

JumpStart

Common Data Quality Issues in SDTM Data

PhUSE US Connect
June 14-18, 2021

Presented By:

Alan Shapiro MD PhD

Food and Drug Administration (FDA)

Michael Xue

International Business Machines (IBM)

CDER Office of Computational Science

Disclaimer Statement

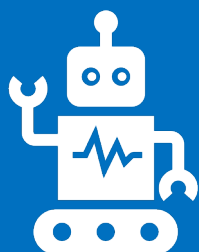


Disclaimer

The views and opinions presented here represent those of the speakers and should not be considered to represent advice or guidance on behalf of the U.S. Food and Drug Administration.

SDTM Assessment Overview

- ❖ The JumpStart service provides CDER review teams an assessment of the composition, conformance, and quality of the clinical trial data submitted within an application.



JumpStart uses automated tools (i.e. DataFit) to assess the data quality and composition of SDTM data submission



Analysts investigate data quality issues and assess the impact these issues have on the ability to conduct a thorough and complete review



Standard reports for each study examine supplemental contents, deaths, and coding quality within the AE and DS domains

SDTM Assessment Results

- ❖ Over time, as JumpStart provided support to more and more review teams, it was able to identify frequently occurring data quality issues.
 - Some issues warrant further scrutiny and/or may hinder the reviewer process.
 - Ambiguity in the data, coding discrepancies, or faulty metadata may require clarification from applicants, resulting in delays.
- ❖ Common findings discussed in this presentation are the following:
 - Reference Exposure Start Dates Provided for Subjects Not in Safety Population
 - Reference Participation End Date Not Implemented Correctly
 - Missing and Inconsistent Serious Criteria Qualifiers
 - Missing Adverse Event Severity and Toxicity Information
 - Discrepancies Between the Sponsor's Treatment Emergent Flag Derivation and Records Flagged as Treatment Emergent
 - Disposition Terms Mapped to Inappropriate Disposition Codes
 - Exposure Records After Latest Disposition Dates
 - Subjects Missing Disposition Start Dates
 - Missing Baseline Liver Records
 - Missing Liver Test Records

Reference Exposure Start Dates Provided for Subjects Not in Safety Population

Findings	- There are subjects with Reference Exposure Start Dates (RFXSTDTC) but were excluded from the safety population (SAFFL = 'N') - Study designs may have a run-in period prior to a subject receiving the study drug
Impact	Additional investigation may be required to understand which subjects actually received the study drug. Using the safety population flag by itself, may exclude other potential subjects who were actually treated.

❖ Below are examples of subjects who are not in the safety population but have a reference exposure start date:

DM				EX					ADSL
USUBJID	ACTARM	RFXSTDTC	RFXENDTC	EXSEQ	EXTRT	EXDOSE	EXSCAT	EXSTDTC	SAFFL
ABC-123	Not Treated	2019-09-09	2019-09-09	1	PLACEBO	0	RUN-IN PERIOD	2019-09-09	N

Reference Participation End Dates Not Implemented Correctly

Findings	▶ Date of End of Participation (RFPENDTC) in the Demographics (DM) domain was not implemented according to SDTM guidance
Impact	▶ Using this variable in analysis without considering all SDTM datasets may lead to unexpected results

- ❖ Below is the definition for RFPENDTC from the SDTM Implementation Guide:
 - **Date of End of Participation (RFPENDTC):** *Date when subject ended participation or follow-up in a trial [...]. Should correspond to the last known date of contact.*
- ❖ Below is an example where RFPENDTC is not the last known date of contact:

DM					DS			
USUBJID	ACTARM	RFXSTDTC	RFXENDTC	RFPENDTC	DSSEQ	DSCAT	DSDECOD	DSSTDTC
ABC-123	XYZ-12	2020-01-12	2020-01-31	2020-02-09	6	DISPOSITION EVENT	ADVERSE EVENT	2020-02-14

In this example this subject has a record in the Disposition (DS) domain that is later than the RFPENDTC. Hence, RFPENDTC is incorrect and does not reflect the latest date of contact with the subject.

