STUDY DATA
TECHNICAL CONFORMANCE GUIDE

Technical Specifications Document

This Document is incorporated by reference into the following Guidance Document(s):

Guidance for Industry *Providing Regulatory Submissions in Electronic Format – Standardized Study Data*

For questions regarding this technical specifications document, contact CDER at cder-edata@fda.hhs.gov or CBER at cber-edata@fda.hhs.gov

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

September 2021

STUDY DATA
TECHNICAL CONFORMANCE GUIDE

September 2021

Revision History		
Date	**Version**	**Summary of Revisions**
January 2014	1.0	Initial Version
December 2014	2.0	Revisions based on the public comment period (February 2014 – May 2014); and CDER/CBER internal review May 2014 – December 2014
March 2015	2.1	Revisions based on comments received to version 2.0. Updates to Sections 2.2 Study Data Reviewer's Guide (SDRG) SDRG, 2.3 Analysis Data Reviewer's (ADRG), 3.3.1 SAS Transport Format, 3.3.2. Dataset Size and a revision of Section 4.1.4.5 Data Definition File
June 2015	2.2	Revisions based on comments received to version 2.1. Updates on Table of Contents; Sections 4.1, 4.1.1.2, 4.1.2.3. Updates to Trial Design. Added Exposure as Collected (EC Domain) and Death Details (DD) subsections. Updates to 4.1.2.2, 4.1.2.3, 4.1.2.4, 4.1.2.5, 4.1.2.6, 4.1.2.8, 4.1.2.9.1, 4.1.2.9.2, 4.1.4 (header and all sub-headers updated to specify which standards apply), 4.1.4.5, and 4.1.4.6. Added 5.1 subsection; 6.7, 6.7.1, 6.7.1.1. Updates on Section 7, 8.2.2 and Glossary.
October 2015	2.3	Updates to Section 1.3, Exposure as Collected (EC Domain) and Death Details (DD Domain). Reorganization of Section 4.1.2 and corresponding updates to appropriate sub-sections. Updates to Sections 4.1.4.5 and 5.1. Added Sections 7.1 and 7.2.
March 2016	3.0	Section 2.2 (Study Data Reviewer's Guide) - Updated link for SDRG in Footnote 10 Section 3.3.2 (Dataset Size) - Increased Data Set Size Section 4.1.1.2 (SDTM General Considerations) - Updated to reflect define.xml file and SDRG reference. Section 4.1.2.2 (Analysis Data Model - General Considerations) - Updated to reflect define.xml file and SDRG reference. Section 4.1.3.2 (Standard for Exchange of Nonclinical Data - General Considerations) - Updated to reflect define.xml file and SDRG reference. Section 4.1.4.5 (Data Definition Files for SDTM, SEND, and AdaM) - Updated to reflect define.xml version 2.0 and data definition specification details Section 5.1 (Therapeutic Area Standards – General) - Updated to reflect more detailed information related to Therapeutic Area Standards Section 5.2 (Supported Therapeutic Area Standards) - Added information related to acceptance testing on the standard Section 5.2.1 (Chronic Hepatitis C) - Added Section for this information. Section 5.2.2 (Dyslipidemia) - Added Section for this information. Section 6.1.2.1 (Use of the specific controlled term "OTHER") - Added information related to controlled terminology and the mapping to "Other" Section 8.3.1 (Study Data Traceability Overview) - Update to Study Data Traceability flow diagram reference.
July 2016	3.1	Section 2.1 (Study Data Standardization Plan) Updated to reflect acronym SDSP (Study Data Standardization Plan) and added footnote 10. Section 4.1.1.3 (SDTM Domain Specifications) – Updated Trial Design Model (TDM) Section 4.1.3.3 (SEND Domain Specification) – Added Trial Design (TD) Section 5.2.3 (Diabetes) - Added Section for this information. Section 5.2.4 (QT Studies) - Added Section for this information. Section 5.2.5 (Tuberculosis) - Added Section for this information. Section 8.2.1.1 (Conformance validation) - Created Section Header and expanded information. Section 8.2.1.2 (Quality checks) – Created Section Header and updated to reflect study data standard. Section 8.2.2 (Support on Data Validation Rules) - Expanded information. Section 3.2 (Portable Document Format) & Glossary – Updated International Council for Harmonisation (ICH) name

Revision History		
Date	**Version**	**Summary of Revisions**
October 2016	3.2	Section 2.2.1 (SDRG for Clinical Data) – Added naming convention Section 2.2.2 (SDRG for Nonclinical Data) -Added naming convention Section 2.3 (Analysis Data Reviewer's Guide) – Provided additional information Section 3.3.3 (Dataset Column Length) – Expanded Information Section 4.1.1.2 (SDTM General Considerations) – Expanded Adjudication Data Section 4.1.2.10 (Software Programs) – Added more detail related to software programs Section 4.1.3.2 (General Considerations) – Added VISITDY variable information Section 4.1.3.3 (SEND Domain Specification) – Added Clinical Observations (CL) Domain and Pharmacokinetic Concentrations (PC) Domain. Expanded Trial Arms and Trial Sets information. Section 5.1 (General) – Expanded Information Section 5.2 (Supported Therapeutic Areas) – Expanded Information Section 7.1 (ECTD File Directory Structure) – Referenced the Guidance to Industry Providing Regulatory Submissions in Electronic Format: Certain Human Pharmaceutical Product Applications and Related Submissions Using the Electronic Common Technical Document Specifications and added footnote Section 7.2 (ECTD Sample Submission) – Change header to align with detailed information. Section 8.2.1 (Types of Data Validation Rules) – Expanded Information Section 8.2.1.1 (Conformance validation) – Expanded Information Section 8.2.1.2 (FDA Business Rules) – Added new Section Section 8.2.2 (Support on Data Validation Rules) – updated to reflect conformance rules Section 8.3.1 (Overview (Study Data Traceability) – added relate counts information
November 2016	3.2.1	Section 8.2.2 (Support on Data Validation Rules) – Footnote 50 Added reference to the Standards Webpage. Section 4.1.3.3 (SEND Domain Specification) – Fixed Typo. Global (Updated naming convention for clinical Study Data Reviewer's Guide ("csdrg.pdf") and the non-clinical Study Data Reviewer's Guide ("nsdrg.pdf") to reflect lower case instead of upper case. eCTD requires lower case file names
March 2017	3.3	Section 1.1 (Background) – Updated tense Section 1.4 (Organization and Summary of the Guide) – Clarification on terminology section Section 2.3 (Analysis Data Reviewer's Guide) – Clarification on ADRG Section 4.1.1.3 (SDTM Domain Specifications) – Clarification on DS Domain Section 4.1.2.2 (General Considerations) – Added prior text for desired analysis views for reviews Section 4.1.3.2 (General Considerations) – Clarification on VISITDY for MA, MI, OM in the DS Domain Section 4.1.3.3 (SEND Domain Specification) – Clarification on SUPPQUAL and MI Domain. Expanded PC Domain, Custom Domain, and Trial Design Model information. Added footnote for SENDIG Section 4.1.4.1 (Variables in SDTM and SEND: Required, Expected, and Permissible) – Clarification on SEND datasets and Subject Visits in SDTM. Section 5.1 (General) – Updated and clarified text Section 5.2 (Supported Therapeutic Areas) – Updated and clarified text, added TA section Section 8 (Study Validation and Traceability) –Updated and clarified text Section 8.3.1 (Overview) – Expanded information on traceability Section 8.3.2 (Legacy Study Data Conversion to Standardized Study Data) – Clarification on legacy data conversions Glossary – Additions

