Creating Transparent SDTM-Based Datasets

Jerry Salyers
Senior Consultant

Richard Lewis
Senior Consultant

Fred Wood
Senior Manager and Lead

Data Standards Consulting
Accenture Accelerated R&D Services

High performance. Delivered.

Outline

- What is meant by Transparency within the larger world of drug development?
  - Transparency at the program or study level
  - Data Transparency at the subject level
- Mapping of stored operational study data to SDTM
- Documenting the transparent mapping of study data and metadata for ease of review
  - Define file
  - BlankCRF
  - Study Data Reviewer's Guide
- Conclusions

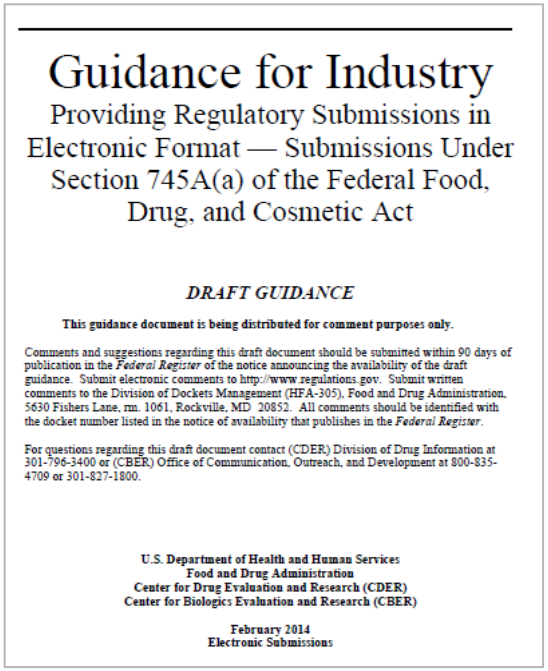
Data Transparency – A High Level look

- In Europe, AllTrials petition to create laws requiring not just the registration of the trial, but the publishing of results.
  - Recently adopted EMA policy to publish clinical study reports; goes into effect 01 Jan 2015
  - Many pharmaceutical companies have declared their support for this initiative
  - Future EMA plans are to make subject-level data available; ensure patient privacy is adequately protected
- Some Pharma companies have their own public registries dedicated to greater clinical trial transparency
  - Results published in peer-review journals whenever possible

Data Transparency – Subject Level Data

- Making data that sits behind the study results available to researchers
- A few companies pioneered this approach within their own trial registries
- These companies and sponsors have now configured a dedicated system that can be used to access anonymized data across sponsors for further analysis and research
    - www.clinicalstudydatarequest.com
- Bayer, Boehringer-Ingelheim, GSK, Lilly, Novartis among others have committed to using this site

New FDA Draft Documents (1)



Guidance for Industry

Providing Regulatory Submissions in Electronic Format — Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

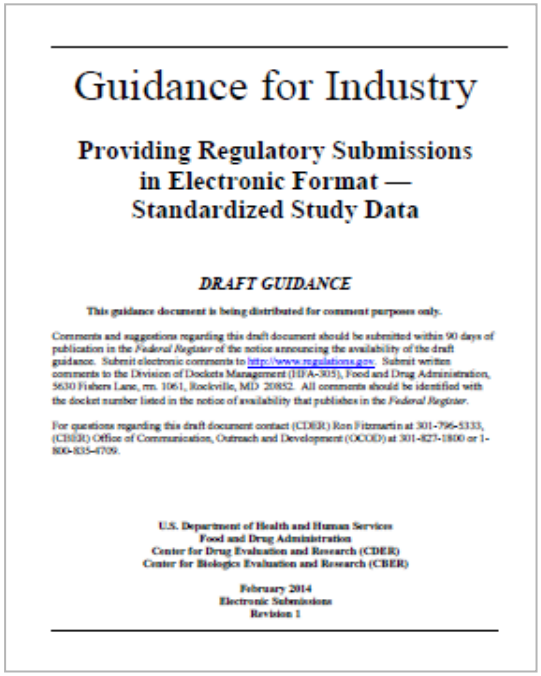
Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to http://www.regulations.gov. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (CDER) Division of Drug Information at 301-796-3400 or (CBER) Office of Communication, Outreach, and Development at 800-835-4709 or 301-827-1800.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