Missing Adverse Event Severity or Toxicity Information

Findings	▶ Adverse Event Severity/Toxicity (AESEV/AETOXGR) is missing for adverse event records in the Adverse Events (AE) domain
Impact	▶ Missing values lack analytical information; the affected records in the AE domain will not be included in severity or toxicity analyses

❖ Below are example subjects missing Adverse Event Severity information :

AE							ADAE
USUBJID	AESQ	AEDECOD	AESEV	AEOUT	AESTDTC	AEENDTC	TRTEMFL
ABC-123	12	Anemia		ONGOING	2020-06-10		Y
DEF-345	8	Swelling face		RECOVERED/ RESOLVED	2020-05-06	2020-11-18	Y

Missing and Inconsistent Seriousness Qualifiers

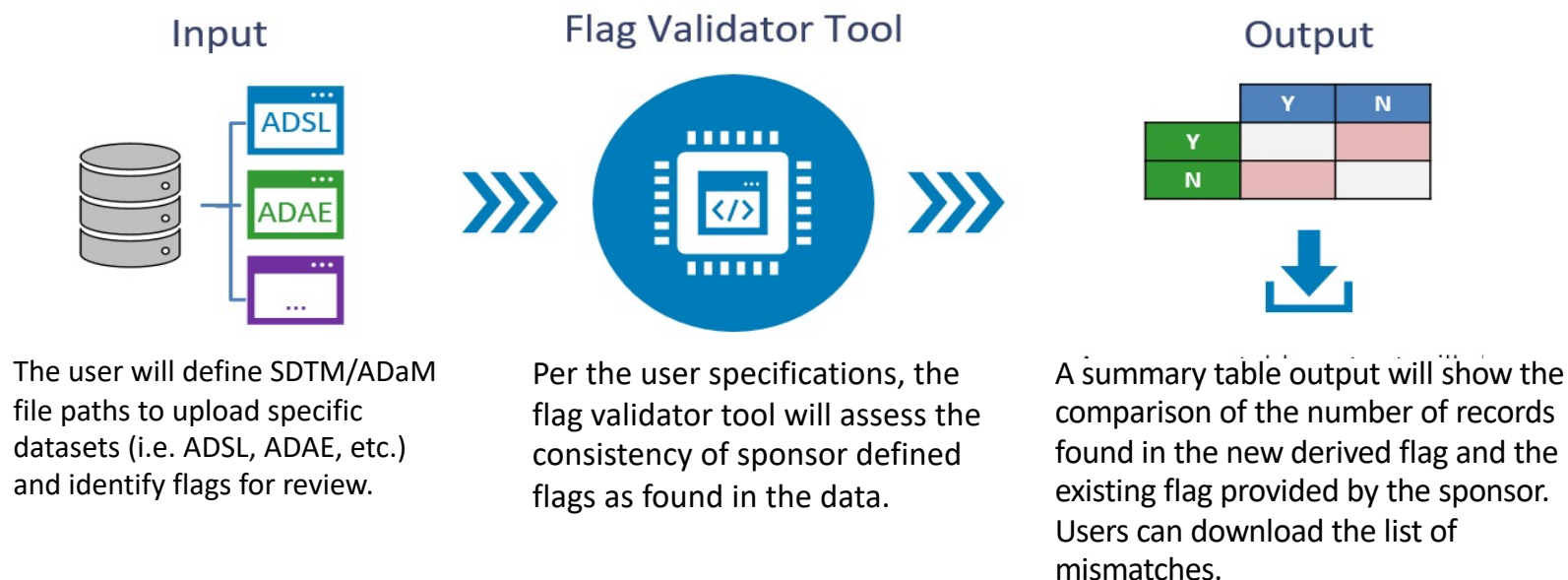
Findings	- Adverse Events have no serious criteria qualifiers set to 'Y' when an adverse event is serious (AESER='Y') - Adverse event records have serious criteria qualifiers marked as 'Y' but the record is not marked as Serious (AESER = 'N')
Impact	It may be difficult to summarize reasons for Serious Adverse Events and independently verify whether an adverse event was serious or not

