Getting It Right: Refinement of SEND Validation Rules

Kristin Kelly, Pinnacle 21

May 28, 2020

PRESENTER — KRISTIN KELLY

- ▸ Principal Consultant at Pinnacle 21

- ▸ More than 10 years experience at both Pharma and CROs

- ▸ CDISC SEND Core Team, SDS Team, GGG voting member

- ▸ Authorized trainer for CDISC in SDTM

- ▸ SME for nonclinical and clinical FDA projects for CDER and CBER

P 21

UPDATE FROM PREVIOUS P21 WEBINAR

‣ 'Confusing Validation Rules Explained' – Michael Beers (April 30, 2020)

 ‣ Stay tuned for future blog posts to answer the questions received!

 ‣ THANKS TO ALL THAT WERE ABLE TO JOIN THIS SUCCESSFUL WEBINAR SESSION!!!

INTRODUCTION (1)

- Why refine the SEND rules at all?
  - When a submission is received at FDA, it is run through FDA DataFit (FDA's instance of Pinnacle 21 Enterprise)
  - If a reviewer requests a Kickstart service, the data packages are also run through FDA DataFit
    - Kickstart Analysts review the validation reports and provide findings to the reviewers
  - A robust rule set that can be automated is very useful to any reviewer that may encounter the data rather than manual checking
    - This has led to efforts to refine the SEND rules as much as possible for both FDA reviewers and sponsors

P21

INTRODUCTION (2)

- 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

CHANGES TO EXISTING RULES

▸ Rule IDs in Pinnacle 21 are assigned for a CDISC 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 model 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.

P21

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

SD0006 – NO BASELINE FLAG IN DOMAIN FOR SUBJECT

▸ 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:

Rule Description
All subjects should have at least one baseline observation ($--BLFL$ = 'Y') in EG, LB, MB, MS, PC and VS domains, except for subjects who failed screening (ARMCD = 'SCRNFAIL') or were not fully assigned to an Arm (ARMCD = 'NOTASSGN') or were not treated (ACTARMCD = 'NOTTRT').

      ▸ Refers to ACTARM = 'Not Treated' and ARM = 'Screen Failure'

        ▸ Concepts do not exist in SEND

▸ SD0006 removed for SEND

SE2319 – NO BASELINE FLAG IN DOMAIN FOR SUBJECT

‣ New rule created to be SEND specific:

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

‣ Rule logic also modified to not fire if there are no records for a subject in that domain

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

P 21

SD1078 – PERMISSIBLE VARIABLE WITH MISSING VALUE FOR ALL RECORDS (1)

- Data quality rule to check variables with 'Core' of 'Permissible' but all values are null in the dataset
- SEND datasets are typically created by the same tool that is collecting the data
  - Sometimes all Permissible variables included 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

SD1078 – PERMISSIBLE VARIABLE WITH MISSING VALUE FOR ALL RECORDS (1)

▸ 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*

  ▸ Keeping them is a sponsor decision

▸ Outcome:

  ▸ Downgraded to a 'Notice' for SDTM

  ▸ Removed for SEND

NEW RULES FOR TRIAL SUMMARY (TS)

▸ P21 already includes rules for parameters that are designated as 'Y' for 'Should Include' in the SENDIG v3.0 and v3.1

Should Include	TSPARMCD	TSPARM	Type	Controlled Terms, Codelist, or Format	CDISC Notes
Yes	ROUTE	Route of Administration	Char	(ROUTE)	The delivery method by which the Test Article is administered to the subjects. Subject-level dosing is recorded in the Exposure (EX) domain. TSGRPID may be used to relate ROUTE records to specific TRT records as needed.
Yes	SNDIGVER	SEND Implementation Guide Version	Char	(SNDIGVER)	The SEND Implementation Guide version used for the submission containing this dataset. Only a single record should be provided in the Trial Summary dataset.
Yes	SNDCTVER	SEND Controlled Terminology Version	Char		The SEND Controlled Terminology version used for the study dataset for this submission. An example of the appropriate format for TSVAL associated with this PARM/PARMCD combination is "SEND Terminology 2011-01-07" where "SEND Terminology" is the base name of the controlled terminology file, and "2011-01-07" is the publication date of the file in ISO 8601 format. The publication date can be found in the "SourceSystemVersion" attribute of the ODM tag in the ODM XML version of the controlled terminology file. Only a single record should be provided in the Trial Summary.

NEW RULES FOR TRIAL SUMMARY (TS)

▸ In the FDA TCG, requested TS parameters are listed for both SDTM and SEND

  ▸ New rules created to check for these additional parameters requested by FDA

FDA Desired - Nonclinical	TSPARMCD	TSPARM	FDA Notes
Y	SNDIGVER	SEND Implementation Guide Version	
Y	SPECIES	Species	
Y	SPLANSUB	Planned Number of Subjects	
Y	SPLRNAM	Test Subject Supplier	
Y	SPREFID	Sponsor's Study Reference ID	
Y	SSPONSOR	Sponsoring Organization	
Y	SSTYP	Study Type	
Y	STCAT	Study Category	
Y	STDIR	Study Director	
Conditional	STENDTC	Study End Date	If applicable.