February 2014
Electronic Submissions

Guidance for Industry:
Providing Regulatory Submissions in Electronic Format – Submissions Under Section 745A (a) of the Federal Food Drug, and Cosmetic Act



Guidance for Industry

Providing Regulatory Submissions in Electronic Format — Standardized Study Data

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

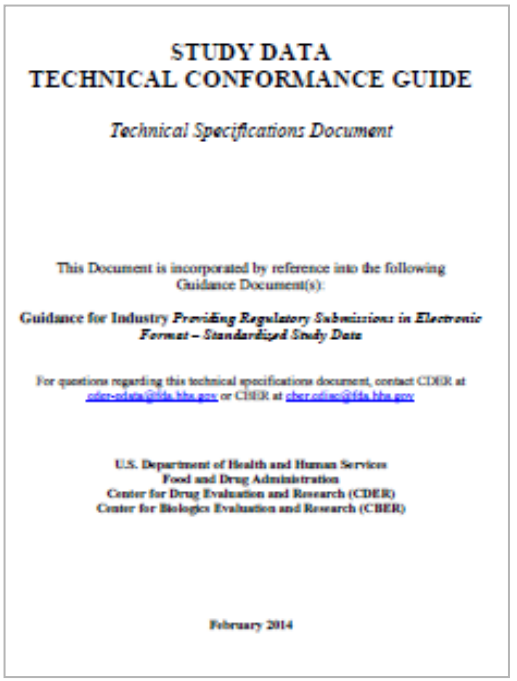
Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to http://www.regulations.gov. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (CDER) Ron Fitzmartin at 301-796-5333, (CBER) Office of Communication, Outreach and Development (OCOD) at 301-827-1800 or 1-800-835-4709.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

February 2014
Electronic Submissions
Revision 1

Guidance for Industry:
Providing Regulatory Submissions in Electronic Format – Standardized Study Data



STUDY DATA
TECHNICAL CONFORMANCE GUIDE

Technical Specifications Document

This Document is incorporated by reference into the following Guidance Document(s):

Guidance for Industry *Providing Regulatory Submissions in Electronic Format – Standardized Study Data*

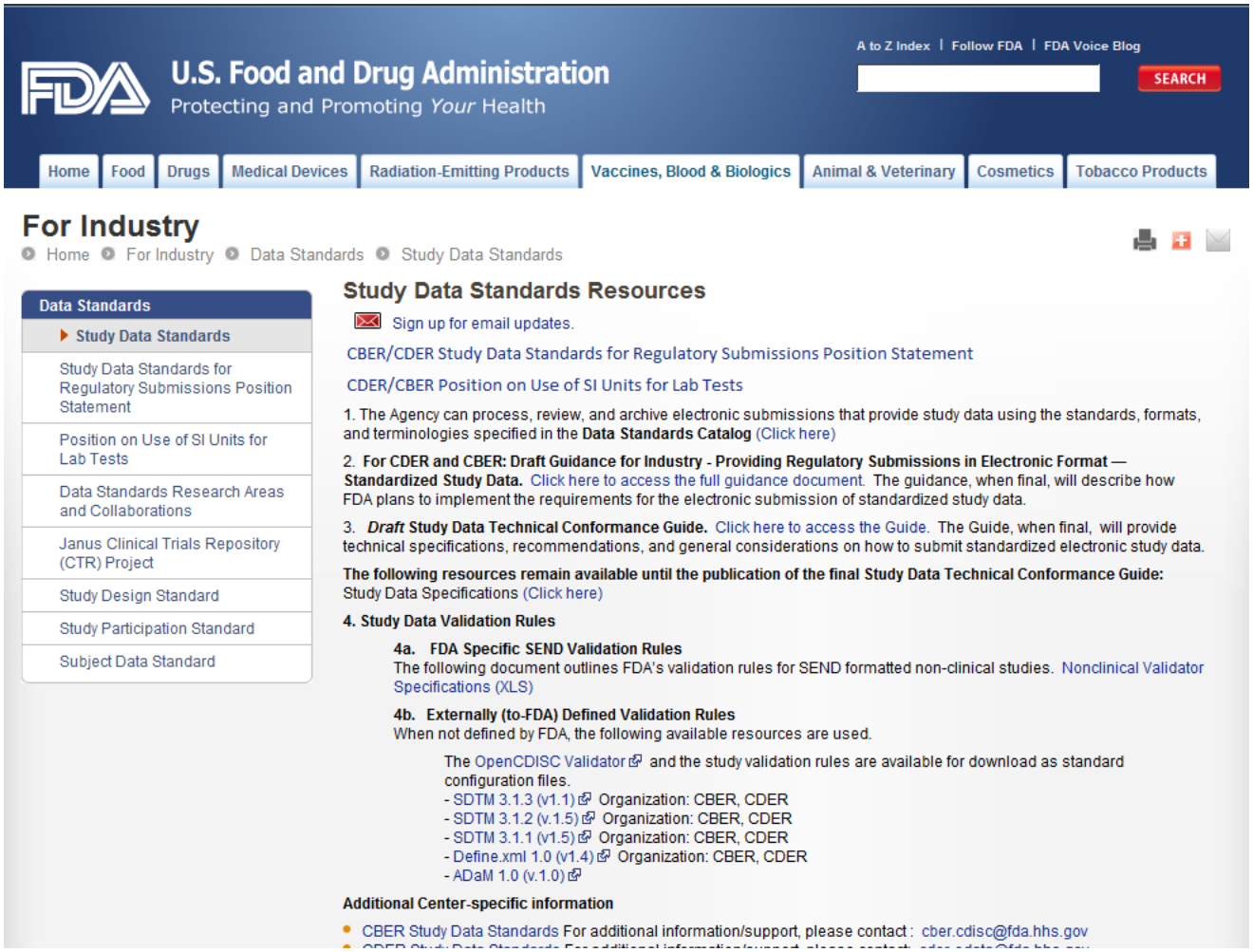
For questions regarding this technical specifications document, contact CDER at cder-sdata@fda.hhs.gov or CBER at cber.ofdsc@fda.hhs.gov