❖ Below are example records of the proper and improper use of serious criteria qualifiers:

AE									
USUBJID	AESQ	AEDECOD	AESER	AESCONG	AESDISAB	AESDTH	AESHOSP	AESLIFE	AESMIE
ABC-123	5	Headache	N						
DEF-345	3	Stroke	Y				Y	Y	
GHI-678	7	Heart Attack	Y						
JKL-910	8	Abdominal Pain	N		Y		Y		

Flag Validator Tool

- ❖ Variable accuracy and consistency has long complicated the application review process. The Flag Validator Tool will validate the **accuracy** of flags submitted by the sponsor so that the review team can use the flags with **confidence** in analysis.



Treatment Emergent Flag Discrepancies

Findings	- There adverse event records where the sponsor's Treatment Emergent Flag (TRTEMFL) in the ADAE dataset equals 'Y' but and the Flag Validator Tool (FVT) derived flag equals N - Discrepancy is due to adverse events that occur on the same day but before the treatment start and still have TRTEMFL = 'Y' even though the Define.xml derivation states - Sample treatment emergent flag (TRTEMFL) Define.xml derivation: <i>TRTEMFL = 'Y', when the adverse event start date/time (ASTDTM) occurs on or after treatment start date/time - (ASTDTM >= TRTSDTM)"</i>
Impact	Using the TRTEMFL in the ADAE will include non-treatment emergent adverse events that occurred prior to the treatment start date/time in AE analyses. Additional time will be needed to review these records

❖ Below are adverse events with TRTEMFL = 'Y' with adverse events occurring before treatment start:

ADSL					ADAE					
USUBJID	ACTARM	TRTSDT	TRTSDTM	SAFFL	AESEQ	AEDECOD	ASTDT	ASTDTM	AENDT	TRTEMFL
ABC-123	XYZ-12	04Mar2020	04Mar2020 5:00:00 PM	Y	3	Nasopharyngitis	04Mar2020	04Mar2020 11:15:00 AM		Y
DEF-456	Placebo	03Jan2021	03Jan2021 11:54:00 AM	Y	5	Vomiting	03Jan2021	03Jan2021 8:45:00 AM	19Jan2019	Y

Inappropriate Disposition Code Mapping

Findings	- There are instances where values for Reported Term for the Disposition Event (DSTERM) are mapped to seemingly inappropriate Standardized Disposition Term (DSDECOD) values - There standard CDISC DSDECOD terms that a sponsor may be used to code certain DSTERMs such as: - ADVERSE EVENT, COMPLETED, DEATH, WITHDRAWAL BY SUBJECT, LACK OF EFFICACY, LOST TO FOLLOWUP, PROTOCOL DEVIATION, PHYSICIAN DECISION
Impact	More appropriate coded values may exist for these dispositions. Additional time should be set aside to review these records

❖ Below is an example of Reported Term for the Disposition Event (DSTERM) values that may have more appropriate Standardized Disposition Event (DSDECOD) terminology:

DS				
USUBJID	DSSEQ	DSTERM	DSDECOD	DSSTDTC
ABC-123	1	LACK OF EFFICACY	OTHER	2020-08-09
DEF-456	3	PATIENT WANTED TO WITHDRAWAL FROM STUDY	OTHER	2020-05-23

CDISC-Controlled Terminology such as 'LACK OF EFFICACY' or 'WITHDRAWAL BY SUBJECT' may be more appropriate for these records instead of 'OTHER'

Missing Disposition Start Date and Study Day Information

Findings	There are records missing Start Date/Time of Disposition Event (DSSTDTC) and Study Day of Start of Disposition Event (DSSTDY) in the Disposition (DS) domain
Impact	Without a time point, it may not be possible to determine when these disposition events occurred. Missing timing information could also affect time to disposition event analysis

