

Getting It Right: Refinement of SEND Validation Rules

Kristin Kelly, Pinnacle 21, LLC

PhUSE US Connect 2020

Paper SA05

Agenda

Introduction

Changes to Existing Rules

New Rules for SEND

Future Improvements

Conclusion

Introduction

- The Study Data Tabulation Model (SDTM) metadata are the foundation for both clinical and nonclinical study data but....
 - Domains and variables are implemented differently via the SENDIG and SDTMIG
- In preparation of submission deliverables, conformance to CDISC standards (e.g., SDTM, SEND, ADaM, Define-XML) is evaluated using automated validation tools
 - Validation rules based on FDA Business Rules and requirements, CDISC rules or other data quality rules
- Though there are several similarities in conformance to SDTM and SEND, there are also differences that need to be accounted for in validation
 - Rules for SEND have been refined by modifying existing rules, removal of others, or creation of new rules

Changes to Existing Rules

- Several rules are applicable to both SDTM and SEND
- Summary of changes to existing rules:
 - Rules that were refined for both standards
 - SEND-specific versions of an existing rule were created
 - Rules deemed to be not applicable for SEND were removed
- Rule IDs in Pinnacle 21 are assigned for a standard designated by a two-character prefix followed by a unique 4-digit number, e.g. SD1148, SE2319
- For SDTM and SEND, these are:
 - **‘SD’** – SDTM rule, can also apply to SEND if it is based on conformance to the SDTM Model or a concept that is identical between the IGs
 - **‘SE’** – SEND rule, only applicable for SEND

SD1117 – Duplicate Records (1)

- Checks for duplicate records in Findings domains where the same test (--TESTCD/--TEST) is done for the same subject (USUBJID/POOLID) at the same timepoint
- Consistently updated by adding Qualifier variables from the SDTM that have a specific definition.
 - Rule logic does not include variables that can be sponsor-defined, e.g., --SPID, --REFID, --LNKID, etc.
- Qualifier and Timing variables recently added considered to be used more often in SEND
 - --ANTREG, --DIR, --PORTOT, VISITDY, --ENDTC, --STINT, --ENINT, --NOMDY

SD1117 – Duplicate Records (2)

- Rule Description also updated to be more generic to account for the number of Qualifier variables that have been added over time

Original Rule Description

The structure of Findings class domains should be one record per Finding Result per subject. No Finding Result with the same Test Short Name (--TESTCD) for the same Subject (USUBJID) and the same Collection Date (--DTC) are expected.

New Rule Description

The structure of Findings class domains should be one record per Finding Result per subject. No Finding Result with the same Test Short Name (--TESTCD) and the same Qualifier variables at the same timepoint for the same Subject (USUBJID) are expected.

SD0006 – No Baseline Flag in Domain for Subject (1)

- Per the TCG, baseline flags, (e.g., last non-missing value prior to first dose) should be submitted for both clinical and nonclinical data, e.g. –BLFL = ‘Y’
 - SDTM - EG, LB, MB, MS, PC, and VS
 - SEND - EG, LB, and VS
- Originally assigned for both SDTM and SEND but.....
 - Returned false positive messages for SEND domains other than EG, LB or VS
 - Rule description is SDTM-focused
 - Refers to ACTARM = ‘Not Treated’ and ARM = ‘Screen Failure’
 - Concepts do not exist in SEND
- SD0006 removed for SEND

SD0006 – No Baseline Flag in Domain for Subject (2)

- New rule created for SEND:

Rule ID	Message	Description	Domains
SE2319	No baseline flag record in domain for subject	All subjects should have at least one baseline observation (--BLFL = 'Y') in EG, LB, and VS domains.	EG, LB, VS

SE2305 – SE2308, SE2318 – Missing Values for All Required Timing Variables Needed for Summarization (1)

- Based on FDAB047, which states:
 - *Required timing variables for identification of the day on which group summaries (group means and incidences) are calculated should be populated for nonclinical data.*
- These rules check that specific timing variables needed for group summaries are populated in the applicable SEND dataset
 - Examples include: --DY, --STDY, VISITDY (v3.0), --NOMDY (v3.1)
 - Originally, the rule logic also included RFSTDTC but was recently updated to only check the timing variables in the corresponding domain