Revision History		
Date	**Version**	**Summary of Revisions**
October 2017	4.0	Section 1.5 (Relationship to Other Documents) – Updated references Section 2.1 (Study Data Standardization Plan) – Clarification on SDSP and added footnotes Section 2.2 (Study Data Reviewer's Guides) – Clarification on Reviewer Guides Section 4.1 (Clinical Data Interchange Standards Consortium) – Clarification on terms SDTM, ADaM, and SEND Section 4.1.1.2 (SDTM General Considerations) – Updated and clarified text Section 4.1.1.3 (SDTM Domain Specifications) – Added QS Domain (Questionnaires) Section 4.1.2.4 (Subject-Level Analysis Data) – Updated and clarified text on baseline characteristics Section 4.1.2.10 (Software Programs) – Updated and clarified text Section 4.1.3.1 (Definition) – Updated and clarified text Section 4.1.3.2 (General Considerations) – Clarification on variable usage Section 4.1.3.3 (SEND Domain Specification) – Clarification and added text Section 4.1.4.1 (Variables in SDTM and SEND: Required, Expected, and Permissible) – Added text Section 4.1.4.6 (Annotated Case Report Form (aCRF) for SDTM) – Updated and clarified text. The recommendation to use the SDTM Metadata Submission Guidelines was removed pending further FDA review. Section 5.2 (Supported Therapeutic Areas) –Added TA sections Section 6.3.1.1 (General Considerations) – Updated and clarified text Section 6.7.1.1 (General Considerations) – Added clarification text Section 8.3.2 (Legacy Study Data Conversion to Standardized Study Data) – Added clarification text Section 8.3.2.2 (Legacy Data Conversion Plan and Report) – Added clarification text
March 2018	4.1	Section 2.1 (Study Data Standardization Plan) – Clarified text Section 2.2 (Study Data Reviewer's Guides) – Updated footnote text Section 2.3 (Analysis Data Reviewer's Guide) – Clarified text Section 3.3.1 (SAS Transport Format) – Updated text Section 3.3.6 (Variable and Dataset Names) – Updated text Section 4.1.1.2 (SDTM General Considerations) – Updated and clarified text Section 4.1.1.3 (SDTM Domain Specifications) – Updated and clarified text Section 4.1.2.2 (General Considerations) – Clarified text Section 4.1.2.10 (Software Programs) – Updated and clarified text Section 4.1.3.3 (SEND Domain Specification) – Clarified text Section 4.1.4.1 (Variables in SDTM and SEND: Required, Expected, and Permissible) – Updated and clarified text Section 5.2 (Supported Therapeutic Areas) –Added TA sections Section 6.1.2 (Use of Controlled Terminologies) – Updated and clarified text Section 6.4.2.1 (General Considerations) – Clarified text Section 6.6.1.1 (General Considerations) – Updated text
October 2018	4.2	Section 3.3.5 (Special Characters: Variables and Datasets) – Added clarification text Section 4.1.1.3 (SDTM Domain Specifications) – Additional text and table under DM and Trial Design Model sections, added DV Domain section Section 4.1.2.10 (Software Programs) – Clarified text Section 4.1.3.2 (General Considerations) – Updated text Section 4.1.3.3 (SEND Domain Specification) – Added Lab Test Results, Body Weight, and Comments domains, updated Pharmacokinetics Concentrations Domain and Trial Design Model sections Section 4.1.4.1 (Variables in SDTM and SEND: Required, Expected, and Permissible) – Updated text Section 5.2 (Supported Therapeutic Areas) –Added TA sections Section 6.6.1.1 (General Considerations) – Updated text Section 7.1 (eCTD Specifications) – Updated text Section 8.2.2 (Support on Data Validation Rules) – Updated text Appendix B, C, D, E, F, G - Added
December 2018	4.2.1	Section 4.1.1.3 (SDTM Domain Specifications) – Updated text Section 4.1.3.3 (SEND Domain Specification) – Updated text Appendix F, G - Removed

Revision History		
Date	**Version**	**Summary of Revisions**
March 2019	4.3	Section 2.3 (Analysis Data Reviewer's Guide) – Clarified text Section 4.1.1.3 (SDTM Domain Specifications) – Updated text Section 4.1.3.2 (General Considerations) – Clarified text Section 4.1.3.3 (SEND Domain Specification) – Added, updated, and clarified text Section 4.1.4.5 (Data Definition Files for SDTM, SEND, and ADaM) – Clarified text Section 5.2 (Supported Therapeutic Areas) – Added TA sections Section 6.4.1.1 (General Considerations) – Updated text Section 6.4.2.1 (General Considerations) – Updated text Section 6.5.1 (Medication - Reference Terminology) – Updated text Section 6.5.1.1 (General Considerations) – Updated text Section Appendix B – Updated text Section Appendix C – Added, updated, and clarified text
October 2019	4.4	Section 2.1 (Study Data Standardization Plan) – Added clarifying text Section 2.2 (Study Data Reviewer's Guide) – Added clarifying text Section 2.3 (Analysis Data Reviewer's Guide) – Added clarifying text Section 4.1.1.2 (SDTM General Considerations) – Added clarifying text Section 4.1.1.3 (SDTM Domain Specifications) – Added text Section 4.1.3.3 (SEND Domain Specification) – Added text Section 4.1.2.6 (Key Efficacy and Safety Data) – Updated title and text Section 5.2 (Supported Therapeutic Areas) – Re-ordered by inclusion date, added TA section Section 6.1 (Use of Controlled Terminologies) – Added clarifying text Section 6.7.1.1 (General Considerations) – Updated section Section 7.1 (eCTD Specifications) – Added text Section 8.2.2 (Support on Data Validation Rules) – Added text
March 2020	4.5	Section 4.1.3.3 (SEND Domain Specification) – Updated and added clarifying text Section 5.2 (Supported Therapeutic Areas) – Added TA section
July 2020	4.5.1	Section 5.2.18 (Vaccines Therapeutic Area User Guide v1.1) – Added text Appendix D – Added text
November 2020	4.6	Section 1.2 (Purpose) – Added text Section 4.1.1.2 (SDTM General Considerations) – Text clarified Section 4.1.3.2 (General Considerations) – Updated and added clarifying text Section 4.1.4.1 (Variables in SDTM and SEND: Required, Expected, and Permissible) – Text clarified Section 5.2 (Supported Therapeutic Areas) – Added TA section Section 6.1.3 (Maintenance of Controlled Terminologies) – Added clarifying text Appendix C – Added text and notes Appendix D – Updated text
March 2021	4.7	Section 4.1.1.3 (SDTM Domain Specifications) – Provided clarification on the use of AE domains and the use of laboratory data units. Section 4.1.3.2 (General Considerations) – Specified when FOCID should be utilized Section 4.1.3.3 (SEND Domain Specification) – Provided clarification on the use of BLQ vs BQL, and the use of custom domains Section 7.1 (eCTD Specifications) – Provided clarification on the submission of toxicity study data for nonclinical Weight of Evidence documents Appendix C: Removed TS code for Interim Study Flag, updated the description for TS code Planned Dose Frequency, and added new TS code titled Study Report Status
June 2021	4.7.1	Section 4.1.4.7 (Requirements During Specific Public Health Emergencies Declared by the Secretary of HHS) – Provided clarification about when electronic standardized study data are required as part of a submission during a declared public health emergency
August 2021	4.7.2	Section 8.2.2 (Support on Data Validation Rules) – Provided clarification regarding the submission of clinical study data with the use of simplified ts.xpt and technical rejection criteria
	4.7.2.1	Section 4.1.3.2 (General Considerations) – Correction made to indicate that SEND datasets will not be required for CBER submissions until March 15, 2023
September 2021	4.8	Section 4.1.3.4 (Scope of SEND), Section 4.1.3.4.1 (Scope of SEND for SENDIGs v3.0 and v3.1), Section 4.1.3.4.2 (Scope of SEND for SENDIG - Animal Rule v1.0) – Provided clarification on the expectation of SEND for studies listed in the referenced SENDIGs.

