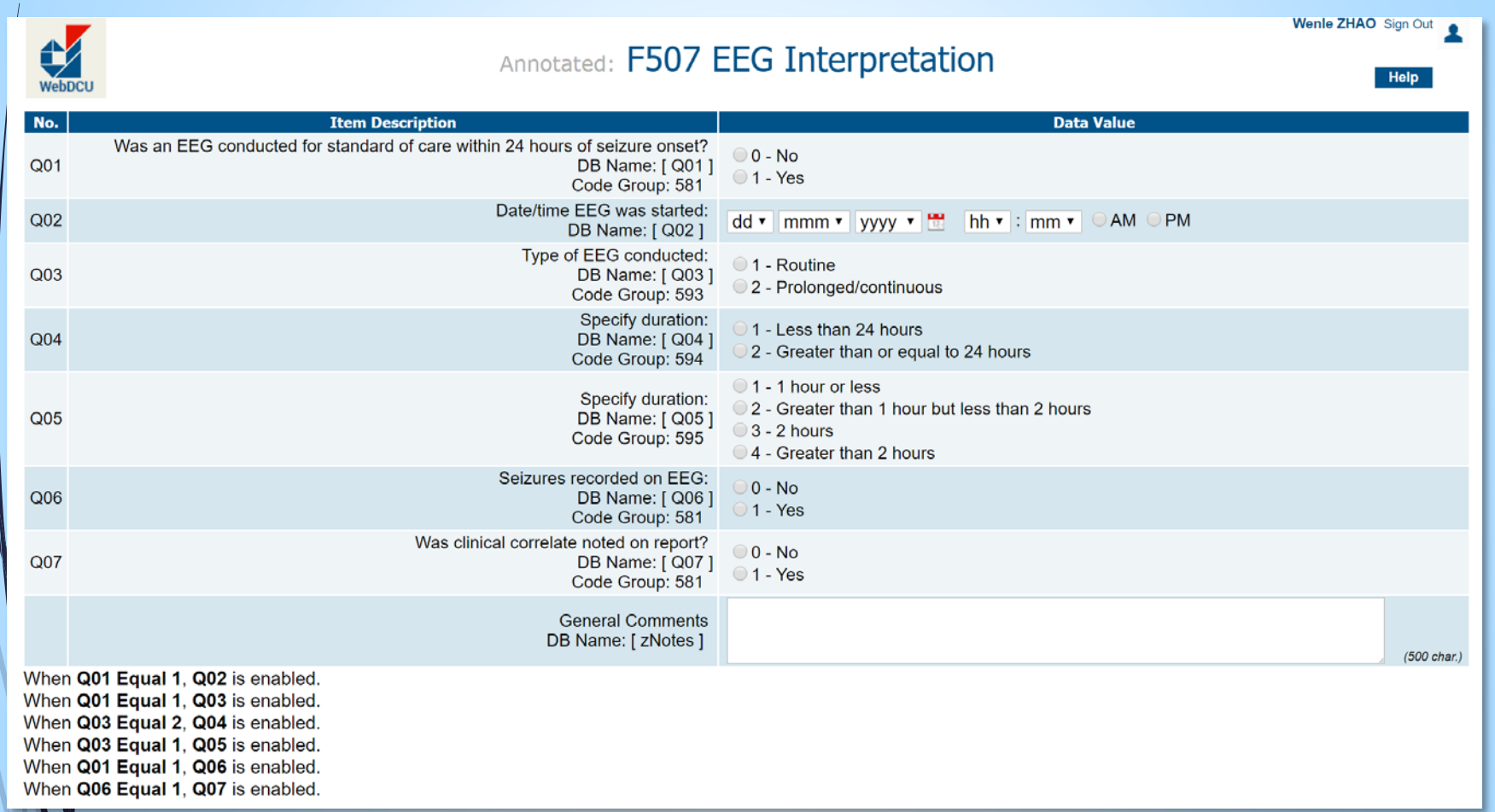
Last updated by [redacted] on [redacted] 2016 4:42PM

Other covers all that are not expected.

Unchecked checkbox = No, not Null.

