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 from a variety of sources, to conduct research on new and existing datasets and metadata. This research was then categorized based on the final project report that was written as a result of the research. The final project report was made available to the public on the PHUSE wiki page for feedback and comments.

The working group identified inconsistencies between datasets and the intended use of the datasets. These inconsistencies were then reported on the PHUSE wiki page where they were open for discussion. The working group then reviewed the comments and then made a final report.

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.

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 reinterpretation 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 DISCREPANCIES*

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

<table>
<thead>
<tr>
<th>Scenario</th>
<th>What Will Be in SEND Dataset</th>
<th>What Will Be in Study Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sammple protocol data may not be present in SEND datasets but not in study report.</td>
<td>Pretest data may be present since systems are filtered out these data.</td>
<td>Pretest data may or may not be present, or only a subset is present.</td>
</tr>
<tr>
<td>Variables for body weight gain are represented differently between SEND dataset and study report.</td>
<td>Values will be reported interval to interval, but intervals may vary.</td>
<td>Values could be reported continually or in another customized way. Report may include additional summary data (e.g., from start to finish or at recovery phases). Report may contain no body-weight gain data.</td>
</tr>
</tbody>
</table>

**From the Technical Conference 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 CL in SEND is consistent with the standardization of findings in CL in the SEND data. The information in CL should be consistent with CL and SEND guidelines when appropriate, should be structured to permit grouping of similar findings, and should support the creation of scientifically interpretable incidence tables. Differences between the representation in CL and the representation of SEND information 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 |
| Modifications for pathology data may not be listed in the study report but should be in the SEND dataset (MA/RH domain). | Modifications will be linked to SEND data in SEND, in addition to ORBIS depending on how data are collected and base processes are defined by the sponsor. | Modifications may or may not be listed in the base findings. May lead to differences in incidence counts. The ORBIS value must be scientifically meaningful but may not match the value only due to the incidence tables. |
| Numerator of severity may differ between SEND dataset and study report. | Will be presented as Controlled Terminology (CT). | Severity will be listed as defined by sponsor. |
| Duplication of pathological findings (controlled terminologies) may have been reviewed and/or communicated to the study report. | Use standard Controlled Terminology. | May or may not use additional pertinent Controlled Terminology. If the differences are important for data interpretation, it is recommended that they be listed out in the SEND dataset. |

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

OUTSTANDING ISSUES

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

- Is referencing this difference in the nSDRG sufficient or is there an expectation that the dataset will have the same number of records as the report related to protected/acceleration data?
- Is it acceptable to simplify the nSDRG that the rounding differences exist?
- Are there cases where these discrepancies 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 SEND dataset sufficient?
- Does the FDA expect manual correlation of RCBIC data for MA/RH? For other domains?
- Why do we have to submit BQ when BQ 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?

CONCLUSIONS

- Differences between the SEND datasets and the study report are likely; list and explain in the nSDRG.
- The SEND landscape is changing rapidly.
- FDA inconsistencies 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.

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.


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.

LEAD AUTHORS: M. Rosentreter; W. Wang

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

CONSISTENCY: SEND Datasets and Study Reports: Request for Collaboration for the Dynamic SEND Landscape