Traceability is the Key to Submission: Common Errors in SEND

Submit with Confidence: Know the common issues to look out for in your SEND datasets

May 29th, 2024

About PointCross + Today's Webinar Agenda

- PointCross, Inc. founded in 1999; serving the Life Sciences industry since 2009
- Platinum members of CDISC, Contributor to PHUSE and formal FDA / Industry dialogues (incl. as an FDA Contractor)
- Xbiom™ technology platform for Nonclinical and Clinical data repository, analysis, Metadata Repository, regulatory submission workflows, and data validation
- 7,800+ studies standardized over 10 years; 350-500 studies per year; 60+ Sponsor / CRO customers annually
- Data Standardization Services for Nonclinical (SEND) and Clinical (SDTM, ADaM)
    - SEND Generation
    - 100% QC of SEND datasets
    - Interim study monitoring and study repository
- Intent and purpose of SEND as an analysis **and** submission artifact influences our view that SEND must be inclusive of all study data contained within the Study Report, even those data types requiring custom domains
- Compliance ≠ Cost: committed to offering industry's best value with an automatic SEND quote generator
- Traceability and consistency with the Study Report as the "GLP-artifact" in line with Section 8.3 of the FDA Technical Conformance Guide
- 1st in a series of webinars dedicated to improving the quality of SEND across industry
- Next in our Nonclinical Series: Visualization of datasets for QC, subject-level analysis, cross-study analysis, and repository

What will we cover in this Webinar?

- Common problems with how SEND is generated
- Most prominent issues found in datasets
- Resolving these problems by Reconciling SEND and Study Report
- Next topic: Visualization

Issues in SEND Datasets Prepared with LIMS and SEND Adaptors

- SEND-ASSURE is an independent verification check of SEND datasets prepared by third parties
- 28 recent studies evaluated
- None of these studies were considered “fit for submission” - all would have severely impeded review analysis
- 2% of the issues found needed correction to the Study Report, 98% were related to SEND preparation quality and consistency issues
- Total number of issues per study was between 17 – 20; no reduction over time
- 8%-10% of the total number of issues were serious enough to consider them ”not-submittable”

48 Serious Issues out of 584 Across 28 Studies Listed by Type	Count	How it was Detected
Domain Missing	5	Reconciliation with DSR Mapping to Collected Data Domain Failed
Exclusion flags missed	4	Reconciliation with DSR Summary Tables
Incorrect Date/DY	6	Reconciliation with DSR Summary Tables
Incorrect observation	2	Reconciliation with DSR Incidence Counts
POOLID/POOLDEF missing	1	Reconciliation with DSR Summary Tables
incorrect "Reason not Done"	3	Discovered during SEND analysis & visualization
STRESN populated for categorical dat	2	Validation Check
Supp--CALCN missing	13	Reconciliation with DSR Summary
TD modelling	9	Review of Protocol in Study Report
Timepoint issue	3	Discovered during SEND analysis & visualization
Average # of SEND Issues per Study	**20.9**	
Std.Dev.	**9.6**	

