Data Consistency - SEND Datasets and Study Reports: Request for Collaboration in the Dynamic SEND Landscape

Lead Authors: M. Rosentreter; W. Wang

Contributors: K. Brown; L. Eickhoff; D. Fish; M. Francomacaro; C. Roy; C. Spicer; F. Wood

ABSTRACT

- The "Data Consistency: SEND Datasets and the Study Report" Working Group was formed at PhUSE CSS 2016 to identify inconsistencies that were being discovered between nonclinical datasets and study reports and provide recommendations for action.
- Team members drew from experience, as well as scenarios that surfaced through the Nonclinical FDA Fit-for-Use Pilot, to draft a list of discrepancies which was then assessed and categorized based on the item's potential impact to study data interpretation. Conclusions drawn from the assessment became the basis for recommendations.
- As work progressed, the team realized that identification of new inconsistencies and our evolving understanding of regulatory needs would quickly make the paper obsolete. With this in mind, the paper is being transformed into a live wiki page where the community is invited to contribute new scenarios and possible solutions. This approach may be applicable to other nonclinical projects. Our poster provides a look into the process, progress, and future of the team's work.

BACKGROUND

SEND datasets are intended to reflect the study design and individual data listings of a nonclinical report. However, differences between the study report and SEND datasets can occur. These factors are generally due to differences in the processes and systems used to collect and report raw data and to generate SEND datasets. Many companies, CROs, and vendors are creating parallel paths for the generation of SEND datasets and data that go into the study reports. As a result, there are and will continue to be inconsistencies between the study reports and the SEND datasets. The effort required to understand, correct, and eliminate differences is still under evaluation and may be significant. Feedback received from the FDA reviewers based on actual SEND submissions may improve our understanding, and impact the initial recommendations provided here. During PhUSE discussions, it became apparent that best practices were needed to help SEND implementers decide what methods should be used to address these differences.

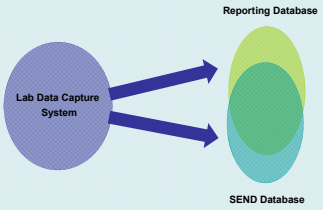


Figure 1 – Parallel Paths For Data Generation

METHODS

SEND Dataset and study report discrepancies were categorized by impact. Low-impact discrepancies are expected to have minimal effect on study data interpretation and conclusions. High-impact discrepancies could lead to misinterpretation of conclusions drawn from reviews of SEND datasets and those presented in study reports and/or data collection process changes by sponsors and CROs. Based on this assessment, this team has identified recommendations described in the next section.

EXAMPLES OF DATA DISCREPENCIES*

Table 1. Low-Impact Discrepancies – Explain discrepancy in the Nonclinical Study Data Reviewer's Guide (nSDRG)

Scenario	What Will Be in SEND Dataset	What Will Be in Study Report
Some pretest data may be present in SEND datasets but not in study report.	Pretest data may be present since some systems are unable to filter out these data.	Pretest data may or may not be present, or only a subset is present.
Values for body weight gain are represented differently between SEND dataset and study report.	Values will be reported interval to interval, but intervals may vary.	Values could be reported cumulatively or in another customized way. Report may include additional cumulative gains (e.g., from start to termination or for recovery phase). Report may contain no body-weight gain data.
CLSTRESC is less detailed than the Study Report**	SENDIG has no single (standard) representation for CL data.	Report format may vary by testing facility and/or data collection/reporting system.

***From the Technical Conformance Guide: Clinical Observations (CL) Domain:*

Only findings should be provided in CL; ensure that events and interventions are not included. Sponsors should ensure that the standardization of findings in CLSTRESC closely adheres to the SENDIG. The information in CLTEST and CLSTRESC, along with CLLOC and CLSEV when appropriate, should be structured to permit grouping of similar findings and thus support the creation of scientifically interpretable incidence tables. Differences between the representation in CL and the presentation of Clinical Observations in the study report which impact traceability to the extent that terms or counts in incidence tables created from CL cannot be easily reconciled to those in the study report should be mentioned in the nSDRG.

Table 2. High-Impact Discrepancies - Reconcile if possible and if not, explain discrepancy in the nSDRG.

Scenario	What Will Be in SEND Dataset	What Will Be in Study Report
Modifiers for pathology data may or may not be listed in the study report but should be in the SEND dataset (MA/MI domains).	Modifiers will be listed in --SUPP and/or --STRESC, in addition to --ORRES depending on how data are collected and base processes are defined by the sponsor.	Modifiers may or may not be listed as part of base finding. May lead to differences in incidence counts. The STRESC value must be scientifically meaningful but may not match the value in the incidence tables.
Nomenclature of severity may differ between SEND dataset and study report.	Will be present as Controlled Term because it is mapped to Controlled Terminology (CT).	Severity will be listed as defined by sponsor.
Textual differences (controlled terminology) between study report and datasets.	Uses standard for Controlled Terminology.	May or may not use standard for Controlled Terminology. If the differences are impactful to data interpretation, it is recommended that they be listed out in the nSDRG.