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

February 2014

Study Data Technical Conformance Guide

Study Data Standards Resources



U.S. Food and Drug Administration
Protecting and Promoting *Your* Health

A to Z Index | Follow FDA | FDA Voice Blog
SEARCH

Home | Food | Drugs | Medical Devices | Radiation-Emitting Products | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Tobacco Products

For Industry

Home › For Industry › Data Standards › Study Data Standards

Data Standards

- Study Data Standards
- Study Data Standards for Regulatory Submissions Position Statement
- Position on Use of SI Units for Lab Tests
- Data Standards Research Areas and Collaborations
- Janus Clinical Trials Repository (CTR) Project
- Study Design Standard
- Study Participation Standard
- Subject Data Standard

Study Data Standards Resources

Sign up for email updates.

CBER/CDER Study Data Standards for Regulatory Submissions Position Statement

CDER/CBER Position on Use of SI Units for Lab Tests

1. The Agency can process, review, and archive electronic submissions that provide study data using the standards, formats, and terminologies specified in the **Data Standards Catalog** (Click here)

2. **For CDER and CBER: Draft Guidance for Industry - Providing Regulatory Submissions in Electronic Format — Standardized Study Data.** Click here to access the full guidance document. The guidance, when final, will describe how FDA plans to implement the requirements for the electronic submission of standardized study data.

3. ***Draft* Study Data Technical Conformance Guide.** Click here to access the Guide. The Guide, when final, will provide technical specifications, recommendations, and general considerations on how to submit standardized electronic study data.

The following resources remain available until the publication of the final Study Data Technical Conformance Guide:
Study Data Specifications (Click here)

4. Study Data Validation Rules

4a. FDA Specific SEND Validation Rules
The following document outlines FDA's validation rules for SEND formatted non-clinical studies. Nonclinical Validator Specifications (XLS)

4b. Externally (to-FDA) Defined Validation Rules
When not defined by FDA, the following available resources are used.

