Recommendations for SEND Dataset QC Best Practices

phuse.eu

Contents

Revision History

Version	Date	Summary
0.1	2021-Jun-30	Initial draft of SEND dataset QC best practices recommendation white paper
0.2	2021-Jul-26	Review by Nonclinical Topics Working Group Leads

Doc ID: WP-047 | Version: 1.1 | Working Group: Optimizing the Use of Data Standards | Date: 12-Feb-2021

1. Overview: Purpose of this Document

There are a wide variety of quality control (QC) checks for SEND datasets based on accepted practices across the pharmaceutical industry and regulatory environment. This is currently done without a commonly acknowledged scope and completeness of these checks and is based on each company's own experience and approaches with SEND datasets. For regulatory reviewers to be more confident that the quality of the SEND datasets across different stakeholders is checked consistently, this document seeks to define best practices for SEND dataset QC procedures. The aim is to promote more consistent checks across industry that will help reviewers be more confident that the data submitted accurately represent the data in the study report.

2. Scope

The scope of this document is as follows:

- Explore the variety of QC procedures and tools among different stakeholders.
- Identify a practical amount of SEND dataset QC checks (including the QC checks for the comparison of SEND datasets and the study report).
- Identify a suite of tools and procedures (e.g. visualisation, documentation) for QC.
- Develop recommendations for efficient and effective SEND QC practices with the long-term goal of establishing a foundation of commonly agreed upon QC procedures.

3. Definitions

- CDER: Center for Drug Evaluation and Research
- CDISC: Clinical Data Interchange Standards Consortium
- nSDRG: nonclinical Study Data Reviewer's Guide
- US FDA: United States Food and Drug Administration
- SEND: Standard for Exchange of Nonclinical Data
- SENDIG: Standard for Exchange of Nonclinical Data Implementation Guide
- CJUG: CDISC Japan User Group
- SAP: Statistical Analysis Plan

4. Problem Statement

The QC procedures for SEND datasets vary widely across industry. To collect information and quantify this variety, the SEND Dataset QC Best Practices project team prepared questions and reached out to the Nonclinical Topics Working Group members via a survey during late 2019. Because there is interest at regulatory agencies to gain a better understanding of what QC checks sponsors complete on studies prior to regulatory submissions, this white paper proposes standardised SEND dataset QC checks to ensure more uniform checks across industry and to assure regulatory authorities that the SEND datasets accurately represent the information in the study report.

5. Background

This project team formed to address ongoing questions from regulators as to what steps sponsors take to ensure SEND datasets match data presented in the study report. To gauge current practices, the team initiated a survey of Nonclinical group members in the PHUSE network. There were 51 individual respondents, with not everyone responding to every question. Please note, results and statements in this paper are condensed and summarised, thus might not include individual data. Most respondents were based in the United States. The results showed that most QC checks are completed manually rather than utilising validators or other automated tools. Further, there is wide variation across industry as to which variables, domains and values are checked as part of a SEND dataset QC. For example, most of the respondents said they check specific domains and variables in the SEND dataset as well as the SEND datasets against the nSDRG and the study report. In addition to the team's review, three members from the CDISC Japan User Group (CJUG) SEND team reviewed and commented on the survey results to provide more global industry inclusion. Generally, CJUG SEND team subgroup participants felt it important to share a common target quality goal across the different stakeholders, implement QC processes during the creation phase by SEND dataset creators, reduce manual QC processes and standardise the toxicology study processes across more than just the SEND datasets (to include the study protocol, SAP and study reports).

Based on the survey results, commercially available software was the most common choice for SEND dataset production, with SEND-as-a-service representing a good alternative. Generally, commercially available tools with manual checks are used to validate and verify data. Organisations that receive complete or partial SEND datasets from third parties use mostly manual checks. The most common areas that are checked are within specific domains or variables, but it is unclear which domains or variables since the question was not asked in a specific enough way in the survey. While manual checks are most likely performed when errors and warnings are represented by automated validators, the results also implied that manual checks are performed on all SEND datasets, including Trial Design domains, regardless of the validator outcome. Many sponsor companies check SEND datasets manually against summary and individual data in the study report. The biggest challenges faced by industry, surrounding QC checks, are 1) how resource intensive it is and 2) lack of clarity about what is considered a best practice for QC.

The Working Group took this survey information and CJUG feedback and broke out into smaller subgroups, with a goal to determine best practices for three groups of domain sets: Trial Design domains, CL/EX/LB, MA/MI/OM/RELREC/SUPP and CV/RE/VS/EG/BW/FW. Each group compiled their recommendations and presented them to the rest of the project team. Generally, it was determined that domains and variables are reviewed for three primary quality measures: conformance to the SEND standard, appropriate use of Controlled Terminology (CT) and accuracy with the study report. Additionally, general record counts are performed across the CL/EX/LB domains for a study to ensure the appropriate number of records is available along with the variables being populated. Companies utilise

varying strategies to address QC goals, such as outsourcing QC checks, performing 100% manual checks and using an automated tool to check all domains. The variation across project team members' companies echoed the survey results throughout this exercise.

