

JumpStart



Common Data Quality Issues in SDTM Data

PhUSE US Connect
March 8-11, 2020

CDER Office of Computational Science
Better Data | Better Tools | Better Decisions

Introductions

Andrea Arbelaez

U.S. Food and Drug Administration
Silver Spring, MD, USA

Aaron Belowich

IBM
Silver Spring, MD, USA

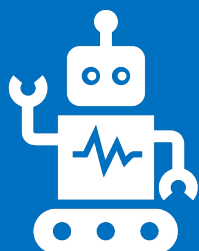
Disclaimer Statement



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



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

Table of Contents

- ❖ Implementation of Death Flag
- ❖ Deriving Treatment Start/End Dates
- ❖ Missing Screen Failures in IE
- ❖ Missing Reference Ranges in LB
- ❖ Reference Participation End Date
- ❖ Implementation of ACTARM
- ❖ Medical History after Subject Start Date
- ❖ AE Outcomes and End Dates
- ❖ Inconsistent Units in LB

Implementation of Death Flag

Findings	Subjects who appear to have died are missing Death Flag (DTHFL) in the Demographics (DM) domain
Impact	Incomplete and/or inconsistent death information may prevent a complete analyses of deceased subjects for the safety evaluation

- ❖ Example records for two subjects who appear to have died, however Subject XYZ-123 is missing death information in DM:

DOMAIN	USUBJID	RFSTDTC	RFENDTC	DTHFL	DTHDTC
DM	XYZ-123	2016-08-12	2016-11-11		
DM	ABC-456	2017-03-12	2017-07-06	Y	2017-07-06

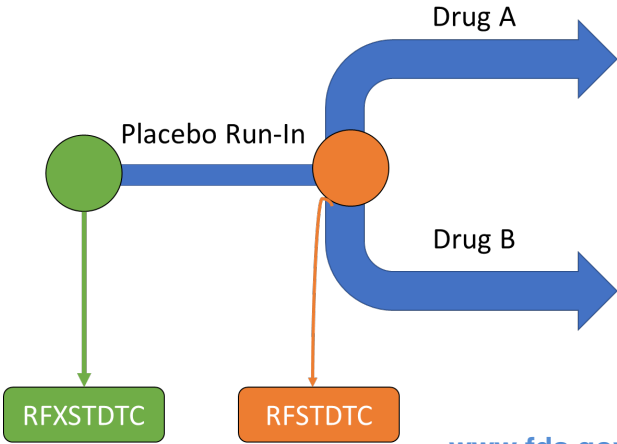
DOMAIN	USUBJID	AEDECOD	AESER	AESDTH	AESTDTC
AE	XYZ-123	Stroke	Y	Y	2016-12-01
AE	ABC-456	Heart Failure	Y	Y	2017-07-06

DOMAIN	USUBJID	DSCAT	DSDECOD	EPOCH	DSSTDTC
DS	XYZ-123	Disposition Event	Death	Follow-Up	2016-12-05
DS	ABC-456	Disposition Event	Death	Treatment	2017-07-06

Deriving Treatment Start/End Dates

Findings	- Subject Reference Start/End Date/Time (RFSTDTC/RFENDTC) represent the <i>study treatment</i> start and end dates, while Date/Time of First/Last Study Treatment (RFXSTDTC/RFXENDTC) represent start and end dates to any records in the exposure domain - RFSTDTC/RFXSTDTC and RFENDTC/RFXENDTC often are equivalent, but the difference is that RFSTDTC/RFENDTC are restricted to study treatment (i.e. records from a placebo run-in period are excluded)
Impact	Deriving treatment start and end dates incorrectly may create confusion about the duration a subject was treated or when events occur as all Study Day (--DY) variables are derived using RFSTDTC

❖ The diagram to the right depicts a study design in which subjects are entered into a placebo run-in period before being assigned to either Drug A or Drug B for the study:



Missing Screen Failures in IE

Findings	- Screen Failures are missing in the Inclusion/Exclusion (IE) domain - Screen Failures must meet at least one exclusion criteria or not meet at least one inclusion criteria as specified in the protocol
Impact	Without explanations for why a subject is classified as a Screen Failure, it will not be possible to confirm a lack of bias in patient selection

❖ Example records of Screen Failure subjects who are included in the IE domain:

USUBJID	IETESTCD ^A	IETEST ^B	IECAT ^C	IEORRES ^D
ABC-123	IN02	Demonstrated ability to properly operate the eDiary	INCLUSION	N
XYZ-456	EX05	Participation in another investigational trial during the 30 days prior to the Run-in Period or during this study	EXCLUSION	Y

^A The IETESTCD variable is the code pertaining to the specific inclusion/exclusion criteria as described in sponsor documentation

^B The IETEST variable is the specific inclusion/exclusion criteria as described in sponsor documentation

^C The IECAT variable indicates whether the criteria in IETEST is an “Inclusion” or “Exclusion” criteria

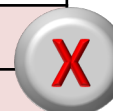
^D The IEORES variable indicates whether the subject met the criteria (“Y”) or did not meet the criteria (“N”) as written in the IETEST variable