Table of Contents

STUDY DATA TECHNICAL CONFORMANCE GUIDE

This technical specifications document represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, send an email to cder-edata@fda.hhs.gov or cber-edata@fda.hhs.gov.

1. Introduction

1.1 Background

This Study Data Technical Conformance Guide (Guide) provides specifications, recommendations, and general considerations on how to submit standardized study data using FDA-supported[1] data standards located in the **FDA Data Standards Catalog** (Catalog).[2] The Guide supplements the guidance for industry *Providing Regulatory Submissions in Electronic Format — Standardized Study Data* (eStudy Data). The eStudy Data guidance implements the electronic submission requirements of section 745A(a) of the Food, Drug, & Cosmetic (FD&C) Act with respect to standardized study data contained in certain investigational new drug applications (INDs), new drug applications (NDAs); abbreviated new drug applications (ANDAs); and certain biologics license applications (BLAs) that are submitted to the Center for Drug Evaluation and Research (CDER) or the Center for Biologics Evaluation and Research (CBER).[3]

1.2 Purpose

This Guide provides technical recommendations to sponsors[4] for the submission of animal and human study data and related information in a standardized electronic format in INDs, NDAs, ANDAs, and BLAs[5]. The Guide is intended to complement and promote interactions between sponsors and FDA review divisions. However, it is not intended to replace the need for sponsors to communicate directly with review divisions regarding implementation approaches or issues relating to data standards.

[1] For the purposes of this document, "supported" means the receiving Center has established processes and technology to support receiving, processing, reviewing, and archiving files in the specified file format.
[2] Available at http://www.fda.gov/eStudyResources.
[3] See *Providing Regulatory Submissions in Electronic Format — Standardized Study Data* (section II.A) available at http://www.fda.gov/eStudyResources.
[4] For the purposes of this document, the term "sponsor" refers to both "sponsors" and "applicants" who are submitting study data to the Agency.
[5] Docket No. FDA-2018-D-1216

Because of the inherent variability across studies and applications, it is difficult to identify all data needed by a review division prior to a scientific regulatory review. We recommend that as early as the pre-IND meeting, sponsors should use the established regulatory process to discuss with the review division the key data necessary to support a submission, the data elements that should be included in each dataset, and the organization of the data within the datasets.

Some data standards may not require the use of all defined data elements to be collected in any given study. For example, the Study Data Tabulation Model Implementation Guide (SDTMIG)[6] classifies variables as required, expected, or permissible. *What* data are collected and submitted is a decision that should be made based on scientific reasons, regulation requirements, and discussions with the review division. However, all study-specific data necessary to evaluate the safety and efficacy of the medical product should be submitted in conformance with the standards currently supported by FDA and listed in the Catalog.

This document applies to submissions to CDER and CBER, however some review offices and multi-disciplinary review teams may have specific technical guidance which provides additional details on preparing and submitting information that may differ from this document. In those cases the specific technical guidance should be followed instead of the information contained herein. If there is a question regarding a specific submission or a particular data standard implementation, the sponsor should contact the review division for specific submission questions or the appropriate contact for data standards issues (cder-edata@fda.hhs.gov or cber-edata@fda.hhs.gov).

This Guide supersedes all previous Study Data Specifications documents (Versions 1.0 - 2.0) and CDER Study Data Common Issues Documents (Versions 1.0 -1.1).

1.3 Document Revision and Control

FDA intends to post updated versions of the Guide to the **Study Data Standards Resources Web page** (Standards Web page)[7]. The plan is to publish updated versions in March and October of each calendar year. However, the Guide will be posted sooner if important issues arise. The revision history page of the Guide provides information on the changes made to previous versions.

1.4 Organization and Summary of the Guide

This document is organized as follows:

Section 1: **Introduction** – provides information on regulatory policy and guidance background, purpose, and document control.

[6] See http://www.cdisc.org.
[7] The Standards Web page can be accessed at http://www.fda.gov/eStudyResources.

Section 2: **Planning and Providing Standardized Study Data** – recommends and provides details on preparing an overall study data standardization plan, a study data reviewer's guide and an analysis data reviewer's guide.

Section 3: **Exchange Format: Electronic Submissions** – presents the specifications, considerations, and recommendations for the file formats currently supported by FDA.

Section 4: **Study Data Submission Format: Clinical and Nonclinical** – presents general considerations and specifications for sponsors using, for example, the following standards for the submission of study data: Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM), and Standard for Exchange of Nonclinical Data (SEND).

Section 5: **Therapeutic Area Standards** – presents supplemental considerations and specific recommendations when sponsors submit study data using therapeutic area extensions of FDA-supported standards.

Section 6: **Terminology** – presents general considerations and specific recommendations when using controlled terminologies/vocabularies for clinical trial data or nonclinical study data.

Section 7: **Electronic Submission Format** – provides specifications and recommendations on submitting study data using the electronic Common Technical Document (eCTD) format.

Section 8: **Study Data Validation and Traceability** – provides general recommendations on conformance to standards, data validation rules, data traceability expectations, and legacy data conversion.

1.5 Relationship to Other Documents

This Guide integrates and updates information discussed previously in the Study Data Specifications and the CDER Common Data Standards Issues documents. As noted above, this Guide supersedes all previous Study Data Specifications documents (Versions 1.0 - 2.0) and CDER Study Data Common Issues Documents (Versions 1.0 -1.1). The examples of issues and concerns discussed in the Guide are intended as examples only of common issues, and not an inclusive list of all possible issues.

This Guide is incorporated by reference into the Guidance to Industry *Providing Regulatory Submissions in Electronic Format: Standardized Study Data*. In addition, sponsors should reference the following:

- Study Data Standards Resources Web page (See section 1.3)
- FDA Data Standards Catalog (See section 1.1)
- FDA Portable Document Format Specifications (See section 3.2)

- Specifications for File Format Types Using eCTD Specifications[8]
- Guidance to Industry Providing Regulatory Submissions in Electronic Format: *Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act*[9]
- Guidance to Industry *Providing Regulatory Submissions in Electronic Format: Certain Human Pharmaceutical Product Applications and Related Submissions Using the Electronic Common Technical Document Specifications*[10]

2. Planning and Providing Standardized Study Data

2.1 Study Data Standardization Plan

For clinical and nonclinical studies, sponsors should include a plan (e.g., during the early stages of product development conducted under the IND) describing the submission of standardized study data to FDA. The Study Data Standardization Plan (SDSP) assists FDA in identifying potential data standardization issues early in the development program. Sponsors may also initiate discussions at the pre-IND stage. For INDs, NDAs, and BLAs, the SDSP should be located in eCTD sections 1.13.9 General Investigational Plan or 1.20 General investigational plan for initial IND. Although a specific template is not specified, an example SDSP is available.[11]

The SDSP should be updated in subsequent communications with FDA as the development program expands and additional studies are planned. Updates to the SDSP should not be communicated each time a study is started. The cover letter accompanying a study data submission should describe the extent to which the latest version of the SDSP was executed. An SDSP should be provided with pre-NDA and pre-BLA meetings.

In addition, for clinical studies that will be submitted to CBER, the SDSP appendix should be provided to the review office no later than the End-of-Phase 2 (EOP2) meeting. The CBER SDSP appendix should include tables of proposed SDTM domain/variable usage, supplemental domain usage and proposed analysis.

2.2 Study Data Reviewer's Guides

The preparation of the relevant Reviewer Guides (RG)[12] is recommended as an integral part of a standards-compliant study data submission. An RG should describe any special

[8] See https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formssubmissionrequirements/electronicsubmissions/ucm347471.pdf

[9] https://www.fda.gov/downloads/drugs/guidances/ucm384686.pdf

[10] www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm333969.pdf

[11] A specific template for a Study Data Standardization Plan is not specified. However, an example can be found at https://www.phuse.eu/css-deliverables The PhUSE SDSP template has been reviewed by FDA and published in the Federal Register https://www.federalregister.gov/documents/2016/11/08/2016-26913/intent-to-review-a-study-data-standardization-plan-template-notice-of-availability-establishment-of. FDA prefers but does not require its use.

