

Common Data Quality Issues Seen During FDA’s JumpStart Service

Alan Shapiro¹; Tejas Patel¹; Jesse Anderson¹; Kathryn Matto²;
David Jacobs²; Aaron Belowich²; Sarah Ferko²; Sarah Rodgers²

¹FDA CDER OTS Office of Computational Science, ²IBM



Background

The JumpStart service provides CDER review teams an assessment of the composition, conformance, and quality of the clinical trial data submitted within an application. Since inception, the JumpStart team has assessed data quality for over 220 applications using automated tools paired with additional investigation to review errors, issues, and warnings that may impact the review. JumpStart empowers reviewers to make data-driven decisions on data quality and improve 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 five common, impactful issues that the JumpStart team has identified during the data quality assessment.

Implementation of Death Information

Death information is often found in several domains, including Demographics (DM), Death Details (DD), Adverse Events (AE), and Disposition (DS). Inconsistent and/or missing date information may affect the ability to conduct a full safety evaluation. Below is an example subject with inconsistent information:

DOMAIN	USUBJID	RFSTDTC	RFENDTC	DTHFL	DTHDTC
DM	XYZ-123	2016-08-12	2016-11-11		
DOMAIN	USUBJID	AEDECOD	AESER	AESDTH	AESTDTC
AE	XYZ-123	Stroke	Y	Y	2016-12-01
DOMAIN	USUBJID	DSCAT	DSDECOD	EPOCH	DSSTDTC
DS	XYZ-123	Disposition Event	Death	Follow-Up	2016-12-05

JumpStart: Subject XYZ-123 experienced a stroke resulting in death (AESDTH = ‘Y’) on 2016-12-01, but the DS domain indicates the subject died on 2016-12-05 and the DM domain does not have the Subject Death Flag (DTHFL) or Date/Time of Death (DTHDTC) populated. Discrepancies between dates can make it difficult to ascertain causality of death; in this example, using DTHFL to review subjects who died would exclude this subject.

Guidance for Implementing ACTARM

The Demographics (DM) domain contains treatment arm variables including, Description of Planned Arm (ARM) and Description of Actual Arm (ACTARM). The ACTARM variable indicates if a subject received the planned treatment and can affect safety analyses if populated incorrectly. For the randomized and not treated subject in the example below, ACTARM is incorrectly populated:

DOMAIN	USUBJID	RFSTDTC	RFENDTC	ARM	ACTARM
DM	XYZ-135			Drug B 100 mg	Drug B 100 mg

JumpStart: The presence of ‘Drug B 100 mg’ in ACTARM portrays that the subject was treated, however the Reference Start/End Date/Time (RFSTDTC/RFENDTC) variables are not populated. By validating that the subject does not have a record indicating treatment was received in the Exposure (EX) domain, the subject can be confirmed as randomized and not treated. Using ACTARM to create analyses would include this subject in the treatment population, when the subject should be excluded. Per SDTMIG v3.2, not treated subjects should have ACTARM = ‘Not Treated’.

Adverse Event Outcomes and End Dates

Outcome of Adverse Events (AEOUTs) are populated using CDISC controlled terminology, including values such as ‘Recovered /Resolved’, ‘Recovering/Resolving’, ‘Not Recovered/Not Resolved’, ‘Fatal’, or ‘Unknown’. The outcome of a given adverse event determines whether an End Date/Time of Adverse Event (AEENDTC) should be provided. Inconsistences may exist where adverse events listed as ‘Recovered/Resolved’ are missing end dates while records listed as ‘Not Recovered/Not Resolved’ have end dates provided, below is an example:

DOMAIN	USUBJID	AEDECOD	AEOUT	AESTDTC	AEENDTC
AE	123-123	Fatigue	NOT RECOVERED/NOT RESOLVED	2016-01-24	
AE	123-123	Headache	RECOVERED/RESOLVED	2016-01-25	2016-01-25
AE	456-456	Erythema	NOT RECOVERED/NOT RESOLVED	2016-05-14	2016-05-16
AE	456-456	Congestion	RECOVERED/RESOLVED	2016-03-21	

JumpStart: It is contradictory to list that an adverse event is resolved and not to provide an end date, or classify an adverse event as not recovered when an end date is provided.

Missing Laboratory Reference Ranges

Laboratory results are evaluated against the Upper Limit Reference Range (LBSTNRHI) to identify abnormal test results and whether the result is related to the study treatment. When a Numeric Result/Finding (LBSTRESN) is greater than LBSTNRHI, the reference range indicator is considered ‘High’ and may warrant further investigation, such as calculating Upper Limit of Normal (ULN). Below is an example of a subject missing reference ranges:

DOMAIN	USUBJID	LBSEQ	LBSTESTCD	LBSTRESN	LBSTNRLO	LBSTNRHI	LBNRIND
LB	ABC-123	1	ALT	42	30	65	Normal
LB	ABC-123	2	ALT	80	30	65	High
LB	ABC-456	1	ALT	71			
LB	ABC-456	2	ALT	36			

JumpStart: With missing reference ranges, the review team cannot assess whether the results for Subject ABC-456 are low, normal, or high. Reference ranges vary across tests and subjects, so additional considerations may be needed before directly imputing reference ranges.

Medical History after Subject Start Date

The Medical History (MH) domain contains relevant events and medical information for individual subjects captured prior to the study. These records may help contextualize events and observations that occurred during the study. Records that occur after the Subject Reference Start Date/Time (RFSTDTC) can be misplaced in the MH domain, preventing the event from being considered “treatment emergent”. Below are two examples of medical history events that occurred after RFSTDTC:

DM			MH	
USUBJID	RFSTDTC	RFENDTC	MHDECOD	MHSTDTC
DEF-123	2011-07-29	2012-08-06	Chronic Kidney Disease	2012-05
DEF-456	2010-10-30	2011-01-15	Hypertension	2010-12

JumpStart: Events captured after a subject receives treatment should be captured in either the Adverse Events (AE) or Clinical Events (CE) domain rather than the MH domain so that events can be considered for treatment emergent analyses.