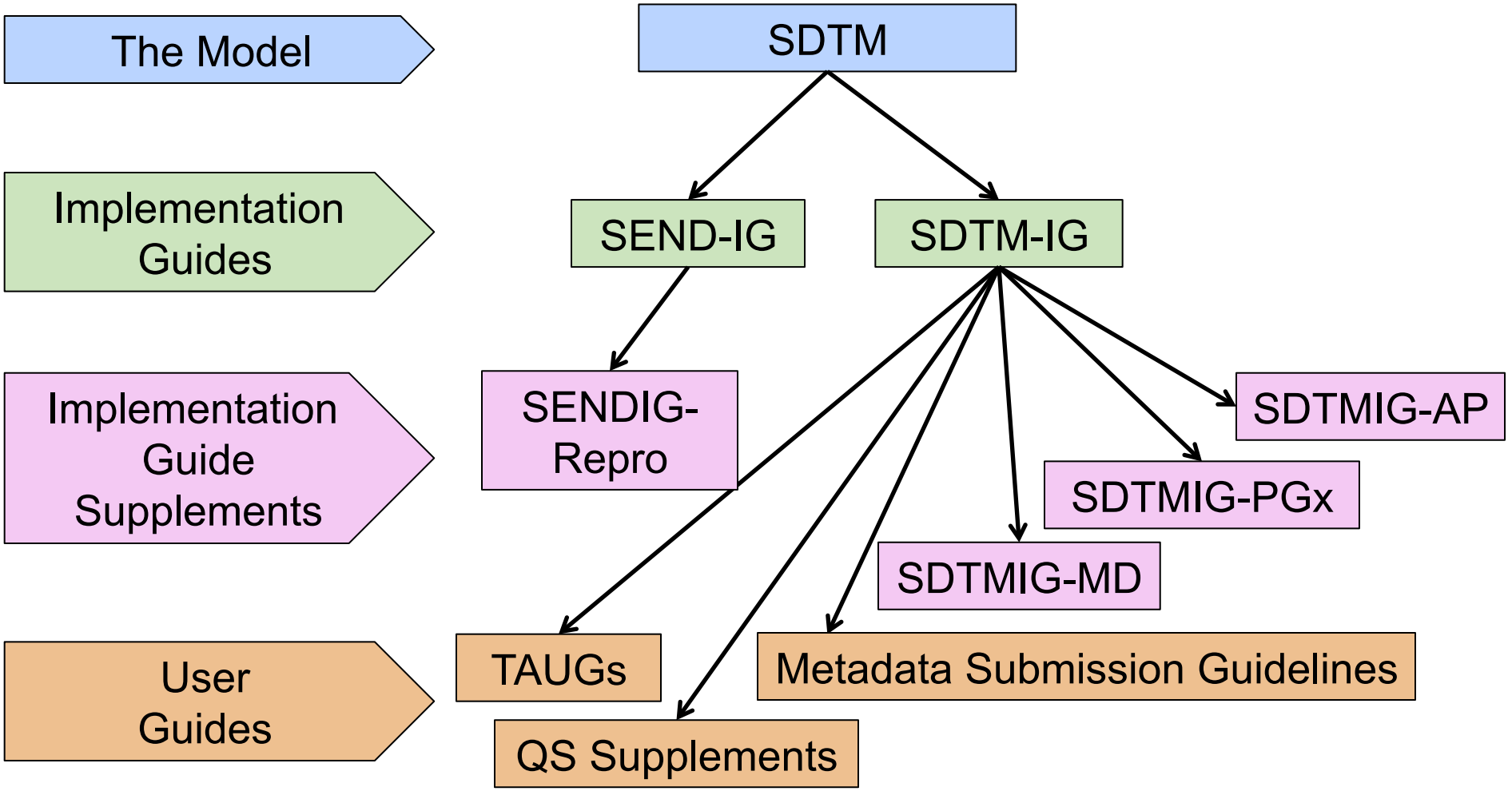
The OpenCDISC Validator and the study validation rules are available for download as standard configuration files.
- SDTM 3.1.3 (v1.1) Organization: CBER, CDER
- SDTM 3.1.2 (v.1.5) Organization: CBER, CDER
- SDTM 3.1.1 (v1.5) Organization: CBER, CDER
- Define.xml 1.0 (v1.4) Organization: CBER, CDER
- ADaM 1.0 (v.1.0)

Additional Center-specific information

- CBER Study Data Standards For additional information/support, please contact : cber.cdisc@fda.hhs.gov

http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm

Transparency at the Data Level – Quick Recap of Current Data Standards



Draft Technical Conformance Guide – Section 4.1.2

- Variable “Core” Designations
  - Required: Column in dataset, no NULL values
  - Expected: Column in dataset, some rows may have NULL values
  - Permissible: Included in dataset ***if collected***
- Some sponsors still think permissible variables are “optional” when it comes to being submitted, even if collected; If a permissible variable is collected or “derived”, it must be submitted.
- Guide shows Epoch as “…should be included for every clinical subject-level observation”

Data Transparency – Real data or a prompt question?