▸ 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

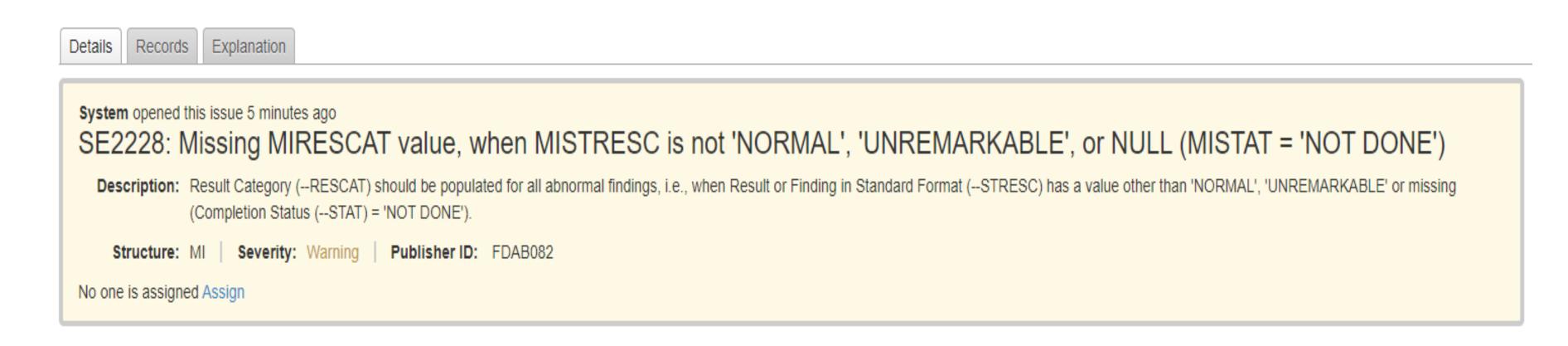# NEW RULES BASED ON FDA BUSINESS RULES (2)

‣ 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').

P21

NEW RULES BASED ON FDA BUSINESS RULES (3)

▸ SE2228 Example:

NEW RULES BASED ON FDA BUSINESS RULES (4)

▸ 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'.

P 21

NEW RULES BASED ON FDA BUSINESS RULES (5)

▸ SE2229 Example:

Details | Records | Explanation

System opened this issue 13 minutes ago
SE2229: MISTRESC/MIRESCAT mismatch when MIRESCAT = 'BENIGN' or 'MALIGNANT'

Description: 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'.

Structure: MI | **Severity:** Warning | **Publisher ID:** FDAB085

No one is assigned Assign

Details | Records | Explanation

🔍 Search   🖨 Print | 📋 Copy | 📥 Download

Record ▲	USUBJID ⇵	MISTRESC ⇵	MIRESCAT ⇵	MIDTC ⇵
1	PDS2014-0001	ADENOSARCOMA, BENIGN	MALIGNANT	2011-01-09T09:59:00
2	PDS2014-0001	BASALIOMA, BENIGN	MALIGNANT	2011-01-09T09:59:00
3	PDS2014-0001	CARCINOMA, C-CELL, MALIGNANT	BENIGN	2011-01-09T09:59:00
7	PDS2014-0001	NORMAL	MALIGNANT	2011-01-09T09:59:00
13	PDS2014-0001	Inflammation chronic, benign	Malignant	2011-01-09T09:59:00
22	PDS2014-0001	NORMAL	BENIGN	2011-01-09T09:59:00
30	PDS2014-0001	Inflammation chronic	BENIGN	2011-01-09T09:59:00

P21

NEW RULES BASED ON FDA BUSINESS RULES (6)

‣ 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 depends 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.

OTHER NEW RULES (1)

▸ SD1352 – Duplicate Records in domain

   ▸ SD1117 - checks for duplicate records in Findings domains

   ▸ SD1201 - checks for duplicate records in Events domains

   ▸ But no rule for Interventions!

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.

OTHER NEW RULES (2)

▸ Data quality checks for −STINT and −ENINT in Pharmacokinetics Parameters (PP)

  ▸ PPTEST contains the standard text 'T1 to T2' indicating a time interval that should be described in PPSTINT and PPENINT

  ▸ −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.

OTHER IMPROVEMENTS (1)

- Updates to existing TS rules to also check Trial Sets (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 additions/changes for variables between versions of the SENDIG (v3.0 to v3.1)
  - Examples:
    - VSORRESU/VSSTRESU changed from the 'VSRESU' codelist to the 'UNIT' codelist
    - MISTRESC in SENDIG v3.1 uses both the NEOPLASM and NONNEO codelists
    - VSTESTCD/VSTEST changed to SVSTSTCD/SVSTST codelist

OTHER IMPROVEMENTS (2)

- In development:
  - Implementation of recently published documents:
    - CDISC SEND v3.0 Conformance Rules v1.0
    - SENDIG – Animal Rule (SENDIG-AR) v1.0
  - Automation of FDA BR that check traceability between tumor.xpt and SEND datasets
  - Additional useful data quality checks
    - New rules that are incorporated into FDA Nonclinical CoreDF reports

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

P 21