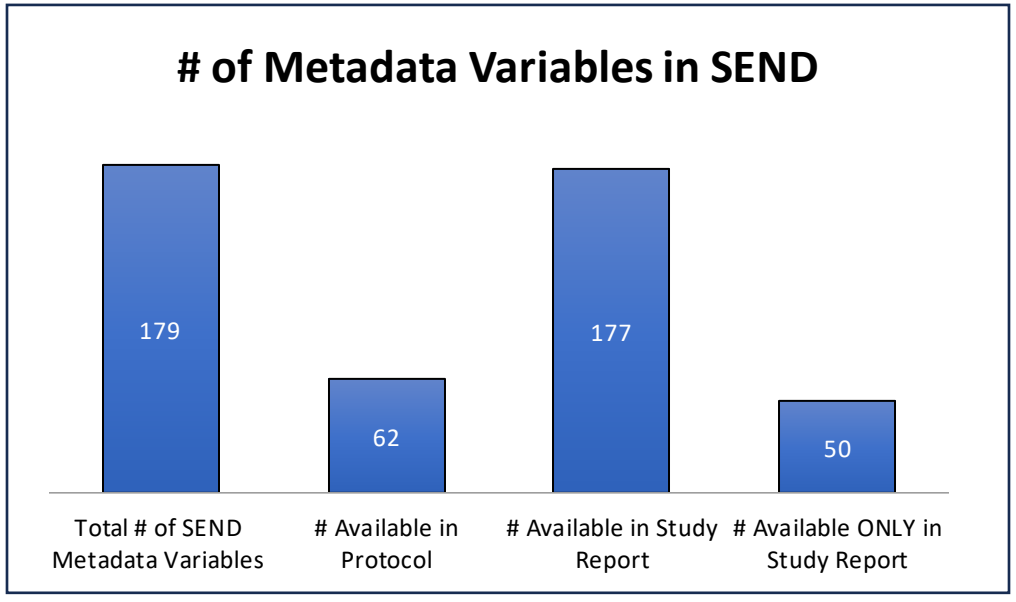
Quality Issues on 41 SEND Datasets from 2 CROs 4-22 to 4-23		
SEND Data Issue Description	**CRO1 (30 Studies)**	**CRO2 (11 Studies)**
Number of Severe Issues Preventing Review of Studies (Total)	**50**	**23**
Incorrect Date/DY, Timing	19	4
Exclusion flags missed	2	
Data Issues - Missing Domains, Duplicate/Excess, Modelling	14	4
Incorrect Trial Design	3	7
Incorrect Terminology Standardization	1	
Incorrect Observation, Reason Not Done, Missing Supp-CALCN	10	7
Define Issues, Incorrect	1	1
Analysis Limiting Issues (Total)	**482**	**200**
Incorrect Date/DY, Timing	26	13
Exclusion flags missed	6	9
Data Issues - Missing Domains, Duplicate/Excess, Modelling	207	78
Incorrect Trial Design	110	35
Incorrect Terminology Standardization	13	2
Incorrect Observation, Reason Not Done, Missing Supp-CALCN	5	4
Define Issues, Incorrect	25	25
Issues with Study Report, nSDRG,	90	34

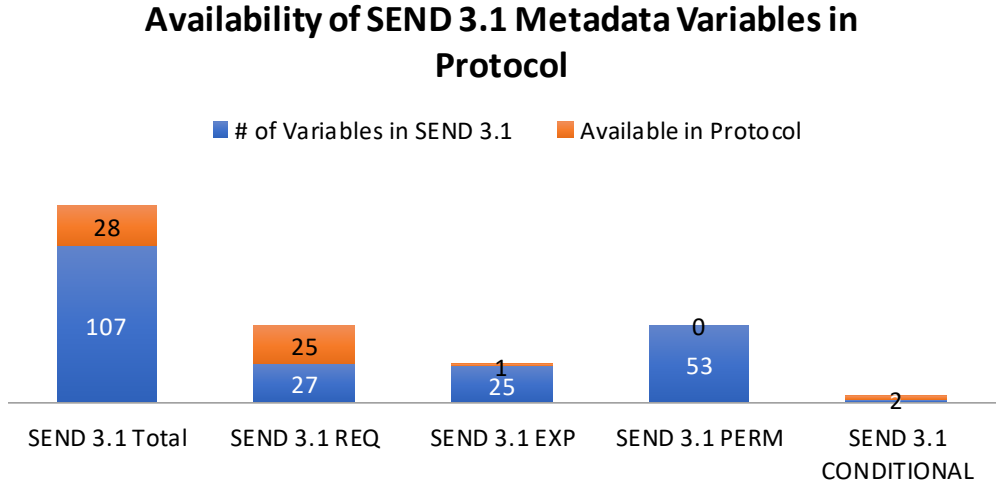
Common problems with how SEND is generated