- As a rule, prompt questions are not submitted in SDTM
  - Questions such as “Does the subject have any relevant medical history to report?”
  - Largely just for data cleaning/monitoring purposes
  - The “No” response will not be part of any descriptive statistics or analysis
  - The “Yes” response is confirmed by the presence of a record
  - Important to understand the distinction between a simple “prompt” and the --OCCUR variable

Medical History – Prompt or OCCUR? (1)

GENERAL MEDICAL HISTORY

Does the subject have any significant medical history within the past 6 months?

☐ **Yes, list the condition(s) below** ☐ **No**

Body System			**Condition**	**End Date** (mm/dd/yyyy)	**Check if Ongoing**
1. Eyes, Ears, Nose, Throat	☐ Yes	☐ No	_____	___/___/_____	☐
2. Respiratory	☐ Yes	☐ No	_____	___/___/_____	☐
3. Gastrointestinal	☐ Yes	☐ No	_____	___/___/_____	☐
4. Endocrine/Metabolic	☐ Yes	☐ No	_____	___/___/_____	☐

Does the “Body System” represent a “Pre-Specified” term? Do we have an --OCCUR variable at all in the SDTM MH domain? The next slide shows the sponsor’s original MH dataset.

Medical History – Prompt or OCCUR? (2)

DOMAIN	MHSEQ	MHTERM	MHCAT	MHSCAT	MHOCCUR	MHENRTPT	MHENTPT
MH	1	Allergies	GENERAL MEDICAL HISTORY		N	U	2012-11-19
MH	2	Dermatological Disease	GENERAL MEDICAL HISTORY		N	U	2012-11-19
MH	3	Endocrine/Metabolic Disease	GENERAL MEDICAL HISTORY		N	U	2012-11-19
MH	4	Neurological Disease	GENERAL MEDICAL HISTORY		N	U	2012-11-19
MH	5	APPENDECTOMY - 1965	GENERAL MEDICAL HISTORY	SURGERY	Y	U	2012-11-19
MH	6	BACK PAIN	GENERAL MEDICAL HISTORY	MUSCULOSKELETAL DISEASE	Y	ONGOING	2012-11-19
MH	7	BLADDER CA	GENERAL MEDICAL HISTORY	GENITO-URINARY DISEASE	Y	ONGOING	2012-11-19
MH	8	BILATERAL CATERACTS	GENERAL MEDICAL HISTORY	HEENT	Y	ONGOING	2012-11-19
MH	9	CONSTIPATION	GENERAL MEDICAL HISTORY	GASTROINTESTINAL DISEASE	Y	ONGOING	2012-11-19

- If a body system didn't have a history, the body system was mapped to MHTERM and MHOCCUR was set to "N'. Is this according to SDTM?
- Also of note, the relative timing variable MHENRTPT is set to "U" (for "Unknown") for those records where MHOCCUR = "N".
- Of course, the question is, Should these be MH records in the first place? Would this dataset fail a validation check?
- The corrected MH dataset is shown on the next slide.

Medical History – Prompt or OCCUR? (3)

DOMAIN	MHSEQ	MHTERM	MHCAT	MHSCAT	MHENRTPT	MHENTPT
MH	1	APPENDECTOMY – 1965	GENERAL MEDICAL HISTORY	SURGERY		
MH	2	BACK PAIN	GENERAL MEDICAL HISTORY	MUSCULOSKELETAL DISEASE	ONGOING	2012-11-19
MH	3	BLADDER CA	GENERAL MEDICAL HISTORY	GENITO-URINARY DISEASE	ONGOING	2012-11-19
MH	4	BILATERAL CATARACTS	GENERAL MEDICAL HISTORY	HEENT	ONGOING	2012-11-19
MH	4	CONSTIPATION	GENERAL MEDICAL HISTORY	GASTROINTESTINAL DISEASE	ONGOING	2012-11-19

- Those records where MHOCCUR = ‘N’ are omitted from the dataset and the variable itself is deleted.
- All of the check boxes on the form are “prompts” and not migrated to SDTM; Body Systems are not Pre-Specified terms.