[12] For the purposes of this document, the term 'Reviewer Guide' refers only to those located in the m4 or m5 eCTD folders.

considerations or directions or conformance issues that may facilitate an FDA reviewer's use of the submitted data and may help the reviewer understand the relationships between the study report and the data.

There are two study data reviewer guides (SDRG): clinical and nonclinical. The SDRG for nonclinical studies (nSDRG)[13] and SDRG for clinical studies (cSDRG)[14] should be placed with the study data in Module 4 and 5, respectively, in the eCTD.[15] The SDRG should be file-tagged as 'data-tabulation-data-definition', with a clear leaf title.

2.2.1 SDRG for Clinical Data

An SDRG for clinical data should be named cSDRG (the prefix 'c' designates 'clinical') and the document should be named 'csdrg' and provided as a PDF file upon submission (csdrg.pdf).

2.2.2 SDRG for Nonclinical Data

An SDRG for nonclinical data should be named nsdrg (the prefix 'n' designates 'nonclinical') and the document should be named 'nsdrg' and provided as a PDF file upon submission (nsdrg.pdf).

2.3 Analysis Data Reviewer's Guide

The preparation of an Analysis Data Reviewer's Guide (ADRG)[16] is recommended as an important part of a standards-compliant analysis data submission for clinical trials. The ADRG provides FDA reviewers with context for analysis datasets and terminology, received as part of a regulatory product submission, additional to what is presented within the data folder (i.e., define.xml). The ADRG also provides a summary of ADaM conformance findings. The ADRG purposefully duplicates limited information found in other submission documentation (e.g., the protocol, statistical analysis plan (SAP), clinical study report, define.xml) in order to provide FDA reviewers with a single point of orientation to the analysis datasets. It should be noted that the submission of an ADRG does not eliminate the requirement to submit a complete and informative define.xml file corresponding to the analysis datasets.

---

[13] A specific template for a Study Data Reviewer's Guide for nonclinical studies is not specified. However, an example can be found at https://www.phuse.eu/css-deliverables. The PhUSE cSDRG template has been reviewed by FDA and published in the Federal Register https://www.federalregister.gov/documents/2015/07/23/2015-18027/intent-to-review-a-study-data-reviewers-guide-template. FDA prefers but does not require its use.

[14] A specific template for a Study Data Reviewer's Guide for clinical studies is not specified. However, an example can be found at https://www.phuse.eu/css-deliverables. The PhUSE nSDRG template has been reviewed by FDA and published in the Federal Register https://www.federalregister.gov/documents/2016/03/04/2016-04791/intent-to-review-a-nonclinical-study-data-reviewers-guide-template. FDA prefers but does not require its use.

[15] The Study Data Reviewer's Guides are separate documents from an overall reviewer's guide which is placed in Module 1 of the eCTD.

[16] A specific template for an Analysis Data Reviewer's Guide is not specified. However, an example can be found at https://www.phuse.eu/css-deliverables.

- The ADRG for a clinical study should be placed with the analysis data in Module 5 of the eCTD. The ADRG should be file-tagged as 'analysis-data-definition', with a clear leaf title.
- An ADRG for clinical data should be called an ADRG and the document should be a PDF file 'adrg.pdf' upon submission.

3. Exchange Format – Electronic Submissions

3.1 Extensible Mark-up Language

Extensible Mark-up Language (XML), as defined by the World Wide Web Consortium (W3C), specifies a set of rules for encoding documents in a format that is both human-readable and machine-readable.[17,18] XML facilitates the sharing of structured data across different information systems. An XML use case is CDISC's define.xml file. All XML files should use .xml as the file extension. Although XML files can be compressed, the define.xml should not be compressed.

3.2 Portable Document Format

Portable Document Format (PDF) is an open file format used to represent documents in a manner independent of application software, hardware, and operating systems.[19] A PDF use case includes, e.g., the annotated CRF (aCRF / blankcrf), and other documents that align with the International Council for Harmonisation (ICH) M2.[20] FDA PDF specifications are located on FDA's eCTD Web site.[21] The Catalog lists the PDF version(s) that are supported by FDA. All PDF files should use .pdf as the file extension.

3.3 File Transport Format

3.3.1 SAS Transport Format

The SAS Transport Format (XPORT) Version 5 is the file format for the submission of all electronic datasets.[22] The XPORT is an open file format published by SAS Institute for the exchange of study data. Data can be translated to and from XPORT to other commonly used formats without the use of programs from SAS Institute or any specific vendor. There should be one dataset per transport file, and the dataset in the transport file should be named the same as the transport file (e.g., 'ae' and ae.xpt, 'suppae' and suppae.xpt).

XPORT files can be created by the COPY Procedure in SAS Version 5 and higher of the SAS Software. SAS Transport files processed by the SAS CPORT cannot be reviewed, processed, or archived by FDA. Sponsors can find the record layout for SAS XPORT

[17] See http://en.wikipedia.org/wiki/XML.
[18] See http://www.w3.org/XML/.
[19] Adobe Systems Incorporated, PDF Reference, sixth edition, version 1, Nov. 2006, p. 33.
[20] See http://www.ich.org/products/electronic-standards.html.
[21] Available at http://www.fda.gov/ectd
[22] See http://www.sas.com

transport files through SAS technical document TS-140.[23] All SAS XPORT transport files should use .xpt as the file extension. There should be one dataset per XPORT file, and the files should not be compressed.

3.3.2 Dataset Size

Each dataset should be provided in a single transport file. The maximum size of an individual dataset that FDA can process depends on many factors. Datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets no larger than 5 GB. Sponsors should submit these smaller datasets, in addition to the larger non-split datasets, to better support regulatory reviewers. The split datasets should be placed in a separate sub-directory labeled 'split' (See section 7.1). A clear explanation regarding how these datasets were split needs to be presented within the relevant data RG.

3.3.3 Dataset Column Length

The allotted length for each column containing character (text) data should be set to the maximum length of the variable used across all datasets in the study except for suppqual datasets. For suppqual datasets, the allotted length for each column containing character (text) data should be set to the maximum length of the variable used in the individual dataset. This will significantly reduce file sizes. For example, if USUBJID has a maximum length of 18, the USUBJID's column size should be set to 18, not 200.

3.3.4 Variable and Dataset Descriptor Length

The length of variable names, descriptive labels, and dataset labels should not exceed the maximum permissible number of characters described below.

Table 1: Maximum Length of Variables and Dataset Elements

Element	Maximum Length in Characters
Variable Name	8
Variable Descriptive Label	40
Dataset Label	40

3.3.5 Special Characters: Variables and Datasets

Variable names, as well as variable and dataset labels should include American Standard Code for Information Interchange (ASCII) text codes only. Variable values are the most broadly compatible with software and operating systems when they are restricted to ASCII text codes (printable values below 128). Use UTF-8 for extending character sets; however, the use of extended mappings is not recommended. Transcoding errors, variable length errors, and lack of software support for multi byte UTF-8 encodings can result in incorrect character display and variable value truncations. Ensure that LBSTRESC and

[23] http://support.sas.com/techsup/technote/ts140.pdf

controlled terminology extensions in LBTEST do not contain byte values 160-191 as some character mappings in that range may interfere with agency processes.

3.3.6 Variable and Dataset Names

Variable names should contain only uppercase letters, numbers, and must start with a letter. Dataset names should contain only lowercase letters, numbers, and must start with a letter. No other symbols or special characters should be included in these names. For legacy studies started on or before December 17, 2016, it is permissible to use the underscore character _ in variable names and dataset names.

