PharmaSUG 2020 - Paper SS-325

Getting It Right: Refinement of SEND Validation Rules

Kristin Kelly, Pinnacle 21, Blue Bell, PA, USA

ABSTRACT

The CDISC SDTM metadata, outlined in the SDTM Model, are used for submission of data from both clinical trials and nonclinical studies. Until recently, many of the Pinnacle 21 validation rules were assigned for both SDTM and SEND domains when in some cases, a specific rule did not apply for SEND data as outlined in the SENDIG. Over the past year, the SEND rule set has been refined through the modification of existing rules, removal of others and creation of new rules. All rules are based on either an FDA Business rule, an FDA Validator rule, or CDISC rules. This paper will discuss some of the changes that have been made in an effort to 'get the rules right' for SEND.

INTRODUCTION

Though the Study Data Tabulation Model (SDTM) metadata are the foundation for both clinical and nonclinical study data, the domains and variables are implemented differently. Because of this, there is an implementation guide for each, the Standard for the Exchange of Nonclinical Data Implementation Guide (SENDIG) for nonclinical and the Study Data Tabulation Model Implementation Guide (SDTMIG) for clinical. When the study data packages are prepared for submission, the study data and other deliverables (e.g., define.xml) must be evaluated for conformance to the standard, in this case, SDTM or SEND. This is typically handled by using automated validation tools, such as Pinnacle 21, that include rules that have been published by FDA or CDISC as well as data quality rules.

Since it is quite recent that FDA and CDISC have begun publishing rules and validation has been performed prior to this, Pinnacle 21 created automated validation rules based on the SDTM, SEND, ADaM, and the Define-XML standards to aid sponsors in preparing for submission. Though there are several similarities in conformance for SDTM and SEND, there are also several differences that need to be accounted for in the validation rules based on the guidance in each implementation guide (IG). Over the past year, the rules for SEND have been refined in an effort to check those aspects that are specific to guidance provided in the SENDIG or are requirements from FDA. This paper will discuss some of the updates that have been made in order to provide a more robust rule set for SEND.

CHANGES TO EXISTING RULES

As mentioned, there are several rules that are applicable to both SDTM and SEND. For some, the rule was refined for both standards but for others, SEND-specific versions of the existing rule had to be created and assigned a new Rule ID. Some SDTM rules were removed completely for SEND as well. Currently, the Rule IDs in Pinnacle 21 are assigned for a specific standard that is designated by a two-character prefix followed by a unique 4-digit number. For SDTM and SEND, these are the following:

- '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

This rule checks for the presence of duplicate records in Findings domains where the same test (--TESTCD/--TEST) is being done for the same subject (USUBJID/POOLID) at the same timepoint. This check is consistently updated by adding Qualifier variables from the SDTM that have a specific definition. Thus, the rule logic does not include variables that can be sponsor-defined, e.g.., --SPID, --REFID, --LNKID, etc.

Some of the Qualifier variables that have recently been added are designated as 'SEND Only' variables or additional Timing variables that also apply to SDTM in order to reduce false positives for nonclinical Findings domains:

- --ANTREG, --DIR, --PORTOT, VISITDY, --ENDTC, --STINT, --ENINT, --NOMDY

The Rule Description for SD1117 was also updated to be more generic due to the additional Qualifier variables that have been added to the rule logic 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.

Updated 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.

SD1078 – PERMISSIBLE VARIABLE WITH MISSING VALUE FOR ALL RECORDS

This is a data quality rule that fires when there is a variable in a dataset that has a 'Core' of 'Permissible' but all the values are null. It is recommended to drop null Permissible variables in an effort to reduce file size. SEND datasets are typically created by the same tool that is collecting the data and may include all Permissible variables in the exported XPT file whether they are populated or not. These null permissibles are typically not dropped from SEND datasets as there is no mapping performed as with SDTM data. Also, there is no guidance in either the SDTMIG nor the SENDIG that states that these variables should be dropped if no data was collected to populate them and further states that keeping them is a sponsor decision. Due to these circumstances, SD1078 was removed for SEND and downgraded from a 'Warning' to a 'Notice' in SDTM.

SD0006 – NO BASELINE FLAG IN DOMAIN FOR SUBJECT

Per the Technical Conformance Guide (TCG)[1], baseline flags, (e.g., last non-missing value prior to first dose) should be submitted for both clinical and nonclinical data. For SDTM, baseline flag should be present for all treated subjects in the following domains: EG, LB, MB, MS, PC, and VS. For SEND, this flag should be present in EG, LB, and VS.