SDTM variables being used or mapped incorrectly (1)

- Consider the following MH dataset that includes both “pre-specified” MHTERMs as well as spontaneously reported terms
- For the pre-specified terms, the date in the MHSTDTC variable is the date that the Parkinson’s Disease specific history (or symptom) was first reported
- Notice the difference in the dates for the MHDTC variable between the pre-specified terms and the spontaneously reported terms. Is the variable being used correctly or has it been “hijacked” for another purpose?
  - ❑ *To avoid confusing or mis-leading a reviewer, care should be taken never to change the meaning of a variable*

SDTM variables being used or mapped incorrectly (2)

MHSEQ	MHTERM	MHCAT	MHPRESP	MHOCCUR	MHDTC	MHSTDTC	MHENRTPT	MHENTPT
1	ATYPICAL	PARK	Y	N	2006-05-01			
2	BALANCE	PARK	Y	Y	2006-05-01	2006-09-21		
3	BRADYKINESIA	PARK	Y	Y	2006-05-01	2006-09-21		
4	DYSKINESIA	PARK	Y	Y	2006-05-01	2007-02-12		
5	ABNORMAL GAIT	PARK	Y	Y	2006-05-01	2007-02-12		
7	L-DOPA TREATMENT	PARK	Y	Y	2006-05-01	2006-12-15		
8	RIGIDITY	PARK	Y	Y	2006-05-01	2007-08-21		
9	TREMOR	PARK	Y	Y	2006-05-01	2006-04-01		
10	BPH	GEN			2010-09-30		ONGOING	SCREENING
11	DEPRESSION	GEN			2010-09-30		ONGOING	SCREENING
12	GLAUCOMA	GEN			2010-09-30		ONGOING	SCREENING

A better solution would have been to create a separate MH record to represent the date of original diagnosis of Parkinson's disease, rather than having the date appear on every record as above.

MHSEQ	MHTERM	MHCAT	MHPRESP	MHOCCUR	MHDTC	MHSTDTC	MHENRTPT	MHENTPT
1	PARKINSON'S DISEASE	PARK	Y	Y	2010-09-30	2006-05-01		

Transparency in the EX (Exposure) Dataset

- As a required domain for any study where subjects receive investigational product, it is imperative that sponsors tell the complete “story” as to how subjects took drug
- Despite its importance in understanding a drug’s safety, our experience indicates that exposure data is one of the most vulnerable to poor and incomplete data collection and representation.
- Often, the CRF doesn’t capture enough information or the right information in order to fully represent a subject’s dosing
- Also, sponsors often fail to properly utilize the data that is collected on the CRF
- On the next slide is shown an EX dataset for a study where subjects took drug daily for 2 weeks with 3 of the doses being given in the clinic.

Representing the Complete Dosing Period

Daily Dosing with Only On-Site Doses Collected

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSFRQ	EXSTDTC	EXENDTC
ABC0001	EX	0001-101	1	DRUG A	AT SITE	150	QD	2012-01-08	2012-01-08
ABC0001	EX	0001-101	2	DRUG A	AT SITE	150	QD	2012-01-15	2012-01-15
ABC0001	EX	0001-101	3	DRUG A	AT SITE	150	QD	2012-01-22	2012-01-22
ABC0001	EX	0001-102	1	DRUG A	AT SITE	150	QD	2012-01-08	2012-01-08
ABC0001	EX	0001-102	2	DRUG A	AT SITE	150	QD	2012-01-15	2012-01-15
ABC0001	EX	0001-102	3	DRUG A	AT SITE	150	QD	2012-01-22	2012-01-22

- Is this data sufficient to give the reviewer an adequate view of the subject's exposure to study treatment? What could be added?

