Assessing Data Quality with Focus on Content

Michael Beers, Pinnacle 21

PhUSE SDE, Ridgefield, CT
July 26, 2018

PINNACLE 21

P21

FDA Definition for Data Quality

- *"Quality data would therefore be defined as data that sufficiently support conclusions and interpretations equivalent to those derived from error-free data"*

  Janet Woodcock, M.D., 1999

- *"Study data validation helps to ensure that the study data are compliant, useful, and will support meaningful review and analysis"*

  FDA TCG, 2018-03

P21

Increasing Demand for Data Quality

- FDA goal is to make drugs available ASAP
- Increased automation is needed for this goal
- Manual review process is inefficient
  - Data review time is limited
  - Data quality evaluation is difficult
- Automation relies on high quality data
  - Value of standard reporting and analysis for Reviewers increases demand for data quality
- Standardized data allows early evaluation and enforcement of data quality

Standardized Collected Data

- Standardization should be done at collection
- Standardized digital data, with no or limited manual conversion, should be the goal
- Any touch of data could change original meaning of collected information
- Relying on conversions by many different preparers leads to different interpretations and implementations
- CDASH provides standardization for some types of data, but gaps need to be filled
- Sponsors need to use these collection standards

CDISC Standards (and extensions)

- CDASH
- SDTM
- ADaM
- SEND
- Define.xml
- TAUGs
- Provisional IGs

P21

CDISC Standards and their adoption

- 1997 – CDISC started as a Volunteer group
- 2005 – SDTM IG 3.1.1, Define-XML 1.0
- 2008 – OpenCDISC
- 2009 – SDTM IG 3.1.2, ADaM IG 1.0
- 2011 – SEND IG 3.0
- 2012 – SDTM IG 3.1.3
- 2013 – SDTM IG 3.2, Define-XML 2.0
- 2016 – Required by FDA for new studies
- 2016 (2020) – Required by PMDA

P21

SDTMIG versions used – newest version used for a study in a submission

- SDTMIG v.3.2
  - ~30%
- SDTMIG v.3.1.3
  - ~45%
- SDTMIG v.3.1.2
  - ~20%
- SDTMIG v.3.1.1
  - A few

Older versions of standards (< SDTMIG v.3.1.3) cause problems

- Missing important information:
  - DM domain
    - RFXSTDTC/RFXENDTC, ACTARMCD/ACTARM, DTHFL/DTHDTC, RFICDTC
  - AE and MH domains
    - Coding information as suppquals (or missing)

New versions of standards may introduce problems

- Inconsistency with how to map ARM values for screen failures

Example of older SDTMIG version causing problems

- Missing important information for <v3.1.3 studies:
  - DM domain
    - RFXSTDTC/RFXENDTC, ACTARMCD/ACTARM, DTHFL/DTHDTC, RFICDTC
- Missing DTHFL/DTHDTC causes this Death Reconciliation report to fail and may result in considerable effort to manually investigate

Death Reconciliation

Search

Print | Copy | Download

USUBJID	ARMCD	RFSTDTC	DTHFL	DTHDTC	AETERM	AEOUT	AESDTH	AESTDTC	AEENDTC	DSTERM	DSDECOD	DSCAT	DSSTDTC
DUMMY-001	SCRNFAIL		Y	2013-06-21	OVERDOSE	FATAL	Y	2013-06-19	2013-06-19	DEATH	SCREEN FAILURE	DISPOSITION EVENT	2013-06-21
DUMMY-002	B	2014-07-08T08:00	Y	2014-07-17	INTOXICATION	FATAL	Y	2014-07-17	2014-07-17				

Cross-study data quality assessment

- As implementation of standards improves, we are able to expand data quality assessment to cross-study and cross-product review
- Consistency across studies requires adherence to both CDISC standards as well as sponsor standards
- Inconsistency across studies in a submission leads to manual effort in review
- Consistency across studies is important for ISS preparation
- PMDA has begun utilizing their accumulated data for cross-product review

Cross-study data quality assessment – example 1

- Inconsistent coding of AETERM across (and within) studies
  - Requires all levels of MedDRA coding to be provided
  - Inconsistency in coding increases the difficulty in analyzing adverse event data