Following the survey result analyses, the project team continued to build recommendations for SEND QC best practice by reviewing questions posed by industry stakeholders within the team, including regulatory agency staff, SEND producers, sponsors, SEND consultants and software developers. The team provided feedback and opinions on those questions to further refine recommendations. FDA CDER team participants provided general feedback on the current SEND consumer experience and areas where reliable QC practices could make a difference in the data review experience, such as 1) how participants currently populate different variables and associated clarity with current standards 2) variation in what and how categories and scales are reported (if at all) in the nSDRG or Define-XML file and 3) variability as to whether finding modifiers and expected variables are filled out as expected. A variety of other factors appear to impact the consistency of the data, including multiple test facilities performing studies for a single application, data collection in real time and timing of dataset creation, relative to associated standards release. There is also variation in the amount of information included in the SEND datasets (e.g. minimum amount required based on the standard, or everything). The nSDRG does not always say what has been omitted from the SEND datasets, which would be helpful for dataset consumers or reviewers to understand.

As there might be differing characterisations of QC check procedures, the project team aimed to pinpoint specific definitions. Again, a subgroup was formed, and the outcome of their work was presented and discussed during a full group meeting. In essence, here are general definitions of high-quality SEND data packages:

Complies with regulatory requirements and guidances:
- Any reasonable non-compliance (e.g. data warnings/errors) is fully explained in the Nonclinical Study Data Reviewer's Guide (nSDRG).

Conforms to the SENDIG:
- Conformance rules are followed, and study information is reported in the correct variable based on the definition of the variable. This also includes SEND Controlled Terminology (CT), where CT is required and expected by the SENDIG.

Correctly Represents the study:
- The datasets are consistent with the study report when the information is in the study report; they are also consistent with information recorded in other study records not included in the study report, e.g. some --DTC information; any discrepancies between the datasets and the report are described in the nSDRG.

Consistent across datasets:
- The same data is represented in the same way across domains (e.g. test article, vehicle).

Understandable:
- There is enough information in the nSDRG and Define-XML file that the consumer can understand what was received.

Fit for Use:
- Consumer requirements are met, which may be highly dependent on expected data use by the consumer.

6. Recommendation

After reviewing feedback from across industry and the US FDA, via a survey and team members' personal opinions, the team encourages industry to adopt the following best practices for SEND dataset QC processes.

- Review datasets generally for conformance/compliance to the SEND standard by, at minimum, checking against:[1]
  - FDA Business Rules[2]/Validator Rules[3]
  - CDISC SEND Conformance Rules[4]
  - Technical Rejection Criteria[5]

  The FDA Validation Rules already include verification as part of the Technical Rejection Criteria, but it is important to additionally check for the second part of eCTD 1736: presence of the Define-XML file.

- Ensure the validator tools are using the latest validation rules version and/or ensure gaps in the validation rules versions are checked manually.

- Use CT available for variables and ensure those values are consistent across the SEND dataset package and are correctly aligned with terminology in the study report.

- The FDA recommends using CT when it is available for a variable in the Study Data Technical Conformance Guide (sdTCG).

- Recommend an approach to QC similar information across domains and points that will impact reviewers' understanding of these data in the context of the study.

- SEND dataset creators should check variables that contain similar information across SEND domains to ensure consistency within the overall dataset package.
  - Example: 'Treatment Vehicle' in TS (TSPARMCD=='TRTV') v EX (variable 'EXTRTV')

- Ensure SEND datasets are consistent with information in the study report.
  - Depending on the SEND dataset creation process, if SEND datasets and listings/tables for the study report are generated with data from a joint LIMS, technical validation of the applications and their interfaces could reduce the necessary amount of QC checks for SEND datasets. If SEND datasets and listings/tables are generated from separated databases, QC checks should be more thorough. However, in many cases, consistency checks will be done manually if the study report or its listings/tables are not machine-readable.

  - Utilise tools to perform QC checks while limiting the number of manual QC checks to a small subset of variables not populated by LIMS systems.

  - Include details in the nSDRG on any gaps found in QC checks that could not be corrected, and any information that is not included in the SEND dataset package based on the study report (e.g. in Section 6.2: Differences between SEND Datasets and Study Report).