- *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



WebDCU

Wenle ZHAO Sign Out

Annotated: F507 EEG Interpretation

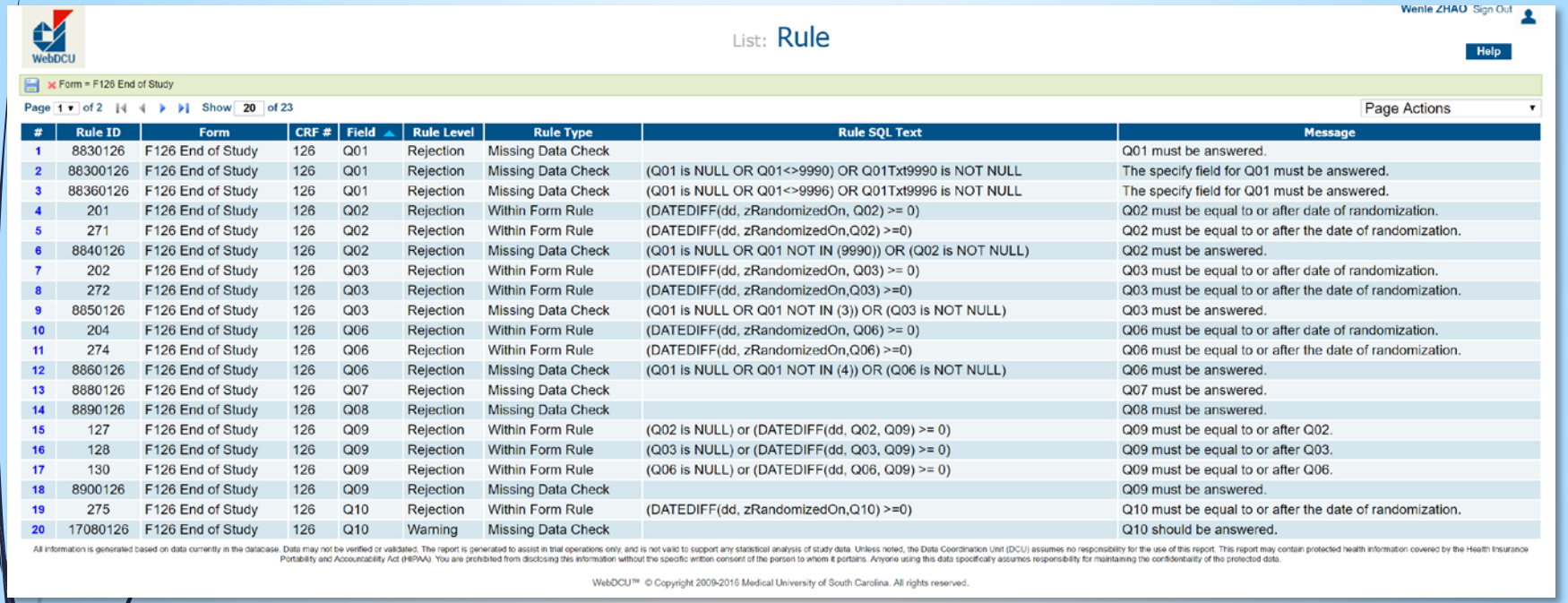
Help

No.	Item Description	Data Value
Q01	Was an EEG conducted for standard of care within 24 hours of seizure onset? DB Name: [Q01] Code Group: 581	0 - No 1 - Yes
Q02	Date/time EEG was started: DB Name: [Q02]	dd ▾ mmm ▾ yyyy ▾ hh ▾ : mm ▾ AM PM
Q03	Type of EEG conducted: DB Name: [Q03] Code Group: 593	1 - Routine 2 - Prolonged/continuous
Q04	Specify duration: DB Name: [Q04] Code Group: 594	1 - Less than 24 hours 2 - Greater than or equal to 24 hours
Q05	Specify duration: DB Name: [Q05] Code Group: 595	1 - 1 hour or less 2 - Greater than 1 hour but less than 2 hours 3 - 2 hours 4 - Greater than 2 hours
Q06	Seizures recorded on EEG: DB Name: [Q06] Code Group: 581	0 - No 1 - Yes
Q07	Was clinical correlate noted on report? DB Name: [Q07] Code Group: 581	0 - No 1 - Yes
	General Comments DB Name: [zNotes]	(500 char.)

When **Q01 Equal 1**, **Q02** is enabled.
When **Q01 Equal 1**, **Q03** is enabled.
When **Q03 Equal 2**, **Q04** is enabled.
When **Q03 Equal 1**, **Q05** is enabled.
When **Q01 Equal 1**, **Q06** is enabled.
When **Q06 Equal 1**, **Q07** is enabled.

- *Define skip pattern controls to avoid basic logic conflicts.*
- *Enforce skip patterns in EDC user interfaces.*

Example 13: Define data validation rules



WebDCU

Wenle ZHAO Sign Out

List: Rule

Help

Form = F126 End of Study

Page 1 of 2 Show 20 of 23

Page Actions

#	Rule ID	Form	CRF #	Field	Rule Level	Rule Type	Rule SQL Text	Message
1	8830126	F126 End of Study	126	Q01	Rejection	Missing Data Check		Q01 must be answered.
2	88300126	F126 End of Study	126	Q01	Rejection	Missing Data Check	(Q01 is NULL OR Q01<>9990) OR Q01Txt9990 is NOT NULL	The specify field for Q01 must be answered.
3	88360126	F126 End of Study	126	Q01	Rejection	Missing Data Check	(Q01 is NULL OR Q01<>9996) OR Q01Txt9996 is NOT NULL	The specify field for Q01 must be answered.
4	201	F126 End of Study	126	Q02	Rejection	Within Form Rule	(DATEDIFF(dd, zRandomizedOn, Q02) >= 0)	Q02 must be equal to or after date of randomization.
5	271	F126 End of Study	126	Q02	Rejection	Within Form Rule	(DATEDIFF(dd, zRandomizedOn,Q02) >=0)	Q02 must be equal to or after the date of randomization.
6	8840126	F126 End of Study	126	Q02	Rejection	Missing Data Check	(Q01 is NULL OR Q01 NOT IN (9990)) OR (Q02 is NOT NULL)	Q02 must be answered.
7	202	F126 End of Study	126	Q03	Rejection	Within Form Rule	(DATEDIFF(dd, zRandomizedOn, Q03) >= 0)	Q03 must be equal to or after date of randomization.
8	272	F126 End of Study	126	Q03	Rejection	Within Form Rule	(DATEDIFF(dd, zRandomizedOn,Q03) >=0)	Q03 must be equal to or after the date of randomization.
9	8850126	F126 End of Study	126	Q03	Rejection	Missing Data Check	(Q01 is NULL OR Q01 NOT IN (3)) OR (Q03 is NOT NULL)	Q03 must be answered.
10	204	F126 End of Study	126	Q06	Rejection	Within Form Rule	(DATEDIFF(dd, zRandomizedOn, Q06) >= 0)	Q06 must be equal to or after date of randomization.
11	274	F126 End of Study	126	Q06	Rejection	Within Form Rule	(DATEDIFF(dd, zRandomizedOn,Q06) >=0)	Q06 must be equal to or after the date of randomization.
12	8860126	F126 End of Study	126	Q06	Rejection	Missing Data Check	(Q01 is NULL OR Q01 NOT IN (4)) OR (Q06 is NOT NULL)	Q06 must be answered.
13	8880126	F126 End of Study	126	Q07	Rejection	Missing Data Check		Q07 must be answered.
14	8890126	F126 End of Study	126	Q08	Rejection	Missing Data Check		Q08 must be answered.
15	127	F126 End of Study	126	Q09	Rejection	Within Form Rule	(Q02 is NULL) or (DATEDIFF(dd, Q02, Q09) >= 0)	Q09 must be equal to or after Q02.
16	128	F126 End of Study	126	Q09	Rejection	Within Form Rule	(Q03 is NULL) or (DATEDIFF(dd, Q03, Q09) >= 0)	Q09 must be equal to or after Q03.
17	130	F126 End of Study	126	Q09	Rejection	Within Form Rule	(Q06 is NULL) or (DATEDIFF(dd, Q06, Q09) >= 0)	Q09 must be equal to or after Q06.
18	8900126	F126 End of Study	126	Q09	Rejection	Missing Data Check		Q09 must be answered.
19	275	F126 End of Study	126	Q10	Rejection	Within Form Rule	(DATEDIFF(dd, zRandomizedOn,Q10) >=0)	Q10 must be equal to or after the date of randomization.
20	17080126	F126 End of Study	126	Q10	Warning	Missing Data Check		Q10 should be answered.