At first, SD0006 was assigned for both SDTM and SEND but later, it was reported as a false positive for SEND when the rule fired for domains other than EG, LB, or VS. Also, the Rule Description for SD0006 is very SDTM-focused in that it refers to screen failures (ARMCD = 'SCRNFAIL') and not treated subjects (ACTARMCD = 'NOTTRT'). These concepts do not exist in SEND. For these reasons, SD0006 was removed for SEND and a SEND-specific rule was created to check for baseline flags:

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

SE2305 – SE2308, SE2318 – MISSING VALUES FOR ALL REQUIRED TIMING VARIABLES NEEDED FOR SUMMARIZATION

These rules are based on the FDA Business rule, FDAB047, which states the following:
Required timing variables for identification of the day on which group summaries (group means and incidences) are calculated should be populated for nonclinical data.[2]

These rules check that specific timing variables needed for group summaries are populated in SEND domains. In earlier versions of this rule, the presence of RFSTDTC for each subject was also included in the rule logic but then they were updated to only check timing variables in the corresponding domain based on feedback from the FDA Nonclinical Working Group. The updated rules are outlined in the table below:

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

ADDITION OF NEW RULES

New rules that have recently been added include either creation of a SEND-specific rule that used to be handled by one rule for both SDTM and SEND or based on an FDA Business rule.

NEW RULES FOR TRIAL SUMMARY DOMAIN (TS)

Beginning with v4.2 of the TCG, TS parameters requested to be included were listed for both SDTM and SEND. Based on this, several rules were created to check that a record for that particular parameter was included in TS. P21 had already included rules for parameters that were designated as 'Y' for 'Should Include' in the SENDIG v3.0 and v3.1. In the TCG, additional parameters that were not required per the SENDIG were included, thus, new rules were created to check for these additional parameters requested by FDA. Please note that only those parameters that were 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.

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.

SEND-SPECIFIC RULES FOR UNIQUE TIMEPOINTS

In SDTM, --TPT (Planned Timepoint Name) and –TPTNUM (Planned Timepoint Number) should be unique, i.e., a one-to-one relationship, within a domain for the same –ELTM (Planned Elapsed Time from Time Point Ref). Checking for this uniqueness is handled by the three rules noted below. These rules are based on the CDISC rule, CG0240, which states that ''--TPT and --TPTNUM have a one-to-one relationship'.[3]

- SD1125 - Inconsistent value for --TPT within --ELTM
- SD1126 - Inconsistent value for --TPT within --TPTNUM
- SD1127 - Inconsistent value for --TPTNUM within --TPT

But in SEND, timepoint variable pairs need only be unique within –CAT (Category) and –SCAT (Subcategory) and not within an entire domain. Because of this, SD1125 – SD1127 were removed for SEND since false positive messages were generated. Three new SEND-specific rules to check timepoints were then created and assigned to Findings domains for SEND v3.0 v3.1.

Rule ID	Message	Description
SE1125	Inconsistent value for --ELTM within --TPT, --TPTREF, --CAT and/or --SCAT	All values of Planned Elapsed Time (--ELTM) variable should be the same for a given value of Planned Time Point Name (--TPT) variable, Time Point Reference (--TPTREF) and if present, Category and/or Subcategory (--CAT/--SCAT).

Rule ID	Message	Description
SE1126	Inconsistent value for --TPT within --TPTNUM, --DY, --CAT and/or --SCAT	All values of Planned Time Point Name (--TPT) variable should be the same for a given value of Planned Time Point Number (--TPTNUM) variable. Uniqueness for these time points is determined by a combination of the domain, study day, and, if present, category and/or subcategory (--CAT/--SCAT).
SE1127	Inconsistent value for --TPTNUM within --TPT, --DY, --CAT and/or --SCAT	All values of Planned Time Point Number (--TPTNUM) variable should be the same for a given value of Planned Time Point Name (--TPT) variable. Uniqueness for these time points is determined by a combination of the domain, Study Day (--DY), and, if present, category and/or subcategory (--CAT/--SCAT).

RULES BASED ON NEW FDA BUSINESS RULES

Beginning with the FDA Business Rules (BR) v1.4, several new rules were added for nonclinical. Though several were specific to the 'tumor.xpt' file, there were a few that could be automated to check SEND datasets.

The first is FDAB082 that states the following:

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'.[2]

Two validation rules were created and assigned to the Microscopic Findings (MI) domain for SEND v3.0 and v3.1 to determine if the FDA BR was being followed.

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').

Another new FDA BR is FDAB085 which is:

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. [2]

A new rule was implemented that checks the consistency between MIRESCAT and MISTRESC for both SEND 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'.