7. Disclaimer

The opinions expressed in this document are those of the authors and should not be construed to represent the opinions of PHUSE members, respective companies/organisations or regulators' views or policies. The content in this document should not be interpreted as a data standard and/or information required by regulatory authorities.

1 Nonclinical Study Data Reviewer's Guide (nSDRG): https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Nonclinical+Topics/nSDRG+v1.1.zip

2 FDA Business Rules: https://www.fda.gov/media/116935/download

3 FDA Validator Rules: https://www.fda.gov/media/103587/download

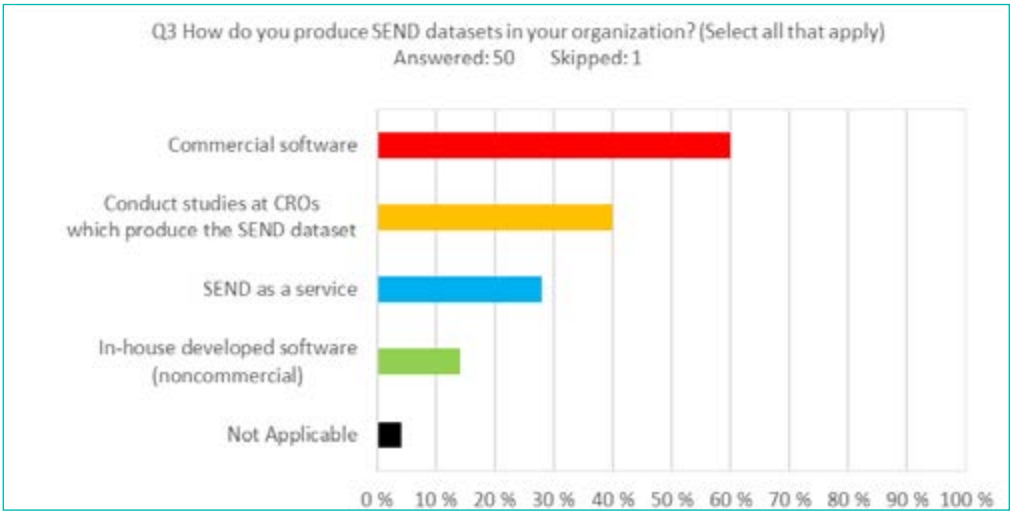
4 SENDIG Conformance Rules v3.0: https://www.cdisc.org/standards/foundational/send/send-conformance-rules-v3-0

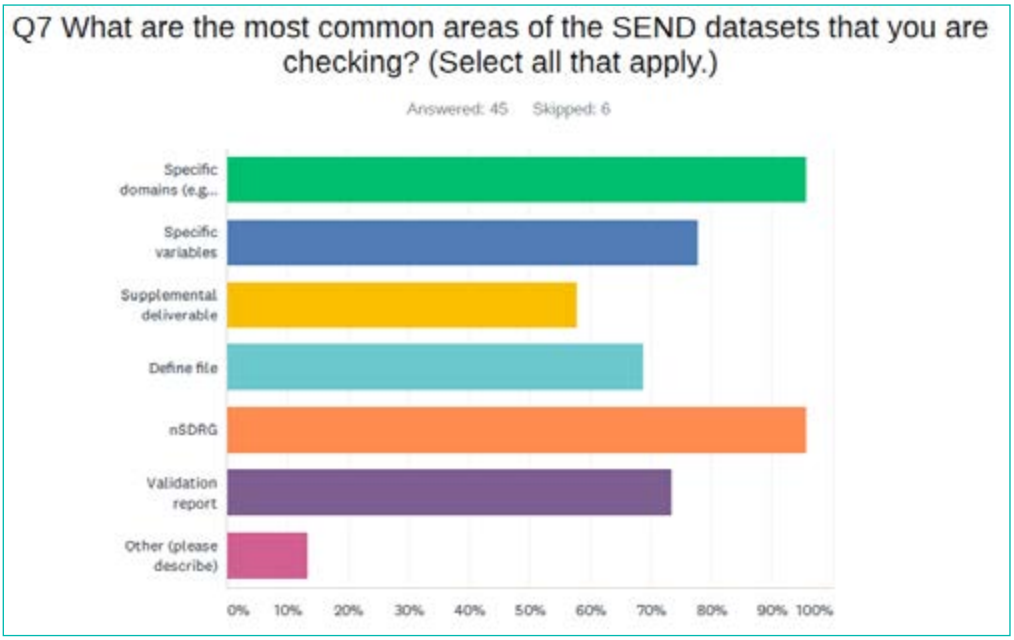
5 Technical Rejection Criteria for Study Data: https://www.fda.gov/media/100743/download