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.

Use validated assessments

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)

CRF change control

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.

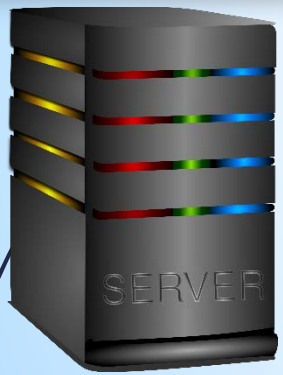
General strategy for CRF changes:

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 re-assessment.

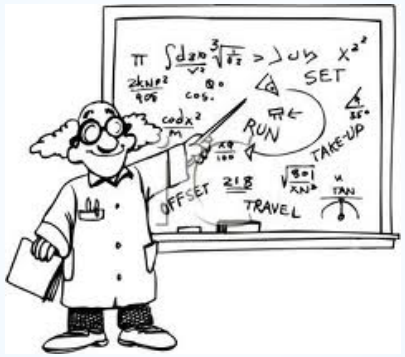
Consider the workload distribution

Employee of the year
2015
Mr. SQL Server

Employee of the year
2016
Mr. SQL Server



Maximize work for
Mr. SQL Server



Reduce work for
Dr. P-value

Minimize burden for
Ms. CDM

Action #3: prevent fake data and data fraud

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.

Summary – Section 2

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

Monitor the study progress

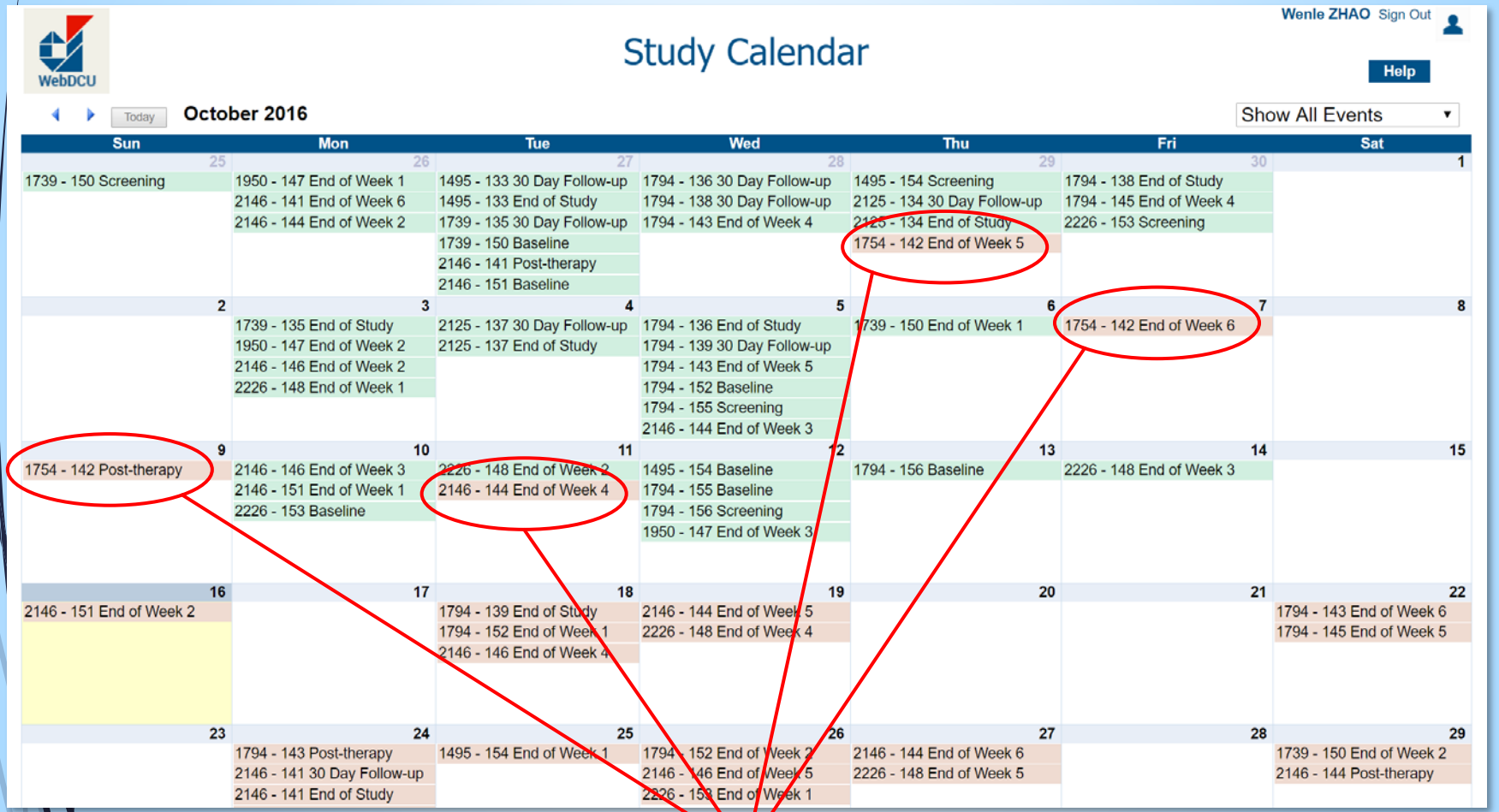
1. Monitor subject study progress. Ensure compliance with study treatments and assessments specified in the protocol.
2. Monitor subject CRF submission timeliness. Ensure data completeness.
3. Review CRF rule violation and protocol violation reports.
4. Review CRF data that can not be validated by computer rules.