3.3.7 Variable and Dataset Labels

Variable and dataset labels can include punctuation characters. However, special characters should not be provided, such as,

1. Unbalanced apostrophe, e.g., Parkinson's.
2. Unbalanced single and double quotation marks.
3. Unbalanced parentheses, braces or brackets, e.g.,‘(‘, ‘{‘and ‘[‘.
4. ‘<’ less-than sign and ‘>’ greater-than sign.

4. Study Data Submission Format – Clinical and Nonclinical

4.1 Clinical Data Interchange Standards Consortium

Clinical Data Interchange Standards Consortium **(**CDISC) is an open, multidisciplinary, neutral, nonprofit standards development organization (SDO) that has been working through consensus-based collaborative teams to develop global data standards for clinical and nonclinical research.[24]

Data format specifications for the tabulation datasets of clinical and nonclinical toxicology studies are provided by SDTM and SEND, respectively, while data format specifications for the analysis datasets of clinical studies are provided by ADaM. It should be noted that data format specifications for the analysis datasets of nonclinical toxicology studies have not been developed. As noted in section 1.1, the Catalog provides a listing of the currently supported data standards with links to reference materials. For the purposes of this Guide, the terms SDTM, ADaM, and SEND apply to versions only listed and supported by FDA in the Catalog.

Although the SDTM and SEND formats facilitate review of the data, they do not always provide the data structured in a way that supports all analyses needed for review. Analysis files are critical for FDA to understand, on a per subject basis, how the specific analyses contained in the study report have been created. Therefore, sponsors should supplement the SDTM with ADaM analysis datasets as described below.

[24] See http://www.cdisc.org.

There may be instances in which current implementation guides (e.g., SDTMIG, SENDIG) do not provide specific instruction as to how certain study data should be represented. In these instances, sponsors should discuss their proposed solution with the review division and submit supporting documentation that describes these decisions or solutions in the appropriate SDRG at the time of submission.

4.1.1 Study Data Tabulation Model

4.1.1.1 Definition

The Study Data Tabulation Model (SDTM) defines a standard structure for human clinical trials tabulation datasets.

4.1.1.2 SDTM General Considerations

It is recommended that sponsors implement the SDTM standard for representation of clinical trial tabulation data prior to the conduct of the study.

The SDTMIG should be followed unless otherwise indicated in this Guide or in the Catalog. The conformance criteria listed in the SDTMIG should not be interpreted as the sole determinant of the adequacy of submitted data. If there is uncertainty regarding implementation, the sponsor should discuss application-specific questions with the review division and general standards implementation questions with the specific center resources identified elsewhere in this Guide (See section 1.2). Each submitted SDTM dataset should have its contents described with complete metadata in the define.xml file (See section 4.1.4.5) and within the cSDRG as appropriate (See section 2.2).

Except for variables that are defined in the SDTMIG as being coded, numerically coded variables typically are not submitted as part of the SDTM datasets. Numeric values generated from validated scoring instruments or questionnaires do not represent codes, and therefore have no relevance for this issue. There may be special instances when codes are preferred, hence sponsors should refer to the review division for direction, if there are any questions.

Subject Identifier (SUBJID)
The variable SUBJID uniquely identifies each subject that participates in a study. If a single subject is screened and/or enrolled more than once in a study, then the subject's SUBJID should be different for each unique screening or enrollment. For a study with multiple screenings and/or multiple enrollments per subject, SUBJID should be included in other related domains besides DM even though it may cause validation errors. It is recommended to include a table linking each SUBJID for a single subject to that subject's USUBJID with any additional necessary explanation included in the relevant RG.

Unique Subject Identifier (USUBJID)
The variable USUBJID is an identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product[25]. Each individual

[25] CDISC, https://www.cdisc.org/standards/foundational

subject should be assigned a single unique identifier across the entire application. This is in addition to the subject ID (SUBJID) used to identify subjects in each study and its corresponding study report. An individual subject should have the exact same unique identifier across all datasets, including between SDTM and ADaM datasets. Subjects that participate in more than one study should maintain the same USUBJID across all studies. It is important to follow this convention to enable pooling of a single subject's data across studies (e.g., a randomized control trial and an extension study).

Sponsors should not add leading or trailing spaces to the USUBJID variable in any dataset. For example, applications have been previously submitted in which the USUBJID variable for each individual subject appeared to be the same across datasets; however, in certain datasets, the actual entry had leading zeros added, or zeros added elsewhere in the entry. This does not allow for machine-readable matching of individual subject data across all datasets. Improper implementation of the USUBJID variable is a common error with applications and often requires sponsors to re-submit their data.

Adjudication Data

There are no existing standards or best practices for the representation of adjudication data as part of a standard data submission. Until standards for adjudication data are developed, it is advised that sponsors discuss their proposed approach with the review division and also include details about the presence, implementation approach, and location of adjudication data in the SDRG.

Whenever adjudication data are provided, they should be clearly identified so that the reviewer can distinguish the results of adjudication from data as originally collected.

4.1.1.3 SDTM Domain Specifications

SUPPQUAL (Supplemental Qualifier)

A SUPPQUAL dataset is a special SDTM dataset that contains non-standard variables which cannot be represented in the existing SDTM domains. SUPPQUAL should be used only when key data cannot be represented in SDTM domains. In general, variables used to support key analyses should not be represented in SUPPQUAL. Discussion with the review division should occur if the sponsor intends to include important variables (e.g., that support key analyses) in SUPPQUAL datasets, and this should be reflected in the SDRG.

DM Domain (Demographics)

In the DM domain, each subject should have only one single record per study.

Screen failures, when provided, should be included as a record in DM with the ARM, ARMCD, ACTARM, and ACTARMCD field left blank. For subjects who are randomized in treatment group but not treated, the planned arm variables (ARM and ARMCD) should be populated, but actual treatment arm variables (ACTARM and ACTARMCD) should be left blank.[26]

[26] Although this convention is inconsistent with the SDTMIG, FDA recommends its use so that 'Screen Failure', 'Not assigned', and 'Not treated' are not specified as a treatment arm.

For subjects with multiple enrollments within a single study, the primary enrollment should be submitted in DM. Additional enrollments should be included in a custom domain with a similar structure to DM. Clarifying statements in the RG would be helpful.

For subjects with multiple screenings and no subsequent enrollment, include the primary screening in DM with additional screenings in a custom domain with a structure similar to DM.

For subjects with multiple screenings and subsequent enrollment, include the enrollment in DM with screenings in a custom domain with a structure similar to DM.

DS Domain (Disposition)

When there is more than one disposition event, the EPOCH or DSSCAT variable should be used to aid in distinguishing between them. This will allow identification of the EPOCH in which each event occurred or DSSCAT to differentiate if the disposition is for treatment or study. If a death of any type occurs, it should be the last record and should include its associated EPOCH. It is expected that EPOCH variable values will be determined based on the trial design and thus should be defined clearly and documented in the define.xml.

SE Domain (Subject Elements)

The Subject Elements domain should be included to aid in the association of subject data (e.g., findings, events, and interventions) with the study element in which they occurred.

AE Domain (Adverse Events (AE))

The AE domain should include all adverse events, unless otherwise specified in Technical Specification Document(s)[27] appropriate for the indication. The definition of treatment emergent adverse events should be agreed upon with the review division and specified in the protocol (e.g., any AE after first dose of investigational product administration, or any AE after first dose of investigational product administration until X days after the last dose).

The entry of a 'Y' for the serious adverse event variable, AESER, should have the assessment indicated (e.g., as a death, hospitalization, or disability/permanent damage). Frequently, sponsors omit the assessment information, even when it has been collected on the CRF. The criteria that led to the determination should be provided. This information is critical during FDA review to support the characterization of serious AEs.