SE2305 – SE2308, SE2318 – Missing Values for All Required Timing Variables Needed for Summarization (2)

Rule ID	Message	Description	Domains
SE2305	Missing values for all required set(s) of timing variables: VISITDY, --STDY, or --NOMDY	At least one of these variables or sets of variables must be present and populated in DS: (VISITDY, DSSTDY, or DSNOMDY). These are the required timing variables for identification of the day on which group summaries (Group Means and Incidences) are calculated for post-mortem data.	DS
SE2306	Missing values for all required set(s) of timing variables: VISITDY, --DY or --NOMDY	At least one of these variables or sets of variables should be populated: VISITDY, --DY, or --NOMDY. These are the Required timing variables for identification of the day on which group summaries (Group Means and Incidences) are calculated.	BW, CL, CV, EG, LB, PM, RE, VS
SE2307	Missing values for all required timing variables: --DY & --ENDY	These variables must be populated: --DY and --ENDY. These are the required timing variables for identification of the day on which group summaries (Group Means and Incidences) are calculated for data associated with a time interval.	BG, FW
SE2308	Missing values for required timing variables: (VISITDY and --ELTM) or (--NOMDY and --ELTM)	PCNOMDY and PCELTM or VISITDY and PCELTM must be present and populated in PC. These are the required timing variables for identification of the day on which group summaries (Group Means and Incidences) are calculated for pharmacokinetic concentrations.	PC
SE2318	Missing values for all required set(s) of timing variables: VISITDY or --NOMDY	At least one of these variables or sets of variables must be present and populated in PP: VISITDY or PPNOMDY. These are the required timing variables for identification of the day on which group summaries (Group Means and Incidences) are calculated for pharmacokinetic parameters.	PP

New Rules for Trial Summary (TS)

- In the TCG, requested TS parameters were listed for both SDTM and SEND
- P21 already included rules for parameters that were designated as ‘Y’ for ‘Should Include’ in the SENDIG v3.0 and v3.1
- Additional parameters that were not required per the SENDIG were now included
 - New rules were created to check for these additional parameters requested by FDA
 - Note: Only those parameters flagged in the TCG as ‘FDA Desired – Nonclinical’ = ‘Y’ had a rule created to check for presence in TS. Those parameters designated as ‘Conditional’ were not included

New Rules for Trial Summary (TS)

Rule ID	Message	Description
SE2321	Missing SLENGTH Trial Summary Parameter	'Study Length' (SLENGTH) record should be populated in Trial Summary (TS) domain. It is used for Janus Nonclinical data load blocking.
SE2327	Missing SPLANSUB Trial Summary Parameter	'Planned Number of Subjects' (SPLANSUB) record should be populated in Trial Summary (TS) domain. It is used for Janus Nonclinical data load blocking.
SE2328	Missing STENDTC Trial Summary Parameter	'Study End Date' (STENDTC) record should be populated in Trial Summary (TS) domain.
SE2329	Missing DOSENDTC Trial Summary Parameter	'End Date/Time of Dose Interval' (DOSENDTC) record should be populated in Trial Summary (TS) domain.
SE2330	Missing DOSSTDTC Trial Summary Parameter	'Start Date/Time of Dose Interval' (DOSSTDTC) record should be populated in Trial Summary (TS) domain.
SE2331	Missing GLPFL Trial Summary Parameter	'Good Laboratory Practice Flag' (GLPFL) record should be populated in Trial Summary (TS) domain.
SE2332	Missing PCLASS Trial Summary Parameter	'Pharmacologic Class' (PCLASS) record should be populated in Trial Summary (TS) domain.
SE2333	Missing PDOSFRQ Trial Summary Parameter	'Planned Dose Frequency' (PDOSFRQ) record should be populated in Trial Summary (TS) domain.
SE2334	Missing SEXPOP Trial Summary Parameter	'Sex of Participants' (SEXPOP) record must be populated in Trial Summary (TS) domain.
SE2335	Missing TRTUNII Trial Summary Parameter	'Primary Treatment Unique Ingredient ID' (TRTUNII) record must be populated in Trial Summary (TS) domain.