Missing Reference Ranges in LB

Findings	The Laboratory (LB) domain contains numeric records missing the Upper Limit of Normal (LBSTNRHI)
Impact	Without reference ranges, it will not be possible to determine if a record should be considered “Low”, “Normal”, or “High”

- ❖ Example records in the LB domain. The first set of records provides reference ranges (LBSTNRHI/LBSTNRLO), allowing an easy determination of normality. The second set of records is missing reference ranges, leading to ambiguity in the data:

DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBSTRESN	LBSTNRLO	LBSTNRHI	LBNRIND
LB	ABC-123	1	ALT	42	30	65	Normal
LB	ABC-123	2	ALT	84	30	65	High
LB	ABC-456	1	ALT	71			
LB	ABC-456	2	ALT	36			



Reference Participation End Date

Findings	Per the SDTM Implementation Guide¹, the Reference Participation End Date (RFPENDTC) is derived as <i>“Date when subject ended participation or follow-up in a trial [...]. Should correspond to the last known date of contact.”</i>
Impact	Deriving this variable as the last subject visit or last treatment date without considering all SDTM datasets may lead to unexpected results

❖ Consider the following records in the AE, DS, EX and SV domains:

DOMAIN	USUBJID	AEDECOD	AESER	AESTDTC	AEENDTC
AE	XYZ-654	Influenza	Y	2015-04-13	2015-04-22

DOMAIN	USUBJID	DSCAT	DSDECOD	EPOCH	DSSTDTC
DS	XYZ-654	Disposition Event	COMPLETED	Follow-Up	2015-12-05

DOMAIN	USUBJID	EXTRT	EXDOSE	EXSTDTC	EXENDTC
EX	XYZ-654	Drug A	100	2015-02-10	2015-12-01

DOMAIN	USUBJID	VISIT	EPOCH	SVSTDTC	SVENDTC
SV	XYZ-654	Visit 10	Follow-Up	2015-12-15	2015-12-15

The last date identified in the SDTM datasets is used to derive RFPENDTC:

DOMAIN	RFPENDTC
DM	2015-12-25

Implementation of ACTARM

Findings	- Subjects in the Demographics (DM) domain have ACTARM populated but are missing exposure records - It is unclear if the subjects are missing exposure records or were not treated
Impact	Using ACTARM to create safety analyses will include subjects that potentially were not treated

❖ Consider the following two subjects:

USUBJID	RFSTDTC	RFENDTC	ARM	ACTARM
XYZ-123	2016-08-12	2016-11-11	Drug A 50mg	Drug A 50mg
XYZ-135			Drug B 100mg	Drug B 100mg

❖ To clarify that subject XYZ-135 was, not treated, the ARMNRS (Reason Arm and/or Actual Arm is Null) variable can be used to clarify why the subject was not treated:

USUBJID	RFSTDTC	RFENDTC	ARM	ACTARM	ARMNRS
XYZ-135			Drug B 100mg		Assigned, Not Treated

Medical History after Subject Start Date

Findings	Records in the Medical History (MH) domain begin after the subject has started treatment
Impact	These records will be precluded from being considered “Treatment Emergent” and will be excluded from analyses of Adverse and Clinical Events

❖ Example records of Medical History events occurring after the subject begins treatment:

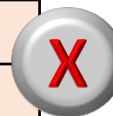
DM			MH	
USUBJID	RFSTDTC	RFENDTC	MHDECOD	MHSTDTC
123123	2011-07-29	2012-08-06	Coronary Artery Disease	2012-05
456456	2010-10-30	2011-01-15	Incisional drainage	2010-12

AE Outcomes and End Dates

Findings	- Records in the Adverse Events (AE) domain listed as “Recovered/Resolved” are missing end dates while records listed as “Not Recovered/Not Resolved” have end dates provided - It may be contradictory to say that an adverse event is not recovered but then to provide an end date, or to say an adverse event has resolved but not to provide an end date
Impact	Missing dates create ambiguity in adverse event data and can make AE duration analysis difficult

- ❖ In the example records below, the first two records provide consistent information while the second two provide contradictory information with respect to the outcome (AEOUT) and the end date (AEENDTC):

AEDECOD	AEOUT	AESTDTC	AEENDTC
Fatigue	NOT RECOVERED/NOT RESOLVED	2016-01-24	
Headache	RECOVERED/RESOLVED	2016-01-25	2016-01-25
Erythema	RECOVERED/RESOLVED	2016-03-21	
Nasal Congestion	NOT RECOVERED/NOT RESOLVED	2016-05-14	2016-05-16



Inconsistent Units in LB

Findings	- Multiple values for Standard Laboratory Units (LBSTRESU) exist within the same lab test (LBTEST) - This may be due to the sponsor providing both conventional and SI units
Impact	Analyzing lab results across domains may require data manipulation to address inconsistencies; conducting analyses without manipulation may lead to unexpected results

❖ Example ALB records with inconsistent values for standard unit:

DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBSTRESN	LBSTRESU	VISIT
LB	ABC-123	1	ALB	4.7	g/dL	WEEK 4
LB	ABC-123	2	ALB	4.3	g/dL	WEEK 8
LB	ABC-456	1	ALB	38	g/L	WEEK 4
LB	ABC-456	2	ALB	47	g/L	WEEK 8