Risk-based monitoring strategy

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

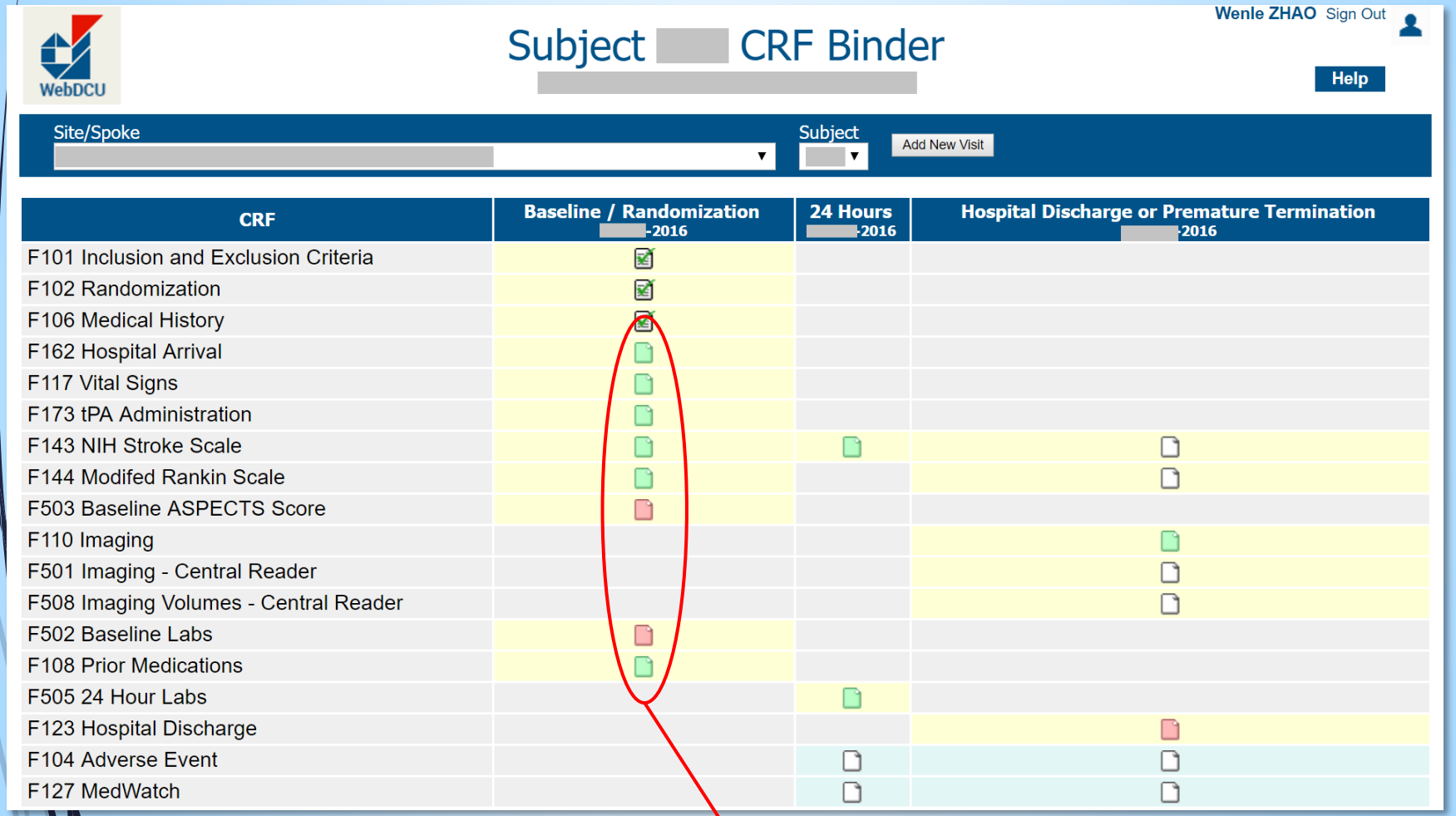
1. Eligibility CRF
2. Randomization CRF
3. Study treatment CRF
4. Adverse event CRF
5. End of study CRF
6. CRFs with primary – secondary efficacy outcomes
7. CRFs with confirmed protocol violations