Custom Domains

The SDTMIG permits the creation of custom domains if the data do not fit into an existing domain. Prior to creating a custom domain, sponsors should confirm that the data do not fit into an existing domain. If it is necessary to create custom domains, sponsors

[27] Technical Specification Document(s) can be found on the FDA Study Data Standards Resources webpage, https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources.

should follow the recommendations in the SDTMIG. In addition, sponsors should present their implementation approach in the cSDRG. To provide study data that do not fit into an existing SDTM domain or draft SDTM domain, consider creating a custom dataset aligned with the Study Data Tabulation Model (SDTM). Questions about custom domains should be addressed in pre-submission meetings and documented in the SDSP.

LB Domain (Laboratory)

The size of the LB domain dataset submitted by sponsors is often too large to process (See section 3.3.2). This issue can be addressed by splitting a large LB dataset into smaller datasets according to LBCAT and LBSCAT, using LBCAT for initial splitting. If the size is still too large, then use LBSCAT for further splitting. For example, use the dataset name lb1.xpt for chemistry, lb2.xpt for hematology, and lb3.xpt for urinalysis. Splitting the dataset in other ways (e.g., by subject or file size) makes the data less useable. Sponsors should submit these smaller files in addition to the larger non-split standard LB domain file. Sponsors should submit the split files in a separate sub-directory/split that is clearly documented in addition to the non-split standard LB domain file in the SDTM datasets directory (See section 7). FDA may require laboratory data using conventional units for reviewing submissions and labeling. Sponsors should discuss with the review divisions what laboratory data should utilize conventional units prior to submission

Trial Design Model (TDM)

Unless a simplified ts.xpt is indicated (see below), all TDM datasets should be included with each SDTM study submission to describe the planned conduct of a clinical study.

When submitting a full ts.xpt, please refer to the appendix section for a list of study parameters that should be submitted where relevant for clinical studies. Additional parameters may be included beyond those listed in the appendix. For clinical studies, study start date (SSTDTC) is the earliest date of informed consent among any subject that enrolled in the study[28].

In addition to the study parameters indicated in the appendix section, if the study data submitted follows a Therapeutic Area User Guide (TAUG) or an FDA Technical Specification[29], use the values for TSPARM/TSPARMCD and TSVAL from the table below in the TS domain. Use of these parameters in TS will allow for tracking and reporting on the submission rates of study data following a particular TAUG or technical specification. At this time, it is also helpful to include the version of the CDISC implementation guide (IG) and model used using the parameters indicated in the table below.

TSPARMCD value	**TSPARM value**	**TSVAL value**
CTAUG	CDISC Therapeutic	*Should be the exact listing in section 5.2 of the TCG for TAUGs* Ex. Chronic Hepatitis C Therapeutic Area User Guide v1.0

[28] https://www.fda.gov/media/82716/download

[29] https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm

	Area User Guide	
FDATCHSP	FDA Tech Spec	*Should be the exact listing on the study data standards resources website for technical specification documents* Ex. Vaccines Technical Specification Guidance v1.0
SDTIGVER	SDTM IG Version	*Should be the exact term listed in column F of the FDA Data Standards Catalog, Submission Data Exchange Stds tab. If multiple SDTM IG Versions are used for a study, then each version used should have a unique value for TSPARM.* Ex. 3.2
SDTMVER	SDTM Version	*Should be the exact term listed in column E of the FDA Data Standards Catalog, Submission Data Exchange Stds tab. If multiple SDTM Versions are used for a study, then each version used should have a unique value for TSPARM.* Ex. 1.4

EC Domain (Exposure as Collected)

The Exposure as Collected domain provides for protocol-specified study treatment administrations, as-collected. The EC domain may address some challenges in providing a subject's exposure to study medication.

DD Domain (Death Details)

The Death Details domain provides for supplemental data that are typically collected when a death occurs, such as the official cause of death. The AE domain variables, AEOUT, AESDTH and AEENDTC/AEENDY should be populated and consistent with the death details.

QS Domain (Questionnaires)

Some items in an instrument may be logically skipped per the instrument's instructions. Responses for logically skipped items should be (1) recorded and/or scored according to the instructions provided in the instrument's user manual, scoring manual, or other documentation provided by the instrument developer and (2) included in the submission dataset.

If instructions on how to record and/or score responses to logically skipped items are available from the instrument developer, then records for logically skipped items should be included in the submission dataset with the following:

- QSSTAT = 'NOT DONE';
- QSREASND = 'LOGICALLY SKIPPED ITEM'; and
- QSORRES, QSSTRESC, and QSSTRESN would be assigned according to the instrument's instructions.

If instructions on how to record and/or score responses to logically skipped items are not available from the instrument developer, then records for logically skipped items should be included in the submission dataset with the following:

- QSSTAT = 'NOT DONE';

- QSREASND = 'LOGICALLY SKIPPED ITEM'; and
- QSORRES, QSSTRESC, and QSSTRESN all set to null.

DV Domain (Protocol Deviations)

The DV domain should be included in your submission. It will be used by reviewers to examine protocol deviation trends of various study sites in order to facilitate the Bioresearch Monitoring Program (BIMO) clinical investigator site selection process, and once FDA tools are developed to extract and format needed data from SDTM, to populate line listings used by the Office of Regulatory Affairs (ORA) investigators during inspections. The following variables besides CDISC required variables should be included in the DV domain when submitting DV data: DVSPID, DVTERM, DVDECOD, DVCAT, DVSCAT, DVSTDTC, DVENDTC and EPOCH.

4.1.2 Analysis Data Model

4.1.2.1 Definition

Specifications for analysis datasets for human drug product clinical studies are provided by the Analysis Data Model (ADaM) and its implementation by the Analysis Data Model Implementation Guide (ADaMIG). ADaM datasets should be used to create and to support the results in clinical study reports (CSRs), Integrated Summaries of Safety (ISS), and Integrated Summaries of Efficacy (ISE), as well as other analyses required for a thorough regulatory review. ADaM datasets can contain imputed data or data derived from SDTM datasets.

4.1.2.2 General Considerations

Generally, ADaM assists FDA review. However, it does not always provide data structured in a way that supports all of the analyses that should be submitted for review. For example, ADaM structures do not support simultaneous analysis of multiple dependent variables or correlation analysis across several response variables. Therefore, sponsors should, as needed, supplement their ADaM datasets after discussions with the specific review division.

One of the expected benefits of analysis datasets that conform to ADaM is that they simplify the programming steps necessary for performing an analysis. As noted above, ADaM datasets should be derived from the data contained in the SDTM datasets. There are features built into the ADaM standard that promote traceability from analysis results to ADaM datasets and from ADaM datasets to SDTM datasets. To ensure traceability, all SDTM variables utilized for variable derivations in ADaM should be included in the ADaM datasets when practical. Each submitted ADaM dataset should have its contents described with complete metadata in the define.xml file (See section 4.1.4.5) and within the ADRG as appropriate (See section 2.3).

4.1.2.3 Dataset Labels

Each dataset should be described by an internal label that is shown in the define.xml file. The label names of ADaM datasets should be different from those of the SDTM datasets. For example, the SDTM adverse event dataset (i.e., AE) and the ADaM adverse event

dataset (i.e., ADAE) should not share the exact same dataset label, such as "Adverse Events."

4.1.2.4 Subject-Level Analysis Data

Subject-Level Analysis Data (ADSL) is the subject-level analysis dataset for ADaM. All submissions containing standard analysis data should contain an ADSL file for each study. In addition to the variables specified for ADSL in the ADaMIG, such as those listed below in the core variables section (See section 4.1.2.5), the sponsor should include multiple additional variables representing various important baseline subject characteristics / covariates presented in the study protocol. Some examples of baseline characteristics / covariates for drug studies include, but are not limited to, disease severity scores such as Acute Physiology and Chronic Health Evaluation (APACHE) scores[30], baseline organ function measurements such as calculated creatinine clearance or Forced Expiratory Volume in 1 second (FEV1), range categories for continuous variables, and numeric date variables in non-International Standards Organization (ISO) formats. Some examples of baseline characteristics for vaccine studies include, but are not limited to, past medical history (e.g., prior infection history), immunosuppressive conditions, prior vaccination history and concomitant medications/vaccines.