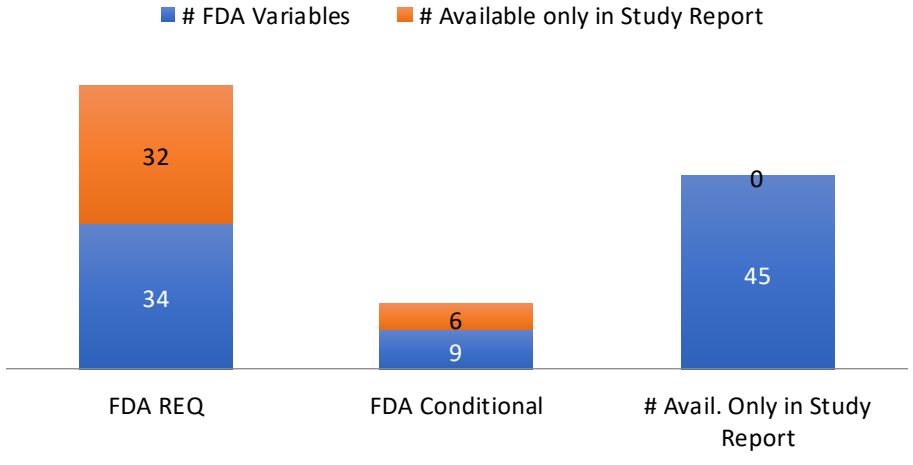
- SEND is generated according to the SEND IG
  - SEND is not GLP
  - SEND may not accommodate domains or data types recorded on-study
- SEND requires the use of data contained in the study report
  - Raw, captured LIMS data = SEND Like, not SEND
    - Subject or individual data point exclusions (by statistics)
    - Specimen conditions or comments made upon analysis
    - Nominal day labels, used for grouping
- SEND is generated by a SEND team proficient in SEND IG standards but rarely exposed to the study
  - LIMS adaptors generate individual subject data in SEND-like format
    - IG and CT chosen by SEND producer may not be the IG & CT needed for submission
    - Premature conversion to standardized terms of CT lose their original observations and result modifiers
    - LIMS do not support every data type collected (PK, ECG, CV data)
    - Combining multiple systems result in discrepancies, i.e. subject ID 001 vs ABC-001

Metadata Variables Defined in SENDIG and Where to Source Them

SEND IG has 98 References to Protocol; but ignores Study Report (3 examples)