Example 14: Subject study visit monitoring



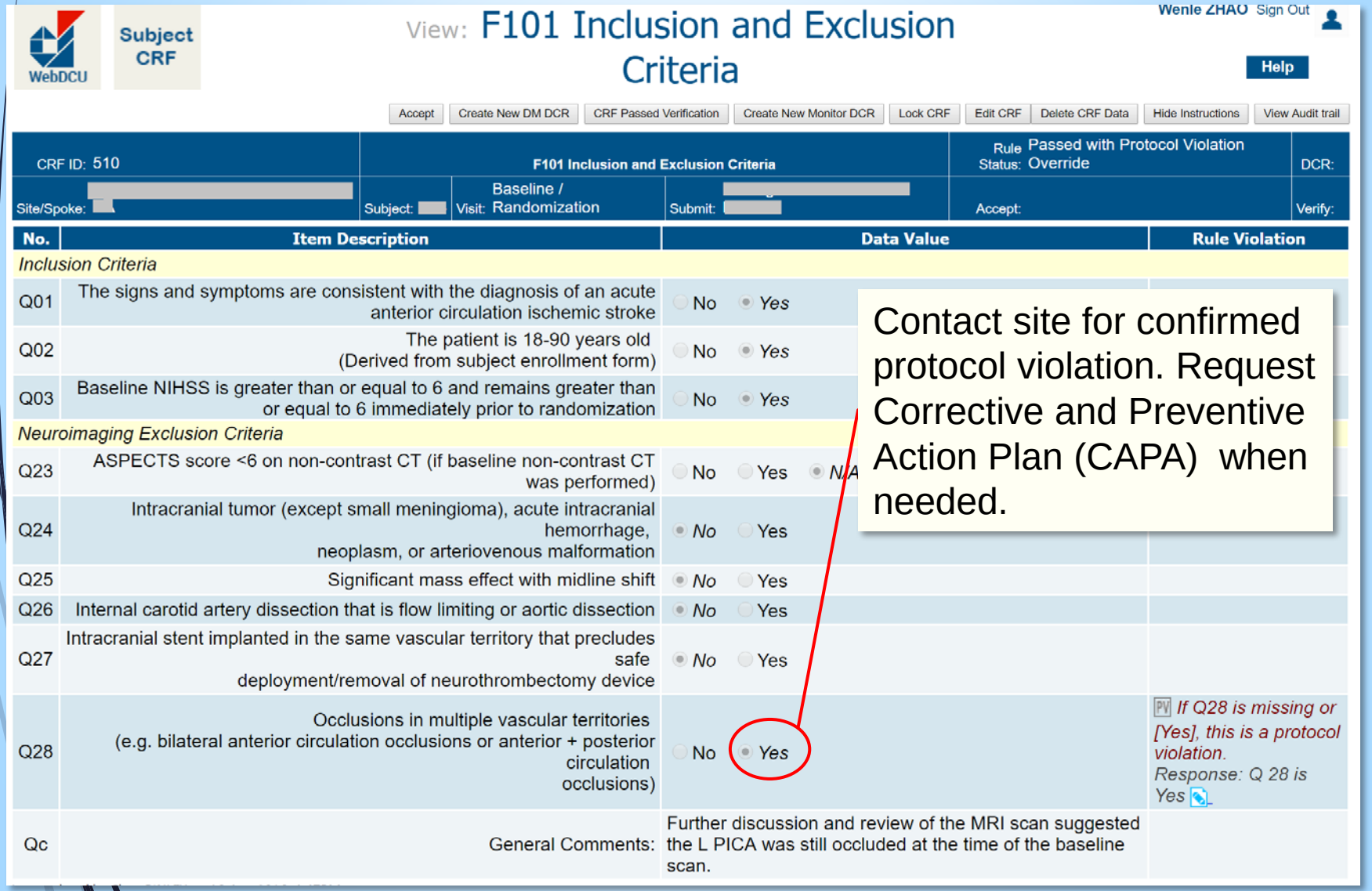
Take immediate action for overdue study visits.

Example 15: CRF completion status monitoring



Contact site study coordinator in cases of delayed CRF submission.

Example 16: Protocol violation review



WebDCU

Subject CRF

View: F101 Inclusion and Exclusion Criteria

Wenle ZHAO Sign Out

Help

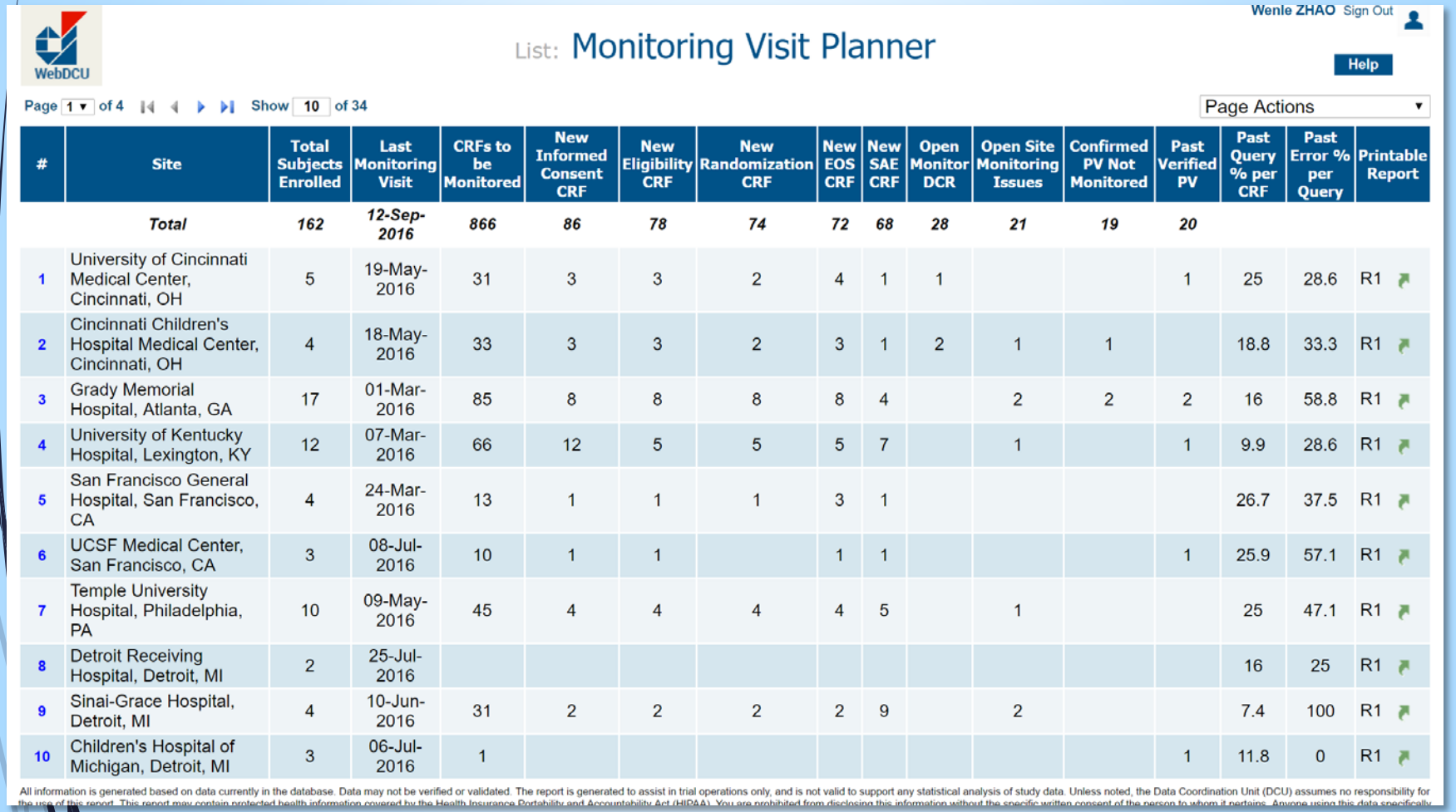
Accept | Create New DM DCR | CRF Passed Verification | Create New Monitor DCR | Lock CRF | Edit CRF | Delete CRF Data | Hide Instructions | View Audit trail

