



CRF Design for Data Standards

David A. Scocca

dave_scocca@rhoworld.com

CRF Design for Data Standards

- Controlled Terminology
- Epochs and Dispositions
- Dates and Times
- Free Text Fields
- Avoiding Unwanted Data
- Collecting Necessary Context

Where CDASH Fits In

- Standard for how to collect data, not for what data to collect
- Provides forms, not a structure within which to present the forms

Controlled Terminology

- Standardized lists of specific values for data
 - Use standard terminology wherever possible
 - Main SDTM list, with CDASH and ADaM extensions
- Current CT workbook is always available from National Cancer Institute
 - <https://www.cancer.gov/research/resources/terminology/cdisc>
- Many lists are “extensible”, but not with synonyms
 - e.g, cannot use “cc” in place of “mL”

Controlled Terminology

SDTM Terminology 2017-12-22.xls [Compatibility Mode] - Microsoft Excel

	A	B	C	D	E	F
	Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
1	C102038	C101806		Abnormal Involuntary Movement Scale Questionnaire Test Name	AIMS01-Upper Extremities	AIMS01-Upper Extremities
47						
48	C66767		No	Action Taken with Study Treatment	ACN	Action Taken with Study Treatment
49	C49503	C66767		Action Taken with Study Treatment	DOSE INCREASED	
50	C49504	C66767		Action Taken with Study Treatment	DOSE NOT CHANGED	
51	C49505	C66767		Action Taken with Study Treatment	DOSE REDUCED	
52	C49501	C66767		Action Taken with Study Treatment	DRUG INTERRUPTED	
53	C49502	C66767		Action Taken with Study Treatment	DRUG WITHDRAWN	
54	C48660	C66767		Action Taken with Study Treatment	NOT APPLICABLE	NA; Not Applicable
55	C17998	C66767		Action Taken with Study Treatment	UNKNOWN	U; Unknown
56	C101865		No	Acute Coronary Syndrome Presentation Category	ACSPCAT	Acute Coronary Syndrome Presentation Category
57	C80383	C101865		Acute Coronary Syndrome Presentation Category	NON-ST ELEVATION MYOCARDIAL INFARCTION	NON-STEMI; NSTEMI
58	C17649	C101865		Acute Coronary Syndrome Presentation Category	OTHER	Other
59	C101888	C101865		Acute Coronary Syndrome Presentation Category	ST ELEVATION MYOCARDIAL INFARCTION	STEMI

Ready | ReadMe | SDTM Terminology 2017-12-22 | 100%

Study Epoch

- Defined: “a period of time that serves a purpose in the trial as a whole”
- Expected for all events, observations, or findings during the trial
- Does not depend on arm — can be determined while data is blinded
- Uses controlled terminology
- No overlap

Epoch Terminology

- SCREENING
- BASELINE
- TREATMENT
- FOLLOW-UP
- RUN-IN
- WASHOUT
- INDUCTION TREATMENT
- CONTINUATION TREATMENT
- BLINDED TREATMENT
- OPEN LABEL TREATMENT
- LONG-TERM FOLLOW-UP

Epoch: What We Need to Know

- What events define the start and end of each epoch?
 - Informed Consent
 - First Dose
 - Last Dose
- What epoch is each observation in?
 - Easy when visit and epoch boundaries match (e.g., Day 1 Predose and Day 1 Dose)
 - Difficult when it requires date comparisons

Dispositions

- What happens to the subject at the end of a study or study period
- “What was the subject’s status?” (CDASH)
 - Either “COMPLETED” or a standardized reason for non-completion
 - Completion determined as per protocol
 - Check the Controlled Terminology
- Can be collected at the end of each epoch

Unscheduled Visits

- Unscheduled does not mean unexpected
- What we need to know:
 - In what epoch is the data being collected?
 - What was the previous scheduled visit?
 - What is the purpose of the unscheduled visit?
 - What was collected together?
- Ideally, allow multiple unscheduled visits so that each visit can be grouped together

Special Case: Rescreening

- Re-screening is not handled well by the data standards
- This is actually a special kind of unscheduled visit
- Use multiple screening visit forms
 - Having a single subject ID with multiple screening visits is much better than establishing an additional ID when a subject is rescreened

Collecting Dates (and Times)

- Events should have a collection date
- Forms should clearly distinguish between a data collection date and a finding date
 - Date of sample collection
 - Date of imaging
 - Date of event/treatment
- Collect times when necessary
 - PK studies
 - Findings repeated within a day

Free Text Fields: Just Say No

- Data for medical coding
 - AETERM, CMTRT, MHTERM
- Implicitly coded data
 - Multiple choice with a “specify” option for some answers
- Comments
 - No reliable way to analyze text content
 - Comment domain is basically a “dumping ground”
 - Real data can be missed or lost

Unwanted Data: Con Meds

- Collect only what you need
- The general intervention form allows for a lot of information
 - Treatment, dose, units, route, frequency, form, indication, start date, stop date
- What is your statistician actually going to do with this data?
 - Collecting, cleaning and tabulating unwanted data is expensive!
 - Context for medical coders

Unwanted Data: Con Meds

- Do we just need to know what the med was and when it was used?
- Can we generalize some choices?
 - Indication: Adverse Event vs. underlying condition vs. Other
- Consider using two CM forms
 - Pain Medications vs. Other Medications
 - Use Indication to validate what goes where

Context: Study Criteria

- Inclusion and Exclusion criteria change over the course of almost every study with protocol revisions.
 - Criteria must be identified at the same level of granularity
 - What protocol version did a subject consent under?
 - Same exact criterion must get same exact result; any different or changed criterion must get something different

CRF Design for Data Standards

- Standardized data is what will be submitted to the FDA or other agency
- We want to make converting the data for submission as straightforward as possible
- Sub-optimal CRF design can cost a lot of time and effort in later parts of the study
- Involve SDTM experts in your CRF development process