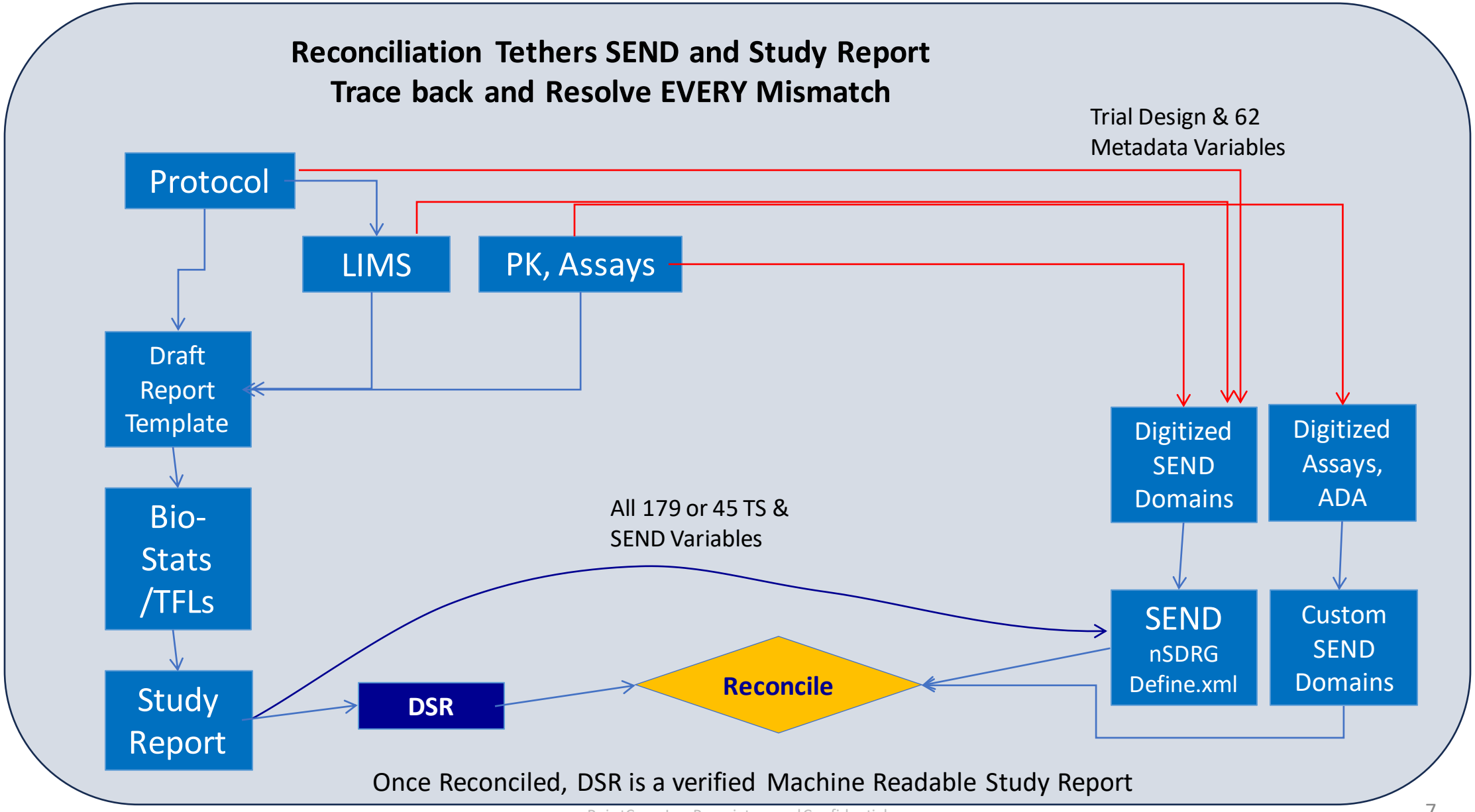
Variables available only in Study Report

- Metadata that is available only after the Study is completed and analyzed for the Report – e.g. NOMDY, Exclusion Flags, --CALCN
- Metadata about what happened during the study execution – Unscheduled Flags, Reason Not Done, Specimen Conditions, Study Dates (actual)

Recommendations

- SEND IG should call out references to Study Report
- Use Protocol During Interim Monitoring of On-Going Studies (SEND is not relevant)
- Use Study Report for SEND for Completed Studies!

Reconciling SEND and Study Report to Ensure SEND is as Good as Study Report



FDA Technical Rejection Criteria (from eCTD)

Currently implements 3 rejection criteria (1734, 1735, 1736)

1734 – Missing ts.xpt

Source: https://www.fda.gov/media/169455/download

Reasons:

1. ts.xpt is actually missing
2. STUDYID mismatch between dataset and ts.xpt
3. Study Start Date is missing

Solutions:

1. Provide a simplified ts.xpt
2. Confirm that your STUDYID is matched between dataset and ts.xpt; there are additional parameters included in ts.xpt for alternative study identifiers
3. A “full” ts.xpt requires a study start date; even if this is unknown, a record should be present with the null flavor value “NA”

1736 – Missing define.xml/dm.xpt

Reasons:

1. Define.xml is missing
2. dm.xpt is missing

Solutions:

For each study submitted to module 4, the requirements to pass 1736 require a define.xml (data definition file) and a dm.xpt (demographics file) to be submitted, along with the ts.xpt mentioned in rule 1734.

For a complete submission…

To have a module 4 study be accepted into the gateway, each study submitted needs:

1. ts.xpt
2. dm.xpt
3. Define.xml

Each file has their own record and study tagging requirements to be evaluated.