* See the Wiki Page for the full list of discrepancies identified.

RECOMMENDATIONS

- **Low-Impact Discrepancies:** Resolution Is Not Necessary or Data Inconsistency is Acceptable
  - The discrepancy has minimal or no impact on reviewer interpretation of SEND datasets versus study reports.
  - The discrepancy should still be called out in the nSDRG.
- **High-Impact Discrepancies:** Resolution Is Necessary or Data Inconsistency Should Be Resolved
  - The discrepancy could potentially lead to different interpretations of the SEND dataset versus the report for a given study.
  - Sponsors should reconcile differences between datasets and study reports using the tool(s) or manual processes available to them to do so when possible.
  - Data reconciliation processes must be sustainable so as not to put an undue burden on industry.
  - Vendors providing the technical tools can have the most impact here by enhancing their products to help industry with this reconciliation.
  - Sponsors and CROs need to consider moving toward collecting data in SEND format wherever possible.
    - Changing lexicons/libraries to use controlled terminology instead of mapping
    - Collecting all data electronically, and setting up studies in ways that will allow consistent, automated population of trial design domains

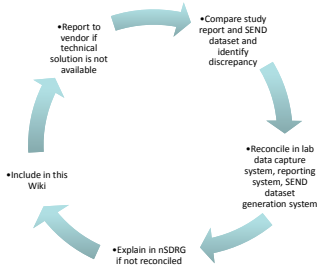


Figure 2 – The Data Discrepancy Lifecycle

OUTSTANDING ISSUES

Direct feedback from the FDA would provide a definitive direction for the industry on a few items categorized and summarized below. An open conversation with a panel of FDA reviewers would be useful in order to develop a go-forward recommendation.

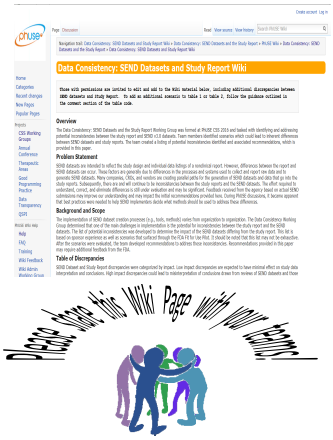
- Is referencing this difference in the nSDRG sufficient or is there an expectation that the datasets will have the same number of records as the report related to pretest/acclimation data?
- Is it acceptable to simply state in the nSDRG that the rounding differences exist?
- Are there cases where these differences are acceptable considering the effort that manual intervention would require?
- Is the inclusion of replaced animal data meaningful to a reviewer? Is there a difference as to timing (i.e., before first dose/after first dose)? Is explanation in nSDRG sufficient?
- Does the FDA expect manual correlation of RELREC data for MA/MI? For other domains?
- Why do we have to submit BG when BW data would allow FDA to analyze the data as desired? How important is it that the intervals in the SEND dataset match those in the study report exactly?

INTERACTIVE & COLLABORATIVE WIKI CONCEPT

The "Data Consistency: SEND Datasets and the Study Report" Working Group strongly encourages your participation in our Wiki page. This will:

- keep this wiki page live and interactive.
- allow additional discrepancies between SEND datasets and study reports to be added.
- provide a forum to ask your questions if issues are encountered.
- provide answers or feedback.

Note: Instructions on how to edit the Wiki page are provided.



http://www.phusewiki.org/wiki/index.php?title=Data_Consistency:_SEND_Datasets_and_Study_Report_Wiki

PROS AND CONS

PROS

- Real time collaboration of SEND experts
- Increase of overall knowledge of data discrepancies
- First step to harmonize approaches to avoid data inconsistencies
- Provides a dynamic forum to stay current with the evolution of SEND

CONS

- Contribution to the Wiki page requires some HTML skills
- The discrepancy list must be maintained to avoid confusion

CONCLUSIONS

- Differences between the SEND datasets and the study report are likely; list and explain in the nSDRG.
- The SEND landscape is changing rapidly.
  - Data discrepancies are due to the immaturity of the tools and the industry inexperience with this technology.
  - Data collection practices and reporting systems will evolve to the requirements for standardized electronic data submissions.
- The Wiki page concept allows the industry to stay abreast of developments and to share real-time information, as well as solutions.