New Rules Based on FDA Business Rules (1)

- Beginning with FDA Business Rules v1.4, several new rules have been added for nonclinical
 - Many are specific to *tumor.xpt* but some others could be automated to check SEND datasets
- FDAB082
 - For carcinogenicity studies, MIRESCAT should be populated unless MISTRESC has a value of 'UNREMARKABLE' (for SENDIGv3.1) or 'NORMAL' (for SENDIGv3.0) or MISTAT has a value of 'NOT DONE'.*
- Based on this, the following rules were created:

Rule ID	Message	Description
SE2228	Missing --RESCAT value, when --STRESC is not 'NORMAL', 'UNREMARKABLE', or NULL (--STAT = 'NOT DONE')	Result Category (--RESCAT) should be populated for all abnormal findings, i.e., when Result or Finding in Standard Format (--STRESC) has a value other than 'NORMAL', 'UNREMARKABLE' or missing (Completion Status (--STAT) = 'NOT DONE').
SE2233	--RESCAT should not be populated, when --STRESC is 'NORMAL', 'UNREMARKABLE', or NULL (--STAT = 'NOT DONE')	Result Category (--RESCAT) should not be populated when Result or Finding in Standard Format (--STRESC) is 'NORMAL', 'UNREMARKABLE' or missing (Completion Status (--STAT) = 'NOT DONE').

New Rules Based on FDA Business Rules (2)

- FDAB085
 - *For findings in MISTRESC using the NEOPLASM controlled terminology list, malignancy status in MIRESCAT should align with any benign or malignant designation in NEOPLASM. Explain any inconsistencies in the nSDRG.*
- New rule implemented that checks the consistency between MISTRESC and MIRESCAT and assigned for SENDIG v3.0 and v3.1

Rule ID	Message	Description
SE2229	--STRESC/--RESCAT mismatch when --RESCAT = 'BENIGN' or 'MALIGNANT'	Result or Finding in Standard Format (--STRESC) values from (NEOPLASM) CT codelist should align with the malignancy status in Result Category (--RESCAT) when --RESCAT = 'BENIGN' or 'MALIGNANT'.

New Rules Based on FDA Business Rules (3)

- FDAB065
 - *DS, CL, EG, EX, LB, MA, MI, PC, PP, and VS should be submitted if collected.*
- Two rules that were historically assigned only for SDTM were recently assigned for SEND v3.0 and v3.1
 - Though most domains in this list would depend on the study design, subjects must always be dosed (EX) and must have a disposition record (DS) in SEND

Rule ID	Message	Description
SD1109	Missing EX dataset	Exposure (EX) dataset should be included in every submission.
SD1110	Missing DS dataset	Disposition (DS) dataset should be included in every submission.

Future Improvements

- In development:
 - Updates to existing TS rules to also check TX
 - Both domains share codelists for –PARMCD/--PARM as well as conventions/formats for populating –VAL per parameter, e.g. date parameters in ISO 8601 format
 - Codelist changes for variables from SEND v3.0 to v3.1
 - Example: VSORRESU/VSSTRESU changed from the ‘VSRESU’ codelist to the ‘UNIT’ codelist
- Future Plans:
 - Implementation of the CDISC SEND v3.0 Conformance Rules
 - Automation of FDA BR that check traceability between *tumor.xpt* and SEND datasets
 - Additional useful data quality checks

Conclusion

- Historically, development of validation rules for SEND has lagged behind SDTM
- SEND rule set has been refined to better check against guidance in the SENDIG as well as FDA Business and Validator rules
- To fully utilize the updated SEND rules, it is recommended to use the latest version of the Validator in order to align with what may be seen by FDA when the data is submitted
- Using the most recent version of the rules may ensure increased turnaround time for review

Questions?

Name: Kristin Kelly

Organization: Pinnacle 21, LLC

E-mail: kkelly@pinnacle21.com