4.1.2.5 Core Variables

Core variables, which include covariates presented in the study protocol that are necessary to analyze data, should be included in each ADaM dataset, and are typically already included in the ADSL dataset (See section 4.1.2.4). The core variables included in an ADaM dataset should be necessary for the analysis need in that dataset. Examples of core variables include study/protocol number, center/site number, geographic region, country, treatment assignment information, sex, age, race, analysis population flags (e.g., Intent-to-Treat (ITTFL), Full Analysis Set (FASFL), Safety (SAFFL), and Per-Protocol (PPROTFL)), and other important baseline demographic variables. Note that all variables that contain coded data should be accompanied by a variable that provides the decoded information.

In addition, it is important to note that SDTM datasets do not have core variables (such as demographic and population variables) repeated across the different domains. The duplication of core variables across various domains can be fulfilled through their inclusion in the corresponding analysis datasets. For example, the SDTM AE dataset does not allow for the inclusion of variables such as treatment arm, sex, age, or race. These and other variables should be included in the adverse event ADaM dataset (i.e., ADAE).

4.1.2.6 Key Efficacy and Safety Data

Sponsors should submit ADaM datasets to support efficacy and safety analyses. At least one dataset should be referenced in the data definition file as containing the primary efficacy variables. Further, variables and parameters pertaining to the primary and secondary endpoints of a study, along with their derivations (as applicable), should be provided as well as documented appropriately (i.e., variable-level metadata or parameter value-level metadata) in the data definition file.

[30] Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985). "APACHE II: a severity of disease classification system." Critical Care Medicine, 13 (10): 818–829.29.

4.1.2.7 Timing Variables

A variable for relative day of measurement or event, along with timing variables for visit, should be included when an ADaM dataset contains multiple records per subject (i.e., repeated measures data).

4.1.2.8 Numeric Date Variables

Numeric date variables are needed for analysis and review purposes. Apply formats to all numeric date variables using a format that is understandable by SAS XPORT Version 5 files as per section 3.3.1 above. The software specific (as opposed to study specific) date of reference used to calculate numeric dates should be specified within the ADRG. In the event of partial dates, imputation should be performed only for dates required for analysis according to the SAP, and appropriate corresponding ADaM imputation flags should be utilized. When numeric time or date time variables are needed, all considerations apply as previously discussed for numeric dates.

For traceability purposes, SDTM character dates formatted as ISO 8601 should also be included in the ADaM datasets.

4.1.2.9 Imputed Data

When data imputation is utilized in ADaM, sponsors should submit the relevant supporting documentation (i.e., define.xml and ADRG) explaining the imputation methods.

4.1.2.10 Software Programs

Sponsors should provide the software programs used to create all ADaM datasets and generate tables and figures associated with primary and secondary efficacy analyses. Furthermore, sponsors should submit software programs used to generate additional information included in Section 14 CLINICAL STUDIES of the Prescribing Information[31], if applicable. The specific software utilized should be specified in the ADRG. Refer to FDA Statistical Software Clarifying Statement for more information[32]. The main purpose of requesting the submission of these programs is to understand the process by which the variables for the respective analyses were created and to confirm the analysis algorithms and results. Sponsors should submit software programs in ASCII text format. Executable file extensions should not be used.

4.1.3 Standard for Exchange of Nonclinical Data

4.1.3.1 Definition

The Standard for Exchange of Nonclinical Data (SEND) provides the organization, structure, and format of standard nonclinical (animal toxicology studies) tabulation datasets for regulatory submission. The SEND Implementation Guide (SENDIG v3.0) supports single-dose general toxicology, repeat-dose general toxicology, and carcinogenicity studies. SENDIG v3.1 additionally supports respiratory and cardiovascular safety pharmacology studies.

[31] https://www.fda.gov/downloads/drugs/guidances/ucm075082.pdf

[32] https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM587506.pdf

4.1.3.2 General Considerations

The SENDIG provides specific domain models, assumptions, and examples for preparing standard tabulation datasets that are based on the SDTM model. If there is uncertainty regarding SEND implementation, the sponsor should discuss the issue with the review division.

The ideal time to implement SEND is prior to the conduct of the study as it is very important that the results presented in the accompanying study report be traceable back to the original data collected. Each submitted SEND dataset should have its contents described with complete metadata in the define.xml file (See section 4.1.4.5) and within the nSDRG as appropriate (See section 2.2).

For nonclinical studies, the define.xml StudyName element value should contain the sponsor's study identifier, consistent with the study identifier [study-id] used in the eCTD study tagging file (STF) referenced under the appropriate subsection of Module 4; refer to Section 7.1 for additional information about the STF. For studies outsourced to a contract test facility, the alternate study identifier assigned to the study by the testing facility, which is typically included in the STUDYID field of the SEND datasets, should be included in the ProtocolName element value in define.xml.

For submissions to CDER, SEND datasets are required when submitting a draft report as these data form the basis of regulatory decisions regarding nonclinical support for clinical development. SEND datasets will not be required for CBER submissions until March 15, 2023. If there are changes to the SEND datasets requiring resubmission with the final study report, resubmit the updated datasets using the 'replace' operator. Information about using the 'replace' operator to update datasets can be found in Section 7.1. SEND datasets would not need to be resubmitted with the final report if there were no changes to the dataset from the draft report.

Sponsor should use the VISITDY or --NOMDY variable appropriate to the selected SENDIG version to group observations for summary analysis. This includes grouping animal data collected over multiple days for a single planned event.

For animals necropsied over multiple grace days for a single scheduled interim, terminal or recovery termination event, the DS dataset VISITDY or DSNOMDY variables should contain a single scheduled day for the event. Postmortem findings in DD, MA, MI, OM, and TF for each planned termination event can then be analyzed together based on the DS dataset VISITDY or DSNOMDY. When in-life observations such as terminal body weight or clinical pathology sample collection are scheduled at the time of necropsy, the VISITDY or --NOMDY associated with those observations should also contain the single planned day for the termination event.

For other in-life observations, when the defined schedule for an observation covers multiple days or the schedule is for a specific day but grace days allowed, and animals are observed/tested over multiple days, VISITY or --NOMDY should contain a single day under which the data should be grouped for analysis. Some examples:

• ECGs are scheduled for week 1, and some animals are tested on day 1, some animals are tested on day 2, and some animals are tested on day 3, all animal ECG results for week 1 should have a common VISITDY or EGNOMDY.
• Urinalysis is scheduled for day 15, but no urine was collected from one animal on that day so the collection attempted again on day 16 and was successful. In the Study Report, the data collected for the day 16 urine sample would be analyzed with the day 15 sample results, so VISITDY or LBNOMDY for day 15 and 16 should be 15.
• Due to the number of animals on study, the protocol allows 1 grace day for physical exams with vital signs scheduled for day 1. Some animals are examined on day 1 and some on day 2. All physical exam and vital sign data should be reported under VISTIDY or --NOMDY day 1.

For tests or observations scheduled relative to dose and having --TPTREF and --RFTDTC should be filled to describe the dosing event, and --TPT, --TPTNUM and --ELTM filled to describe the time relative to dose. VISITDY or --NOMDY should always contain the dose day, not the day of the test or observation. VISITDY should be empty for records with unscheduled tests or observations. In SENDIG v3.0, an empty VISITDY identifies data collected for an unplanned event.

Whenever used, FOCID should be consistently represented across domains for the same focus within a study.

SEND datasets modeled in SENDIG v3.0 are required for single-dose toxicology, repeat-dose toxicology, and carcinogenicity studies initiated after December 17, 2016, that are submitted to CDER in NDAs and BLAs, even when those studies did not require SEND datasets for submission in INDs based on the study initiation date (i.e., initiated on or before December 17, 2017). SEND datasets modeled in SENDIG v3.1 are required for single-dose toxicology, repeat-dose toxicology, and carcinogenicity studies initiated after March 15, 2019, that are submitted in NDAs and BLAs and for such studies initiated after March 15, 2020, that are submitted in INDs.