Create "Blanket" Dosing Records for Entire Dosing Period

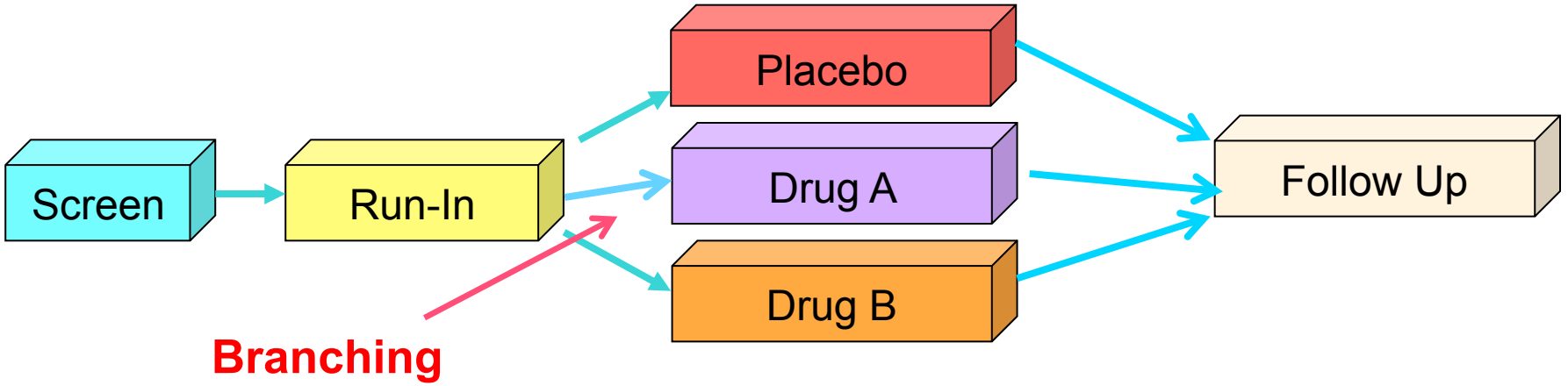
STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSFRQ	EXSTDTC	EXENDTC
ABC0001	EX	0001-101	4	DRUG A		150	QD	2012-01-08	2012-01-22
ABC0001	EX	0001-102	4	DRUG A		150	QD	2012-01-08	2012-01-22

Possibly Use EXCAT = DOSING PERIOD

Data Transparency - Incorrect or incomplete Trial Design datasets

- As the machine-readable representation of the design of a study, correct Trial Design datasets play an important role in how “data transparency” is ultimately measured.
- Trial Design datasets can usually be created straight from the protocol and can be used in CRF design to ensure the data points behind the element start rules are collected.
- The TA table defines the study Epochs from which Epoch is derived onto all subject-level observations (as requested by FDA in the Draft Technical Conformance Guide)
- Trial Design datasets allow a reviewer to understand how subjects transition through the various periods or phases of a study

Trial Design – Trial Arms



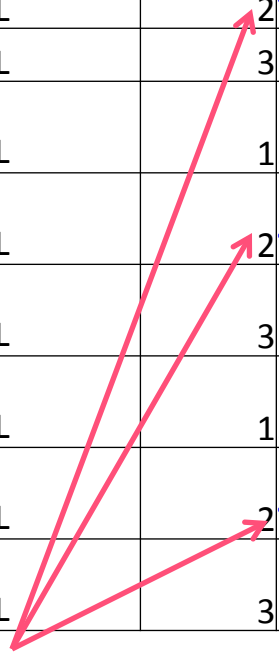
- For each ARM, TA contains one record for each occurrence of an element within the ARM
- TABRANCH highlights “decision points” at the end of elements from which subjects “branch” into an element unique to their assigned ARM
- An Epoch is defined in TA as a vertical slice of time that is independent of Arm; identifies a way to tell what is happening across elements while a trial is blinded

Trial Design – Trial Elements

- An element may appear multiple times in Trial Arms (TA) but appears only once in TE
- “Rules” describe how a subject transitions into and out of the element
- There can be no “gaps” in trial elements
- One element always leads right into the next with no gap in between. The start rule of an element defines the end of the previous element
- If trial is blinded, the start rule for a treatment element needs to differentiate one blinded treatment from another
    - ❑ *A subject is always in a trial element throughout their study participation*

Data Transparency – TA Example (1)