❖ Below are example records of subjects missing DSSTDTC and DSSTDY:

DS					
USUBJID	DSSEQ	DSCAT	DSDECOD	DSSTDTC	DSSTDY
ABC-123	1	PROTOCOL MILESTONE	INFORMED CONSENT OBTAINED	2019-03-04	-9
	2		RANDOMIZED	2019-03-13	1
	3	DISPOSITION EVENT	ADVERSE EVENT		
DEF-456	1	PROTOCOL MILESTONE	INFORMED CONSENT OBTAINED	2018-09-18	-37
	2		RANDOMIZED	2018-10-25	1
	3	DISPOSITION EVENT	COMPLETED		

Exposure Dates After Latest Disposition Date

Findings	Records in the Exposure (EX) domain have dates that are after the latest Disposition Start Date (DSSTDTC)
Impact	Further investigation into the Disposition and Exposure domains may be required to better understand when subjects discontinued from treatment

❖ Below is an example of a subject with exposure dates that are after the latest DSSTDTC:

DM		EX				
USUBJID	ACTARM	EXSEQ	EXTRT	EXDOSE	EXSTDTC	EXENDTC
ABC-123	XYZ-12	1	XYZ-12	50	2020-04-23 T11:32	2020-04-23 T11:32
		2		50	2020-07-22 T10:46	2020-07-22 T10:46
		3		50	2020-10-21 T09:35	2020-10-21 T09:35

DM		DS		
USUBJID	ACTARM	DSSEQ	DSDECOD	DSSTDTC
ABC-123	XYZ-12	3	ADVERSE EVENT	2020-09-19

Missing Baseline Liver Test Records

Findings	Baseline values are missing for at least one of the following tests: ALT, AST, ALP and BILI in the Laboratory (LB) domain
Impact	Without baseline test results, it will be more difficult to discern whether elevated readings are the result of exposure to treatment or typical for a particular subject

❖ Below is an example of a subject missing baseline liver test records:

DM			LB						
USUBJID	ACTARM	RFXSTDTC	LBSEQ	LBTESTCD	LBSTRESN	LBSTRESU	LBBLFL	LB DTC	VISIT
ABC-123	XYZ-12	2019-11-29 T09:30:00	1	ALT	20	U/L	Y	2019-11-28 T10:30:00	BASELINE
			2	ALT	33	U/L		2019-12-13 T10:30:00	VISIT 1
			3	AST	29	U/L		2019-12-13 T10:30:00	VISIT 1
			4	ALP	239	U/L		2019-12-13 T10:30:00	VISIT 1
			5	BILI	3	umol/L	Y	2019-11-28 T10:30:00	BASELINE
			6	BILI	5	umol/L		2019-12-13 T10:30:00	VISIT 1

This subject is Missing Baseline Records
for AST and ALP Laboratory Tests

Missing Liver Test Records

Findings	Subjects are missing test results for AST, ALT, ALP or BILI in the Laboratory (LB) domain
Impact	Missing liver lab test results means the affected subjects in the LB domain will not be included in liver lab analyses. Without liver lab test results, one would not be able to identify candidates that potentially met Hy's Law criteria

❖ Below is an example of a subject missing AST and ALP laboratory test records:

DM			LB				
USUBJID	ACTARM	RFXSTDTC	LBSEQ	LBTESTCD	LBSTRESN	LBSTRESU	LBSTC
DEF-456	XYZ-12	2020-05-29	1	ALT	33	U/L	2020-06-15
			2	BILI	5	umol/L	2020-06-15
GHI-789	XYZ-12	2020-03-28	1	ALT	23	U/L	2020-04-01
			2	AST	33	U/L	2020-04-01
			3	ALP	250	U/L	2020-04-01
			4	BILI	6	umol/L	2020-04-01

Subject DEF-456 is Missing AST and ALP Laboratory Tests