aeterm	aedecod	aehlt	aehlgt	aebodsys	study	count_of_aes
ANGULAR CHEILITIS	ANGULAR CHEILITIS	DENTAL AND ORAL SOFT TISSUE INFECTIONS	INFECTIONS - PATHOGEN UNSPECIFIED	INFECTIONS AND INFESTATIONS	Study 1	3
ANGULAR CHEILITIS	CHEILITIS	ORAL SOFT TISSUE DISORDERS NEC	ORAL SOFT TISSUE CONDITIONS	GASTROINTESTINAL DISORDERS	Study 2	5
ANGULAR CHEILITIS	CHEILITIS	ORAL SOFT TISSUE DISORDERS NEC	ORAL SOFT TISSUE CONDITIONS	GASTROINTESTINAL DISORDERS	Study 3	5
COMMON COLD	NASOPHARYNGITIS	UPPER RESPIRATORY TRACT INFECTIONS	INFECTIONS - PATHOGEN UNSPECIFIED	INFECTIONS AND INFESTATIONS	Study 2	38
COMMON COLD	NASOPHARYNGITIS	UPPER RESPIRATORY TRACT INFECTIONS	INFECTIONS - PATHOGEN UNSPECIFIED	INFECTIONS AND INFESTATIONS	Study 3	9
COMMON COLD	VIRAL UPPER RESPIRATORY TRACT INFECTION	VIRAL INFECTIONS NEC	VIRAL INFECTIOUS DISORDERS	INFECTIONS AND INFESTATIONS	Study 1	66

Cross-study data quality assessment – example 2

- Suppqual consistency across studies
  - Sponsor standards are important too
  - Inconsistent naming of non-standard variables adds additional time to review

qnam	qlabel	study	dataset
LANG	LANGUAGE ADMINISTERED	Study 1	SUPPQS
LANG	LANGUAGE ADMINISTERED	Study 3	SUPPQS
LANG_	LANGUAGE ADMINISTERED	Study 4	SUPPQS
LESNSITE	LESION SITE	Study 1	SUPPTU
LESNSITE	LESION SITE	Study 2	SUPPTU
LESNSITE	LESION SITE	Study 3	SUPPTU
LESNSIT_	LESION SITE	Study 4	SUPPTU
METHOD	METHOD OF ADMINISTRATION	Study 1	SUPPQS
METHOD	METHOD OF ADMINISTRATION	Study 3	SUPPQS
METHOD_	METHOD OF ADMINISTRATION	Study 4	SUPPQS
DSRAND	RANDOMIZATION NUMBER	Study 4	**SUPPDS**
RANDNUM	RANDOMIZATION NUMBER	Study 1	SUPPDM
RANDNUM	RANDOMIZATION NUMBER	Study 3	SUPPDM

Cross-study data quality assessment – example 3

- Potential duplicate subjects
  - Having this information as early as possible is key to determining if the subjects are the same
  - Can use data such as height, and subject initials if available, to narrow down the list

studyid	usubjid	brthdtc	sex	race	country	height (cm)	armcd	actarmcd	rfxstdtc	rfxendtc
study 1	study-1-050001	5/12/1946	F	WHITE	USA	152.4	DRUG B	DRUG B	5/29/2015	7/17/2015
study 3	study-3-008020	5/12/1946	F	WHITE	USA	149.8	DRUG A	DRUG A	8/11/2016	8/1/2017
study 3	study-3-023023	6/11/1950	M	WHITE	USA	183	DRUG A	DRUG A	7/28/2016	7/21/2017
study 2	study-2-047001	6/11/1950	M	WHITE	USA	183	DRUG A	DRUG A	8/19/2016	2/16/2017
study 1	study-1-027004	3/21/1970	F	BLACK OR AFRICAN AMERICAN	USA	160	DRUG B	DRUG B	6/3/2015	8/25/2015
study 1	study-1-021016	3/21/1970	F	BLACK OR AFRICAN AMERICAN	USA	162.5	DRUG A	DRUG A	7/17/2015	10/7/2015
study 2	study-2-034004	3/21/1970	F	BLACK OR AFRICAN AMERICAN	USA	160	DRUG A	DRUG A	7/27/2016	9/13/2016

Extensions of standards - TAUGs

- Approaching level of good-enough standards (SDTM, SEND, ADaM)
- Regulatory agencies beginning to require submission in these standards
- Data standards catalogs ending support for older versions of these standards
- Next focus should be on TAUGs…

Therapeutic Area User Guides

- Released by CDISC
  - 27 - Total
  - 13 – Acceptance by FDA (Referenced in TCG)
  - All* – Acceptance by PMDA

    *Based on PMDA statement at CDISC Japan Interchange: “We basically accept that applicants implement the published TAUGs”
- Released by FDA
  - 2 - Total
    - HIV
    - Vaccines
- Implemented by industry
  - Practically nonexistent
    - 1 out of 32 products (for indications which have a TAUG available)