CRF ID: 510	F101 Inclusion and Exclusion Criteria		Rule Status: Passed with Protocol Violation Override	DCR:
Site/Spoke: [redacted]	Subject: [redacted]	Visit: Baseline / Randomization	Submit: [redacted]	Accept:	Verify:

No.	Item Description	Data Value	Rule Violation
Inclusion Criteria			
Q01	The signs and symptoms are consistent with the diagnosis of an acute anterior circulation ischemic stroke	No / *Yes*	
Q02	The patient is 18-90 years old (Derived from subject enrollment form)	No / *Yes*	
Q03	Baseline NIHSS is greater than or equal to 6 and remains greater than or equal to 6 immediately prior to randomization	No / *Yes*	
Neuroimaging Exclusion Criteria			
Q23	ASPECTS score <6 on non-contrast CT (if baseline non-contrast CT was performed)	No / Yes / *N/A*	
Q24	Intracranial tumor (except small meningioma), acute intracranial hemorrhage, neoplasm, or arteriovenous malformation	*No* / Yes	
Q25	Significant mass effect with midline shift	*No* / Yes	
Q26	Internal carotid artery dissection that is flow limiting or aortic dissection	*No* / Yes	
Q27	Intracranial stent implanted in the same vascular territory that precludes safe deployment/removal of neurothrombectomy device	*No* / Yes	
Q28	Occlusions in multiple vascular territories (e.g. bilateral anterior circulation occlusions or anterior + posterior circulation occlusions)	No / *Yes*	PV *If Q28 is missing or [Yes], this is a protocol violation.* *Response: Q 28 is Yes*
Qc	General Comments:	Further discussion and review of the MRI scan suggested the L PICA was still occluded at the time of the baseline scan.	

Contact site for confirmed protocol violation. Request Corrective and Preventive Action Plan (CAPA) when needed.

Example 17: Plan on-site monitoring visits



WebDCU

Wenle ZHAO Sign Out

List: Monitoring Visit Planner

Help

Page 1 of 4 Show 10 of 34

Page Actions

#	Site	Total Subjects Enrolled	Last Monitoring Visit	CRFs to be Monitored	New Informed Consent CRF	New Eligibility CRF	New Randomization CRF	New EOS CRF	New SAE CRF	Open Monitor DCR	Open Site Monitoring Issues	Confirmed PV Not Monitored	Past Verified PV	Past Query % per CRF	Past Error % per Query	Printable Report
	Total	*162*	*12-Sep-2016*	*866*	*86*	*78*	*74*	*72*	*68*	*28*	*21*	*19*	*20*			
1	University of Cincinnati Medical Center, Cincinnati, OH	5	19-May-2016	31	3	3	2	4	1	1			1	25	28.6	R1
2	Cincinnati Children's Hospital Medical Center, Cincinnati, OH	4	18-May-2016	33	3	3	2	3	1	2	1	1		18.8	33.3	R1
3	Grady Memorial Hospital, Atlanta, GA	17	01-Mar-2016	85	8	8	8	8	4		2	2	2	16	58.8	R1
4	University of Kentucky Hospital, Lexington, KY	12	07-Mar-2016	66	12	5	5	5	7		1		1	9.9	28.6	R1
5	San Francisco General Hospital, San Francisco, CA	4	24-Mar-2016	13	1	1	1	3	1					26.7	37.5	R1
6	UCSF Medical Center, San Francisco, CA	3	08-Jul-2016	10	1	1		1	1				1	25.9	57.1	R1
7	Temple University Hospital, Philadelphia, PA	10	09-May-2016	45	4	4	4	4	5		1			25	47.1	R1
8	Detroit Receiving Hospital, Detroit, MI	2	25-Jul-2016											16	25	R1
9	Sinai-Grace Hospital, Detroit, MI	4	10-Jun-2016	31	2	2	2	2	9		2			7.4	100	R1
10	Children's Hospital of Michigan, Detroit, MI	3	06-Jul-2016	1									1	11.8	0	R1

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

Check number of CRFs for risk-based monitoring.

