



PP01: SEND Dataset QC Best Practices – Common explanations for FDA / CDISC validation issues

Author: M. Rosentreter (Michael.Rosentreter@bayer.com)

ABSTRACT

Nonclinical datasets (SEND) must be in conformance with validator rules published by health authorities and data standards organizations prior to official submission. Although final datasets reach overall conformance, some validation messages cannot be avoided. This can be due to either the study design or company specific data collection LIMS. The PHUSE SEND QC Best Practice project group collected a list of commonly recurring validation issues from their own experiences. The group presents consolidated standard explanations for these issues along with possible occurrence scenarios for additional guidance.

BACKGROUND

- Some validation issues cannot be avoided
- Intrinsic nature of validation rules don’t capture unforeseen deviations from standards
- Caused by study design or LIMS limitation
- To some degree, choice of validator tool determines validation results

MATERIALS & METHODS

- Six months of sample collection from real world validation issues (different validator tools were used)
- Discussion, Compilation, Categorization

FALSE POSITIVES

- Validation issues (by multiple validator tools) considered as false positives

Rule #	Message	Explanation
FDAB016	Missing xxDY variable when xxDTC variable is present	False validator error. DY is populated, and DTC is null (for both LB and PC). Where MIRESCAT is empty, 'Normal' is the MISTRESC value. The upper- and lowercase text in 'Normal' is not recognized as the equivalent of 'NORMAL', however should be considered to be.
FDAB082*	Missing MIRESCAT value when MISTRESC is not 'NORMAL', 'UNREMARKABLE', or 'NULL' (MISTAT = NOT DONE)	False error. There are no time/date stamps in PPRFTDTC, as only nominal days were recorded for PP
CDISC 63	Missing PPTPTREF variable when PPRFTDTC variable is present	A false positive validator error. This error only occurs when the 7 th decimal shows 4 and the 8 th decimal shows 9
FDAB031	PPSTRESN does not equal PPSTRESC	

* Discussed controversial since the SENDIG is quite specific about matching case of standardized terms.

VALIDATION ISSUES AND EXPLANATIONS			
Fixable; should be remedied before submission			
Rule #	Message	Issue and solution	
FDAB011	Value from NY codelist is populated in TSVAL when TSPARMCD='GLPTYP'	Error occurs when study is non-GLP (GLP=N), yet a TS field for GLP type has been left in the TS file. Solution is to change GLPTYP to 'NONE', re-validate.	
FDAB046	Missing PLANMSUB Trial Set Parameter within SETCD	Only female animals were used in this study. Solution is to include PLANMSUB with TSVAL = 0	
Not avoidable; not remediable because of study design or LIMS			
Rule #	Message	Proposed explanation for nSDRG	Example
142, 150, 168, 169, 188, 191, 192, 195, 199, 206, 208	Duplicated records	A subject might have more than one record associated with one time stamp, e.g. when two findings are observed within the same session, both findings are recorded with the same time-stamp	Clinical finding, collected on the same day and time / Multiple findings per specimen in MI
14	Expected variable with missing value for all records	Per section 4.1.3 of the SENDIG, when no data exist for an expected variable, a null column should still be included in the dataset	Missing value in CO domain (CODTC); DSUSCHFL is null; baseline flags
FDAB017, 16, 33, 40, 49, 181, 189	Values not found in extensible codelist	The extensible terms are included in the table in section 3.3, Nonstandard terminology.	Non Standard terms used in domains.
TCG 4.1.4.1, p147	No baseline flag record in LB for subject	No baseline flags were dictated in the final report, or baseline flag were not specified in report or study plan.	LB
FDAB032	Missing values for both --DTC and --DY variables	Data is based on nominal days (--NOMDY / VISITDY / --ELTM), therefore no --DTC or ---DY is provided	Finding domains / SC domain (SCDTC and SCDY)
CG0425	Missing value for --ORRESU, when --ORRES is provided	Some numeric parameters do not have a unit. For further explanation refer to section 4.x	EG, LB, etc.
CG0425	Missing value for LBORRESU, when LBORRES is provided	LBORRES contains parameter counts for which there are no units to be included in LBORRESU.	LB (CG0425 also for LBSTRESC / LBSTRESU)
265, 266, 267, 269, 270	Model permissible variable added into standard domain	Some comments were longer than 200 characters and required the use of COVAL1, COVAL2, etc.	Adding a SDTM variable for a class that is not included in the SENDIG
FDAB046, 242	No Trial Set with TCNTRL Parameter	No control animals were used in this study, hence Trial Set parameter TCNTRL is not populated	TX
FDAB0008	EXSTDTC date is after EXPENDTC in TS domain	Dose administrations were recorded on same day as experimental end date (necropsy). Definition of TS files' EXPENDTC is limited to Year/Month/Day, whereas EXSTDTC records in EX domain also contain hours and minutes. Thus, the EXSTDTC records occur 'after' the null EXPENDTC date in TS domain.	EX
CG0168, 120	CODTC is populated when comment is a child record of another domain	CODTC is populated due to a system limitation for records with an IDVAR of MIREFID. These comments apply to the entire specimen that was collected, not just the records with this IDVAR appearing in the domain.	CO
CG0168	Inappropriate usage of variables in CO domain when comment is a child record of another domain	The variable MIREFID is populated as an IDVAR value for comments made against records from the same specimen. These comments are not attributed to a single record in the parent domain, therefore the specimen identifier in our data capture system has been used.	CO

ACKNOWLEDGE The author would like to thank all active members of the ‘SEND Dataset QC Best Practices’ project group for constant discussions and commitment. Without their engagement and dedication this poster could not have been realized.

RULE LIMITATIONS?

Although it can be recognized that LIMS limitations introduce certain conformance issues during data collection, these can be corrected in a data curation step during SEND conversion. Known systematic hiccups can be identified and solved, so they will no longer lead to systematic validation issues. However, as data transparency and data integrity must be maintained, seemingly ‘flawed’ collected data (such as duplicated --DTCs for two observations) cannot be resolved. Missing data, for instance –DTCs, should not be derived from collected --DYS. Baseline flags, although mentioned in the Study Data Technical Conformance Guide, cannot be artificially introduced to SEND datasets when not specified by study plan or report. So, it appears as if there are limitations to the rules itself. From the author’s perspective, some conformance rules are of general nature, so they do not apply to all different kinds of studies. Should these general rules be retired in order to circumvent validation issues? The author does not believe so. Most, if not all validation rules which the collected examples failed, are legitimate. Duplicated records can accidentally be created. Values can erroneously be missing. Validation rules have their place.

CONCLUSIONS

- Validation of SEND datasets by automated validator tools assist to remedy conformance issues before submission
- All fixable issues should be remedied before submission to avoid delay in the application process
- Re-iteration of validation ensures all fixable issues are caught, and no new issues were introduced
- Some validation issues cannot be remedied because of study design or LIMS
- All non-avoidable validation issues should be described in the nonclinical Study Data Reviewer’s Guide (nSDRG)
- Standard explanation per validation issue should be used to avoid confusion of consumer (applies only to systematic validation issues)
- Validation issue explanations should be clear and precise