TAUG Implementation

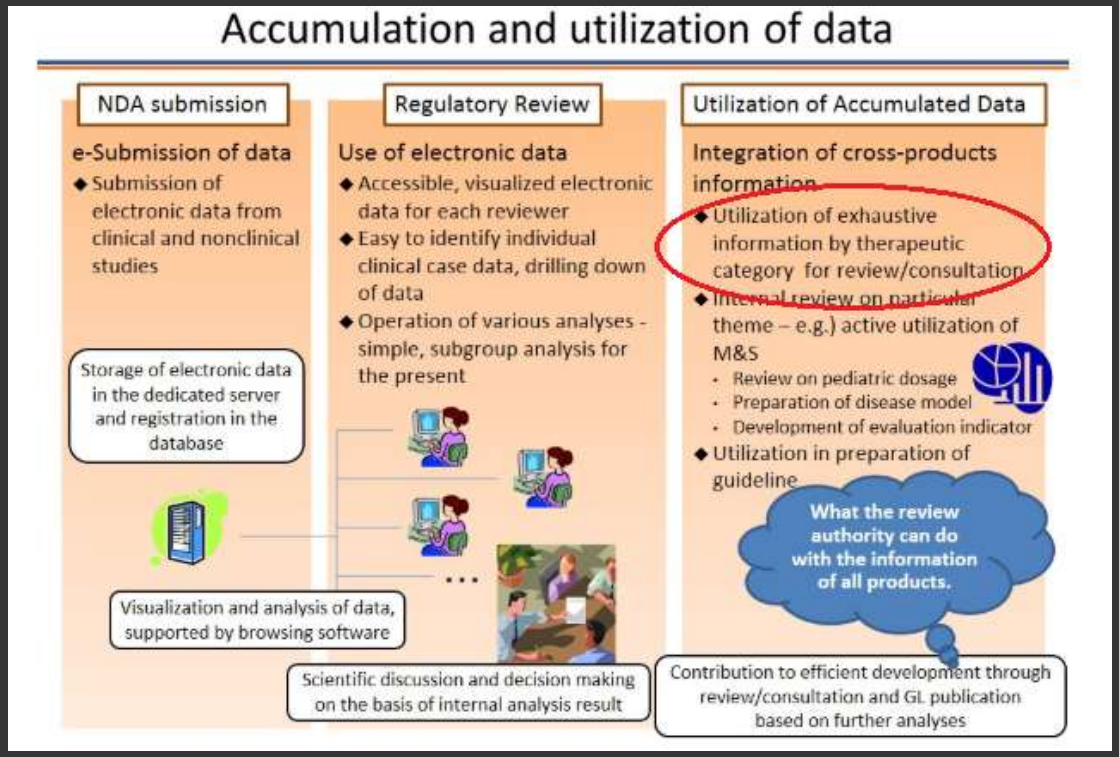
- The Technical Conformance Guide states:
  - "Sponsors may use new TA extensions of a CDISC standard, *but are not required to until the extensions have been incorporated into a SDTMIG version supported by FDA* (the supported SDTMIGs are listed in the Catalog)."
  - "Sponsors should *explain the rationale in the cSDRG for using TA extensions* that are not currently listed in the Guide."

TAUG Implementation

- Lack of TAUG implementation means that although there is guidance for how to map similar data for the same indication, the industry is not mapping it consistently.
- Data collection examples in the TAUGs are not followed

TAUG (lack of) Implementation

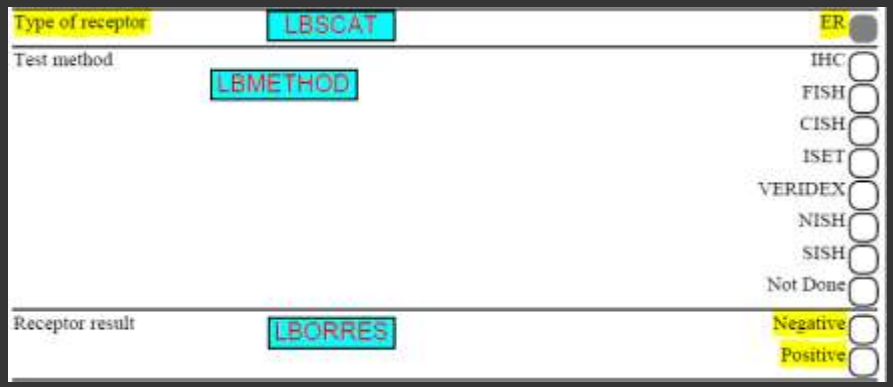
- An important impact of this is that PMDA is doing cross-product analysis from accumulated study data.



http://www.pmda.go.jp/english/review-services/reviews/advanced-efforts/0002.html

TAUG Inconsistency Example 1 – Breast Cancer

- Estrogen receptor data – same data mapped inconsistently



Study A – Mapped to the LB domain

Study B – Mapped to the PF domain

2.	Estrogen Receptor Status [Estrogen Receptor Status] PFTEST	Positive / Negative PFORRES
3.	Pathological Diagnosis Progesterone Receptor Status [Patholog Diag Progester Recep Stat] PFTEST	Positive / Negative PFORRES
4.	Pathological Diagnosis HER2 Status [Pathological Diagnosis HER2 Status] PFTEST	Positive / Negative PFORRES

Study C – Mapped to the MI domain

This matches the example in the TAUG.