Example 18: Central data quality monitoring

WebDCU

Wenle ZHAO Sign Out

Summary: F505 Demographics

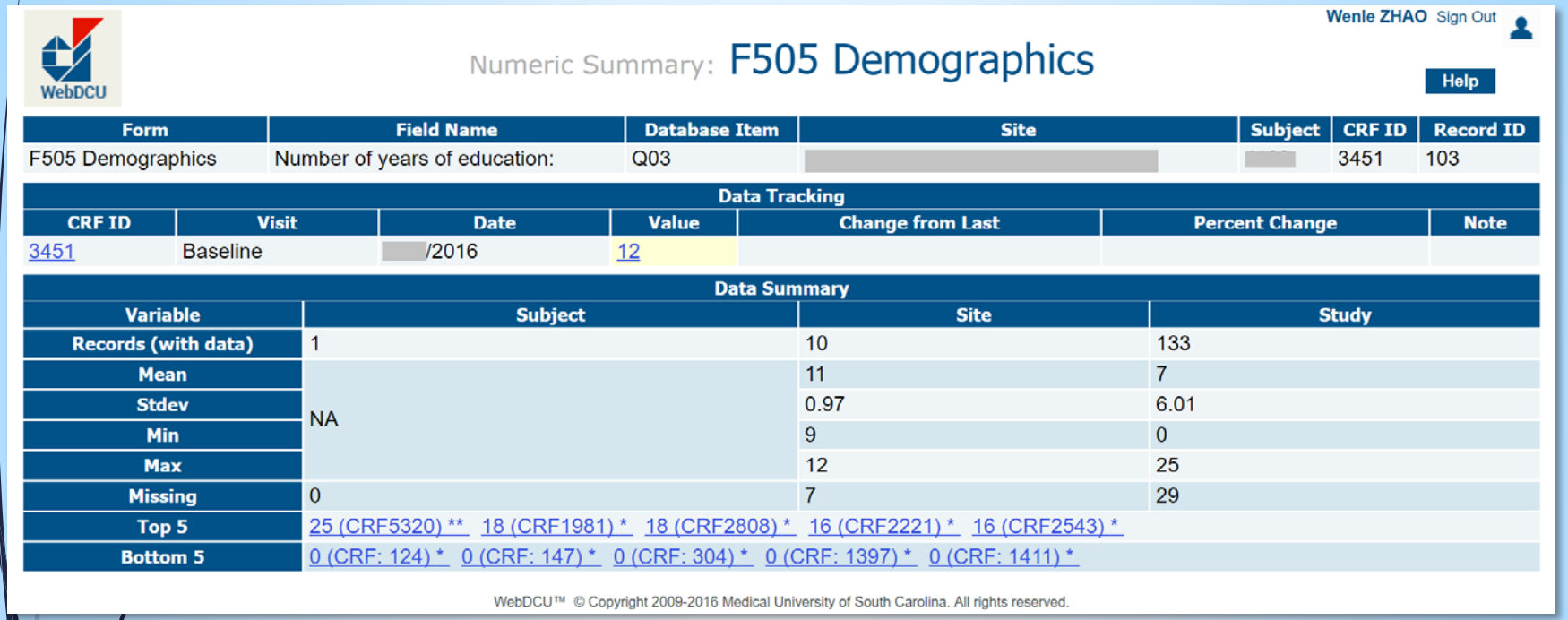
Help

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

DB Field	Data Type	Field Definition	Unit	With Data	Missing	Mean	Stdev	Median	Min	Max	Total
Q03	Number	Number of years of education:		133	29	7	6.01	9	0	25	943

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

Example 19: Central data quality monitoring



WebDCU

Wenle ZHAO Sign Out

Numeric Summary: F505 Demographics

Help

Form	Field Name	Database Item	Site	Subject	CRF ID	Record ID
F505 Demographics	Number of years of education:	Q03			3451	103

Data Tracking

CRF ID	Visit	Date	Value	Change from Last	Percent Change	Note
3451	Baseline	/2016	12			

Data Summary

Variable	Subject	Site	Study
Records (with data)	1	10	133
Mean	NA	11	7
Stdev		0.97	6.01
Min		9	0
Max		12	25
Missing	0	7	29
Top 5	25 (CRF5320) ** 18 (CRF1981) * 18 (CRF2808) * 16 (CRF2221) * 16 (CRF2543) *		
Bottom 5	0 (CRF: 124) * 0 (CRF: 147) * 0 (CRF: 304) * 0 (CRF: 1397) * 0 (CRF: 1411) *		

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

- 1st and 15th day in a month; January and July in a year.
- Rounded to 0 or 5.

