

Common Data Quality Issues Seen During FDA’s JumpStart Service

Alan Shapiro¹; Tejas Patel¹; Jesse Anderson¹; Kathryn Matto²; Caitlin Anderson²; Aaron Belowich²; Sarah Ferko²; David Jacobs²
¹FDA CDER OTS Office of Computational Science, ²IBM



Background

The JumpStart service provides CDER review teams with an assessment of the composition, conformance, and quality of the clinical trial data submitted within an application. From inception, the JumpStart service has provided this assessment to over 170 applications. Using automated tools and additional investigation, the JumpStart team reviews errors, issues, and warnings that may impact the review. JumpStart empowers reviewers to make more data-driven decisions on data quality and readiness for analysis. Some issues warrant further scrutiny and/or may hinder a reviewer’s ability to fully analyze the safety and efficacy of the investigational product. Ambiguity in the data, coding discrepancies, or faulty metadata may require clarification from applicants, resulting in delays. This poster will highlight some common, impactful issues that the JumpStart team has encountered and provide suggestions for improving data quality.

Missing Reference Start Date

Reference Start Date/Time (RFSTDTC) often lists the subject’s first exposure to study treatment in the Demographic (DM) domain. This variable plays a critical role throughout the SDTM datasets – it is used to derive Study Day (--DY) variables and Baseline (--BLFL) flags:

DOMAIN	RFXSTDTC	RFSTDTC	RFXENDTC	RFENDTC	ACTARM
DM					DRUG A

DOMAIN	LBSEQ	LBTESTCD	LBSTRESN	LBSTRESU	LBBLFL	LB DTC	LB DY
LB	1	ALT	42	U/L		2015-01-15	
LB	2	ALT	34	U/L		2015-02-15	
LB	3	ALT	54	U/L		2015-03-15	

Missing baseline flags will make it harder to discern if elevated readings are because of exposure to treatment or typical for a subject. **JumpStart Recommendation:** All randomized and treated subjects should have RFSTDTC populated in order to derive Study Day variables and baseline flags.

Disposition Coding Quality

As part of a Disposition (DS) dataset, sponsors are required to map a Reported Term (DSTERM) to a Dictionary-Derived Term (DSDECOD), which takes values from an extensible codelist. While there is no dictionary (i.e. MedDRA) to convert DSTERM to DSDECOD, coding terms incorrectly can lead to misleading disposition event counts.

DOMAIN	USUBJID	DSSEQ	DSTERM	DSDECOD
DS	ABC-123	1	WORSENING NAUSEA	OTHER
DS	ABC-456	2	STROKE	ADVERSE EVENT
DS	ABC-789	1	SUBJECT WITHDREW	PROTOCOL VIOLATION

In rows 1 & 3, alternative DSDECODs (such as ADVERSE EVENT and WITHDRAWAL BY SUBJECT) would be more appropriate. Row 2 appears to be coded correctly, as the subject withdrew due to an Adverse Event (Stroke). **JumpStart Recommendation:** Additional time should be spent ensuring that DSTERM is mapped appropriately to DSDECOD.

Computational Algorithms in define.xml

The define.xml file provides a complete listing of all datasets, variables, origin information, and codelists. Some variables, such as study days or flag variables, are derived based on the values of assigned variables. Computational algorithms are expected for derived variables. Below is an example of a derived variable from a define.xml:

Variable	Label	Origin	Derivation/Comment
DTHDTC	Date/Time of Death	Derived	ISO8601(AE.AEENDAT_RAW, AE.AEENTIM) if AE.AEOUT = "Fatal" or AE.AESDTH = "Yes" or ISO8601(ACM.DTHDAT_RAW) if ACM.ALIVED14_STD = "N" or ALIVED28_STD = "N"

In this example, the derivation references “raw” datasets not provided to FDA. This prevents FDA’s ability to fully validate the variable in question, in this case the Date/Time of Death

JumpStart Recommendation: All derived variables should contain detailed and reproducible derivation comments from the SDTM data package.

Seriousness Criteria for SAEs

Serious Adverse Events (SAEs) within the ICH definition must meet one of the following criteria to be considered serious: Results in Death, Life-threatening, Hospitalization, Disability or Permanent Damage, Congenital Anomaly/Birth Defect, Required Intervention, or Other. If one or more of these conditions is true, the event should be flagged as serious with a “Y”. Below are examples of SAEs and qualifiers in AE:

AEDECOD	AESER	AESDTH	AESLIFE	AEHOSP	AESDISAB	AESCONG	AESMIE
Headache	N						
Neutropenia	Y						
Stroke	N				Y		
Chest Pain	Y			Y			Y

In these examples, **Neutropenia** is listed as serious but missing a qualifier, and **Stroke** is not listed as serious but has at least one criteria marked as “Y”. These two are incorrect.

JumpStart Recommendation: All SAEs should have >= 1 qualifier marked as “Y”. Non-SAEs should have null values for all qualifiers.

Guidance for Implementing EPOCH

EPOCH is a categorical variable to measure an interval of time within a study (i.e. Screening, Treatment, Follow-Up). The EPOCH variable allows FDA to easily determine which phase of the trial an observation occurred. Consider a subjects records from the Subject Elements (SE) and Adverse Events (AE) domain:

DOMAIN	SESEQ	ELEMENT	EPOCH	SESTDTC	SEENDTC
SE	1	SCREENING	SCREENING	2015-04-15	2015-05-14
SE	2	DRUG A	TREATMENT	2015-05-14	2015-08-20
SE	3	FOLLOW-UP	FOLLOW-UP	2015-08-20	2015-09-01

DOMAIN	AEDECOD	AEBODSYS	EPOCH	AESTDTC	AEENDTC
AE	Influenza	Infections and Infestation		2015-07	2015-08-10

Even though AESTDTC is a partial date, it still falls within the TREATMENT period

JumpStart Recommendation: All records with --DTC occurring within the dates of an element in SE should have EPOCH populated