TAUG Inconsistency Example 2 – Diabetes

- Hypoglycemia signs/symptoms data – same data mapped and collected inconsistently

Q1B - Result in loss of consciousness? FAOBJ = this episode resulted in loss of consciousness

FAORRES

Q1C - Result in a seizure?
If response to ALL of the above is "No", then Severe Hypoglycemia is assessed not to have occurred: jump to Question #3.
If response to ANY of the above is "Yes", complete Question #2 to collect additional data regarding the Severe Hypoglycemia episode AND Question #3 to assess whether this episode also meets criteria for Documented Hypoglycemia.

FAOBJ = this episode resulted in a seizure

FAORRES

Study A – Mapped to the FA domain

Study B – Mapped to the CF domain

4.	Symptoms of Hypoglycemia [Symptoms of Hypoglycemia] CFORRES when CFTESTCD='HYPOGSYM'	[Symptoms of Hypoglycemia] [A:NO SYMPTOMS] No Symptoms [A:MARKEDLY DEPRESSED LEVEL OF CONSCIOUSNESS] Markedly Depressed Level of Consciousness [A:LOSS OF CONSCIOUSNESS] Loss of Consciousness [A:SEIZURE] Seizure [A:OTHER SYMPTOMS NOT LISTED ABOVE] Other Symptoms not Listed Above

CECAT= HYPO SYMPTOMS	Were Signs/Symptoms Present?	No Yes (If yes complete following) CEYN	
CETERM= SWEATING	Sweating	No Yes	CEOCCUR with CEPRESP=Y
CETERM= TREMORS/TREMBLING	Tremors/Trembling	No Yes	
CETERM= DIZZINESS	Dizziness	No Yes	
CETERM= COGNITIVE IMPAIRMENT	Cognitive Impairment	No Yes	
CETERM= LOSS OF CONSCIOUSNESS	Loss of Consciousness	No Yes	
CETERM= CONVULSIONS/SEIZURE	Convulsions/Seizure	No Yes	
CETERM= COMA	Coma	No Yes	
	Other (Specify)	No Yes (if yes enter below)	

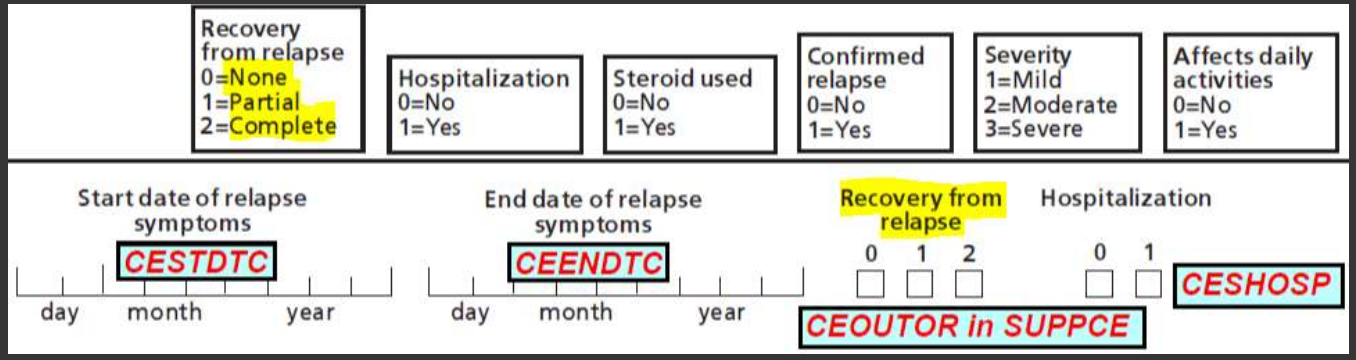
For reference, this is how it is in the TAUG. Mapped to the CE domain and collected differently.

TAUG Inconsistency Example 3 – Multiple Sclerosis

- Relapse data – same data mapped inconsistently, but no guidance available

Study A – 'Recovery from relapse' mapped to a suppqual

Study B – same exact 'Recovery from relapse' data, mapped to a different suppqual



If it is data common to a therapeutic area, shouldn't the TAUG clearly define a mapping?

Conclusion

- Standardization needs to be pushed even farther, as early as possible, and filling in the gaps
- Automated data quality assessments need to be created to ensure high quality standardized data
- Therapeutic Area User Guides
  - SDOs need to create stronger guidance for therapeutic areas, from collection to analysis
  - Regulatory agencies should enforce the use of these extensions of the standards
  - Industry needs to adopt and implement the extensions of the standards
  - Validation rules and data quality assessments need to be created for the compliance to the extensions of the standards

Questions

Michael Beers
mbeers@pinnacle21.net