Errors in Conformance

Errors found by eDataValidator, checked against FDA Validator Rules, CDISC SEND Conformance Rules, CDISC Define Conformance Rules, FDA Technical Rejection Criteria, and more

Trial Design modeling

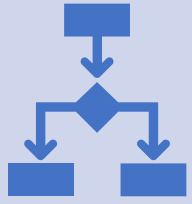
Trial design models are incorrect per the trial design in the report

- Number of elements/arms/sets
- Set-up with recovery element
- Presence of TK groups in TX domain
- Separate male/female dose groups
- Incorrect durations
- No differentiation in arms/sets

A proper trial design is the basis for analyzing a study beyond just the recorded values. Without it, you would compare:

- Dosed values with non-dosed values
- Different dose levels
- Last dose levels after different repetitions of dosing

Missing or multiple parameters



Depending on the study, some parameters can have multiple (“duplicate”) records or be missing required parameters.

Validation systems, as they are programmed, will flag these as for the duplicate records without looking at the true value which we evaluate scientifically.

For each value that appears as a duplicate, the record should be evaluated for:

- Is it necessary for inclusion? Does it add value?
- Does it repeat any information already present elsewhere?
- Can it be described using the define file or nsdrg?

Usually, duplications can be described using the nsdrg to “explain away” the error or warning that is produced.

When parameters are missing, the solution is to either appropriately populate the variable, or explain why it is “blank” in the nsdrg, i.e. no baseline values were recorded.

Observations occurring after removal

It is common that post-mortem records occur after the time of removal, specifically, date/time of organ measurement.

For other post-mortem examinations, such as microscopic examination, the date and time should be the exact date and time of the subject's disposition.

For other records in the dataset, such as food and water consumption, the record should end at or before the time of removal.

The absolute time of removal should be populated in ds.xpt.

The absolute date of removal, with or without the time, should be populated in dm.xpt.

Any observations that occur after the exact time/date of removal will produce an error.

Errors in Traceability

Errors found during reconciliation (visual and tabular) between the dataset and the digitized study report using Xbiom™ technology

Traceability – Reconciliation against Digital Study Report

The nSDRG and define files are the best places to describe anything that might be triggering an error or warning, or to further explain how the dataset was populated.

The nSDRG is a supplemental document specifically provided with the purpose to aide in review of the study. *When in doubt, nSDRG.*

Common missing domains:

- DD: Commonly misunderstood; when a subject's death is premature and a cause of death can be found, a DD domain should be presented.
- SUPP—:
  - SUPPPC: missing PCCALCN when values are used for interpretation
  - SUPPMA/SUPPMI: missing when modifiers are interpreted as part of the original finding for standardization
- POOLDEF: When food consumption is recorded as a pooled group, the pooled group should be present instead of the standardized individual result.
- "Out of scope" domains: While not part of SEND, information collected throughout the study that can be presented in SEND/tabular manner, should (per SENDIG 3.1.1, TCG).

Missing Exclusion/Baseline flags

Baseline flags are populated, typically, based on the last observation before dosing. Baseline flags are permissible, but should be populated if a baseline can be identified.

Exclusion flags are populated based on analysis after recording original data.

Proper reasons for exclusion must be included; a reason of “Excluded” does not provide an appropriate reason for analysis.

Missing Grouping Labels

The “Nominal Day” or “Nominal Label” is used to group together subjects that experienced the same parameter in the same set of days.

A protocol could dictate that in week 4, after a single dose on day 1, all subjects are to receive a set of laboratory tests.

Without the nominal label, even though subjects 001, 006, and 015 received the same treatment and treatment level, they might be evaluated differently because they received laboratory tests on day 25 or 26, depending on their grouping.

Incorrect or missing controlled terminology

1

Controlled terminology exists to make analysis easier – without it, there would not be one consistent way to record and analyze data, specifically, to reduce ambiguity.

2

When controlled terminology is not followed when a codelist exists, the study cannot be compared against the historical control data, potentially impacting a submission.

3

When the controlled terminology is incorrect, either via auto-mapping or human error, the potential exists for the incorrect conclusions to be drawn about an investigated treatment.

Thank you

Contact us at: ask@pointcross.com

www.pointcrosslifesciences.com