There were two validation rules that were historically only assigned for SDTM that were recently added for SEND as well based on FDAB065 that states: *DS, CL, EG, EX, LB, MA, MI, PC, PP, and VS should be submitted if collected.* [2]

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.

For most of the domains in this list, one would not readily know if they were collected and would have to check the study report. However, DS and EX are required for every study in SEND because all animals must be dosed (EX) and must have a disposition record in DS.

OTHER MISCELLANEOUS ADDITIONS/FUTURE DEVELOPMENT

As mentioned earlier, the rule, SD1117 that checks for duplicate records in Findings domains has seen its share of updates by adding more Qualifier and Timing variables. There is also another rule, SD1201, that checks for duplicate records in Events domains. Since there is only one Events domain in SEND, DS, this is only assigned for this domain in SEND v3.0 and v3.1. It was recently requested to add a rule that would check for duplicates in the SEND EX dataset, an Interventions domain. At the time, P21 did not contain a rule for Interventions domains for either SDTM or SEND so a new duplicate records rule has since been created and assigned for both standards.

Rule ID	Message	Description
SD1352	Duplicate records in domain	The structure of Interventions class domains should be one record per Intervention Episode per subject. No Interventions with the same Treatment (--TRT), Decoded Term (--DECOD), Category (--CAT), Subcategory (--SCAT), Route (--ROUTE), Location (--LOC) and Treatment Vehicle (--TRTV) values for the same Subject (USUBJID) and the same Start Date (--STDTC) are expected.

Also recently added for both SDTM and SEND were a few data quality rules that check that –STINT and –ENINT are populated in the Pharmacokinetics Parameter (PP) dataset when A) PPTEST contained the standard text ‘T1 to T2’ indicating a time interval that should be described in PPSTINT and PPENINT and B) that –STINT and –ENINT were both populated in the dataset where applicable.

Rule ID	Message	Description
SD1348	Missing values for --STINT and --ENINT variables when --TEST references "T1 to T2"	Planned Start of Assessment Interval (--STINT) and Planned End of Assessment Interval (--ENINT) should be populated when Name of Measurement, Test or Examination (--TEST) is over a time interval, 'T1 to T2', where 'T1' is represented by the value in --STINT and 'T2' is represented by the value in --ENINT.
SD1350	Missing value for --ENINT, when --STINT is provided	Planned End of Assessment Interval (--ENINT) should not be NULL, when Planned Start of Assessment Interval (--STINT) is populated.
SD1351	Missing value for --STINT, when --ENINT is provided	Planned Start of Assessment Interval (--STINT) should not be NULL, when Planned Start of Assessment Interval (--ENINT) is populated.

In development are updates to some existing Trial Summary (TS) rules to also check the Trial Sets (TX) domain since both domains share the same codelist for –PARMCD/--PARM as well as the same conventions for populating TSVAL/TXVAL for specific parameters, e.g., all date parameters should be in ISO 8601 format, etc.

There were also some codelist changes for variables between SENDIG v3.0 and v3.1 and these updates to the metadata will be implemented in the future. One example of this is for the VSORRESU/VSSTRESU variables in VS that were subject to the VSRESU codelist in SENDIG v3.0 but it was changed to the UNIT codelist in SENDIG v3.1.

Future plans for SEND rules also include implementation of the CDISC SEND Conformance Rules for SEND v3.0 and v3.1. Publication of the CDISC rules for SEND v3.0 is imminent with rules for v3.1 soon to follow. After publication by CDISC, the rules can be implemented in P21 and made available to sponsors to use for validation.

CONCLUSION

Historically, development of validation rules for SEND has lagged behind SDTM but recently, the SEND rule set has been refined to better check against guidance in the SENDIG as well as adherence to FDA Business and Validator rules. Though not all the changes made over time were outlined here, some of the highlights of changes were discussed. 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 may ensure an increased turnaround for review in an effort to get drugs to patients faster.

REFERENCES

[1] U.S. Food & Drug Administration, Study Data Technical Conformance Guide (current version at time of access – v4.4/ October 2019), Accessed January 2020 - https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources

[2] U.S. Food & Drug Administration, FDA Business Rules (current version at time of access – v1.5/June 2019 Accessed October 2019 - https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources

[3] SDTMIG v3.2 Conformance Rules, SDS Sub-Team for SDTM Conformance Rules. Version 1.0, December 2016

[4] Standard for Exchange of Nonclinical Data Implementation Guide: Nonclinical Studies Clinical Data Interchange Standards Consortium (CDISC) Standard for Exchange of Nonclinical Data Team. Version 3.1. June 2016

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Kristin Kelly
Pinnacle 21
kkelly@pinnacle21.com