DOMAIN	ARMCD	ARM	TAETORD	ETCD	ELEMENT	TABRANCH	EPOCH
TA	A	DOCETAXEL	1	SCRN	Screening	Randomized to DOCETAXEL	SCREEN
TA	A	DOCETAXEL	2	TRT	Treatment		TREATMENT
TA	A	DOCETAXEL	3	FU	Follow-Up		FOLLOW-UP
TA	B	DRUG A + DOCETAXEL	1	SCRN	Screening	Randomized to Drug A + DOCETAXEL	SCREEN
TA	B	DRUG A + DOCETAXEL	2	TRT	Treatment		TREATMENT
TA	B	DRUG A + DOCETAXEL	3	FU	Follow-Up		FOLLOW-UP
TA	C	DRUG B + DOCETAXEL	1	SCRN	Screening	Randomized to Drug B + DOCETAXEL	SCREEN
TA	C	DRUG B + DOCETAXEL	2	TRT	Treatment		TREATMENT
TA	C	DRUG B + DOCETAXEL	3	FU	Follow-Up		FOLLOW-UP



Within each Arm, is there an element that makes that Arm unique? Having only a single treatment element doesn't differentiate one Arm from another.

Data Transparency – TA Example (2)

DOMAIN	ARMCD	ARM	TAETORD	ETCD	ELEMENT	TABRANCH	EPOCH
TA	A	DOCETAXEL	1	SCRN	Screening	Randomized to DOCETAXEL	SCREEN
TA	A	DOCETAXEL	2	**DOCET**	Docetaxel		TREATMENT
TA	A	DOCETAXEL	3	FU	Follow-Up		FOLLOW-UP
TA	B	DRUG A + DOCETAXEL	1	SCRN	Screening	Randomized to Drug A plus DOCETAXEL	SCREEN
TA	B	DRUG A + DOCETAXEL	2	**DRGADOC**	Drug A plus Docetaxel		TREATMENT
TA	B	DRUG A + DOCETAXEL	3	FU	Follow-Up		FOLLOW-UP
TA	C	DRUG B + DOCETAXEL	1	SCRN	Screening	Randomized to Drug B plus DOCETAXEL	SCREEN
TA	C	DRUG B + DOCETAXEL	2	**DRGBDOC**	Drug B plus Docetaxel		TREATMENT
TA	C	DRUG B + DOCETAXEL	3	FU	Follow-Up		FOLLOW-UP

Within each Arm, now we have a treatment element that makes each Arm unique. Again, having only a single treatment element doesn't differentiate one Arm from another.

Transparency – Study Level Documentation (1)

- Overall theme should be to develop the study metadata and data guide as early as possible in the process
  - These are the best tools to get the reviewer up to speed as quickly as possible; It's more than just simply fulfilling the regulatory requirement.
  - Remember that just because you can document the mapping in the Define doesn't mean you can use a variable for other than its intended purpose
- With all pieces, goal should be to provide as much detail as possible regarding the collection and reporting of each piece of data.
- At all points, reference the metadata submission guidelines (scheduled to be updated in the near future).

Transparency – Study Level Documentation (2)

- Define.xml
  - Begin assembling early in the study (codelists, value-level metadata)
  - Avoid using pre-conversion or “source” database variable names; FDA has no access to operational database names
- Annotated CRF (BlankCRF)
  - As with the Define, avoid using source variable names
  - Clearly differentiate data points where there is “no data collected” (variable included if data exists) versus those that are “not submitted” (variable not included)
- Data Guide
  - Use the PhUSE developed template to fully explain all aspects of the data; Should be developed throughout the course of the study
  - Essential in explaining any oddities in the data as well as documenting validation errors or warnings

Conclusions

- The level of confidence in a trial's data transparency and/or traceability can be affected by poor or misleading mapping from operational source to SDTM

- Well designed metadata (Define.xml and aCRF) as well as the Data Guide further help to ensure data transparency; these are necessary supplements to the study data

- SDTM, as the submission format for the tabulation data, cannot make up for inadequate data management practices or poor query resolution during study execution

Contact Information

Fred Wood, *Senior Manager and Lead, Data Standards Consulting*

f.wood@accenture.com

Jerry Salyers, *Senior Consultant, Data Standards Consulting*

jerry.j.salyers@accenture.com

Richard Lewis, *Senior Consultant, Data Standards Consulting*

richard.m.lewis@accenture.com