Data comparison over time

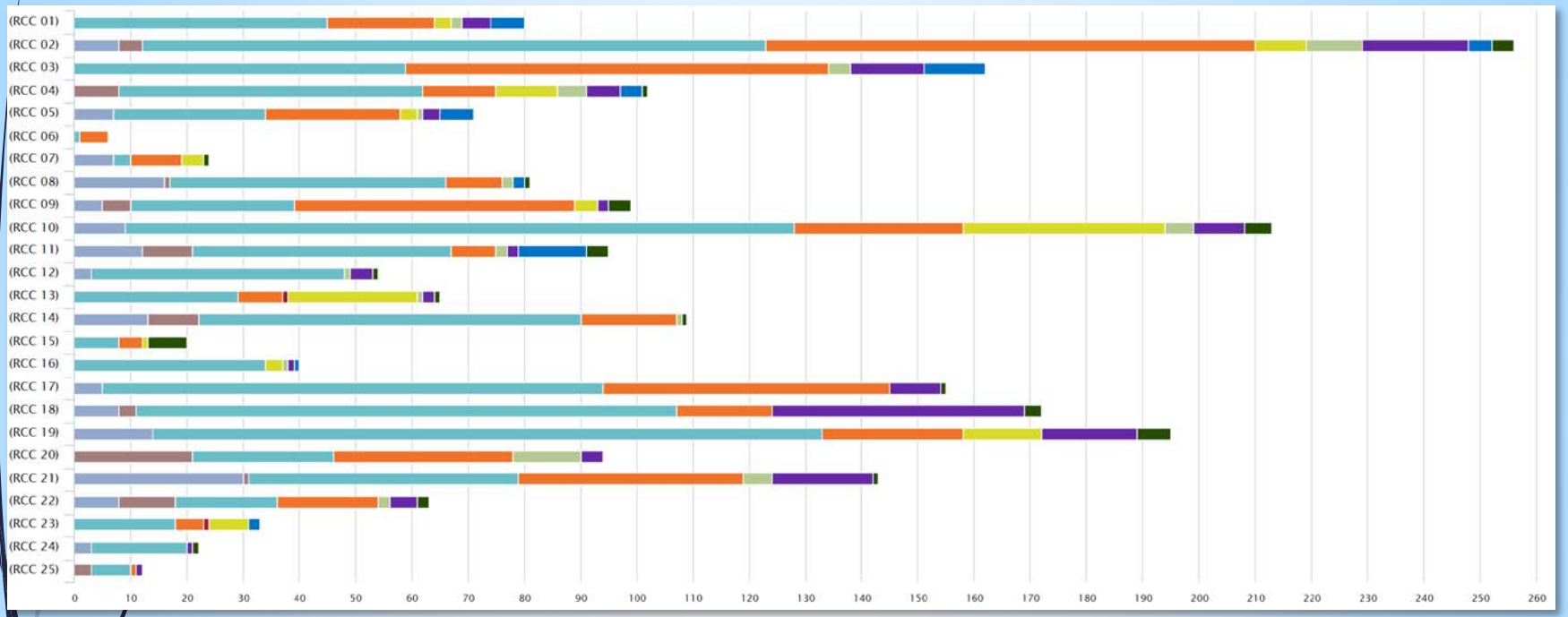
- Loosened eligibility criteria due to slow enrollment.
- Low site team member retention rate.
- Performance disparity among sites.
- Slow site enrollment ➔ less familiar with study protocol.
- Logic error among data on several CRFs.

Use trial operation performance dashboard

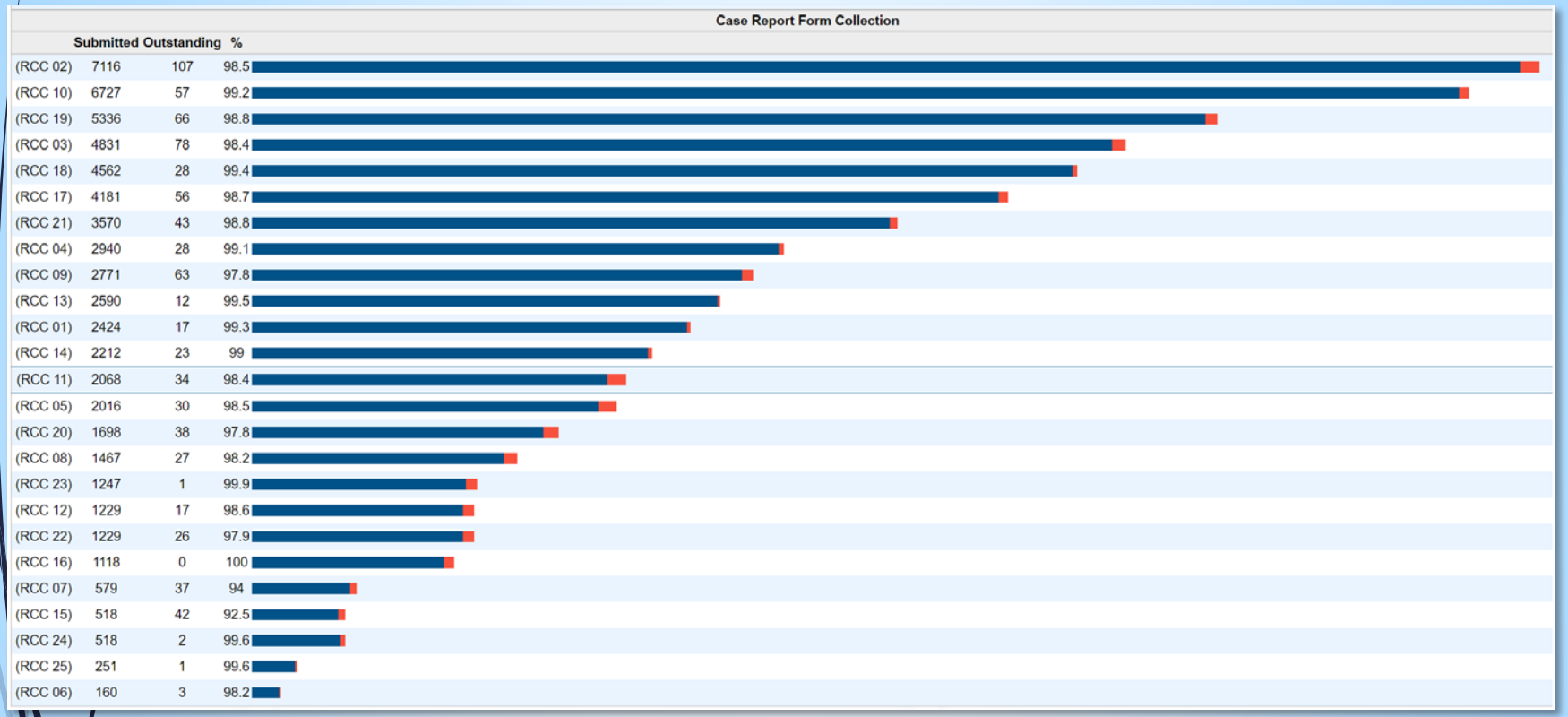
Share data quality report within the study community with information on:

- 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 - Enrollment



Trial performance dashboard – CRF submit



Summary – Section 3

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: zhaow@musc.edu