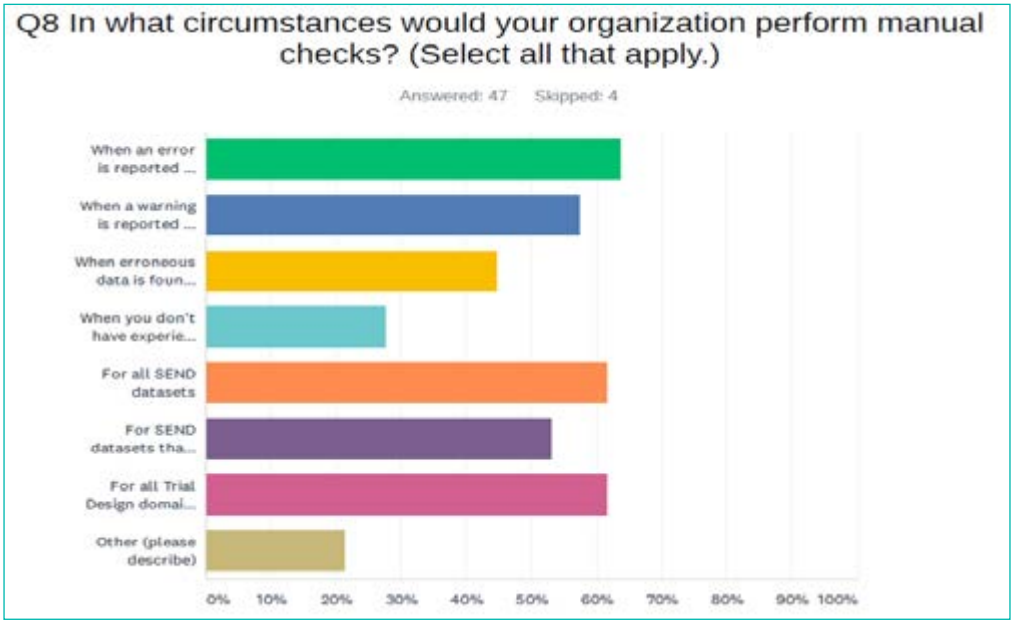
6 Study Data Technical Conformance Guide: https://www.fda.gov/media/147233/download

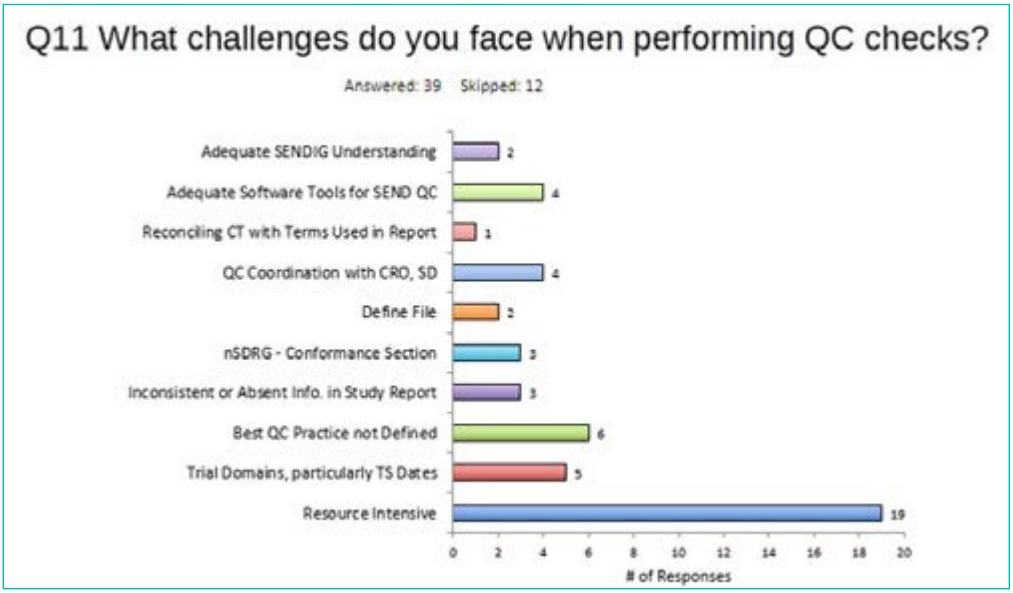
8. Appendices

Below are graphics that show some key results from the industry SEND survey.









9. Project Contact Information

- Jesse Anderson – US Food and Drug Administration
- Michael Rosentreter – Bayer AG
- Contact: workinggroups@phuse.global

10. Acknowledgements

We would like to acknowledge all participants of the PHUSE SEND Dataset QC Best Practices Working Group, who work continuously for better SEND dataset quality. Furthermore, we would like to thank all members of the CDISC Japan User Group (CJUG) SEND team who improved this white paper with their valuable feedback.