Similarly, SEND datasets modeled in SENDIG v3.1 are required for respiratory and cardiovascular safety pharmacology studies initiated after March 15, 2019, that are submitted to CDER in NDAs and BLAs, even when those studies did not require SEND datasets for submission to CDER in INDs based on the study initiation date (i.e., initiated on or before March 15, 2020).

4.1.3.3 SEND Domain Specification

SUPPQUAL (Supplemental Qualifier)

A SUPPQUAL dataset is a special SEND dataset that contains non-standard variables which cannot be represented in the existing SEND domains. Discussion with the review division should occur if the sponsor intends to include important variables (i.e., that support key analyses) in SUPPQUAL datasets and this should be reflected in the nSDRG.

Currently, SUPPMA, SUPPMI, and SUPPTF should be used to capture some collected information (e.g., pathology modifiers) as detailed in the SENDIG.

Microscopic Findings (MI) Domain

Sponsors should ensure that the transformation of findings from MIORRES to MISTRESC closely adheres to the instructions in the SENDIG. When controlled terminology is not required for MISTRESC, non-neoplastic findings should be standardized and limited to only the base pathological process to ensure that data can be tabulated. For suggestions as to what constitutes a base pathological process, refer to the CDISC NONNEO Controlled Terminology list. Result qualifiers for which there are variables available (e.g. MISEV, MIDTHREL, MICHRON) should be placed appropriately and not duplicated in MISTRESC or SUPPMI.

When using a CDISC CT version dated before 2018 and histopathology severity data are collected on a severity scale that cannot be represented using the CDISC MISEV codelist without a loss of scientific accuracy (e.g. data were collected on 3 levels or 4 levels but MISEV specifies 5 levels), severity scores may be represented in MISEV as '1 OF 4' '2 OF 4' or '1 OF 3'as appropriate, where the first number is the score and the second is the number of available severities in the scale. A score of 1 should be the least severe finding. Extend the non-extensible MISEV codelist with the necessary terms to describe the alternative severity scores, include these extended values in the define.xml and nSDRG, and explain any resulting validation error(s) in the nSDRG.

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.

Laboratory Test Results (LB) Domain

Categorical, noncontinuous results reported as incidence counts rather than summary statistics (i.e. mean and standard deviation) should be placed in LBSTRESC, and even if the categories are numbers, LBSTRESN should be null. Specifically, this includes urinalysis tests where the results are values on a scale. For example, if the allowable values for a urine glucose dipstick test are: 'NEGATIVE', '100', '250', '500', '1000', '>2000', results should only be placed in LBSTRESC. Placing categorical results in LBSTRESC allows straightforward creation of incidence tables on LBSTRESC. The full scale used for laboratory tests with categorical results should be included in the nSDRG.

When a laboratory test result is either above or below the limit of quantification (LOQ) for the measurement method and this result was used in calculation of group means in the study report, the value used for calculation should be submitted using the supplemental qualifier variable LBCALCN.

If an animal is fasted prior to collection of a sample for laboratory testing, all results from testing of the collected sample should have LBFAST= Y.

Pharmacokinetics Concentrations (PC) Domain

The PC domain should support creation of time series graphs and automatic calculation of pharmacokinetic parameters from sets of related plasma concentrations. Three elements are necessary:

- Nominal timings relative to the dose in ISO 8601 duration format
- Grouping of each different set of time series measurements used to calculate a related pharmacokinetic parameter
- Identification of the start of each time series relative to the start of exposure

If the nominal times are provided in PCELTM, nulls should be avoided for plasma concentrations used to calculate a profile. PCDTC and PCDY variables should be populated with actual/collected information when it available; however, for GLP single dose, repeat dose, or carcinogenicity studies where actual/collected information are not readily available to be incorporated into the dataset, these variables may be left null or populated with calculated or nominal dates/times. The use of calculated or nominal dates and times should be mentioned in the nSDRG.

When actual dose dates or date/time values are available for PCRFTDTC/PPRFTDTC, they can be included.

When a test result is below a lower limit of quantitation (LLOQ), it should be submitted using the following instructions:

- PCORRES should not contain a specific value. For example, the value in PCORRES may be '<LLOQ', where LLOQ is the numerical value.
- 'BLQ'[33] should be in PCSTRESC to signify that the result is below the LLOQ.
- PCSTRESN should be blank.
- Standardized units for LOQ should be in PCSTRESU.
- PCLLOQ should be populated with the lower limit of quantitation for the analyte.
- When a numeric value has been assigned to a result that is below the LLOQ for the purpose of group summary statistics, that value should be submitted in SUPPPC as QNAM= 'PCCALCN' to allow the group statistics presented in the study report to be reproduced. When a value that is below the LLOQ is excluded from group statistics, no PCCALCN entry is needed.

Custom Domains

To provide study data that does not fit into an existing SEND domain, draft SEND domain, or published SDTM domain, consider creating a custom dataset aligned with the

[33] According to the FDA's Bioanalytical Method Validation Guidance for Industry (May 2018), study samples with concentrations listed below the LLOQ should be reported as 'BQL'; however, 'BLQ', as specified in FDA-supported SENDIG versions, is appropriate to use in SEND datasets to report this data.

SDTM model. Questions about custom domains should be addressed in pre-submission meetings and documented in the SDSP.

When immune response data are collected in toxicology studies intended for submission to CBER, these data ideally should be submitted in a dataset(s); however, these data currently may be submitted as part of the study report. For data submitted using SENDIG v3.1, use of LB domain or a custom IS domain is acceptable. In these cases, when a numeric value has been assigned for calculation purpose to a result of below limit of quantification, the value should be provided in SUPPLB, as QNAM=LBCALCN; or in SUPPIS, as QNAM = ISCALCN.

Trial Design Model (TDM)

All TDM datasets should be included in SEND submissions as a way to describe the planned conduct of a nonclinical study.

Ensure that Trial Arms and Trial Sets represented in TA and TX closely follow the SENDIG examples of study designs with recovery and/or toxicokinetic animals. Recovery and/or toxicokinetic animals should be presented in separate Trial Sets from the main animals. Trial Sets should be defined to contain animals of both sexes if all other experimental parameters are the same.

The Trial Sets domain (TX) should be submitted for each study. Every set in the TX domain should have only one record with each of the following TXPARMCD values: SPGRPCD (sponsor group code associated with the set), GRPLBL (sponsor group label associated with the set), PLANMSUB (planned number of males in set), and PLANFSUB (planned number of females in set). There should be a one-to-one correspondence between GRPLBL and SPGRPCD entries in the TX domain.

See the appendix section for a list of parameters that should be included in the full Trial Summary (TS) dataset where relevant for nonclinical studies. Additional parameters can be included beyond those listed in the appendix. If information for a parameter listed in the appendix of a full TS.xpt file is not available, the parameter should not be included for datasets modeled in SENDIG v3.0. If information for a parameter listed in the appendix of a full TS.xpt file is not available, it can be included with TSVAL blank and TSVALNF filled for datasets modeled in SENDIG v3.1. For nonclinical studies, study start date (TSPARMCD= STSTDTC) is the date on which the study protocol or plan is approved (signed) by the Study Director, also known as the study initiation date[34].

Tumor Dataset

Carcinogenicity studies should include an electronic dataset of tumor findings to allow for a complete review. At this time, sponsors should continue to include the tumor.xpt and associated define.pdf files regardless of whether the study is in SEND format. When

[34] https://www.fda.gov/media/82716/download

both tumor.xpt and SEND are submitted, the sponsor should ensure that data are consistent and traceable between tumor.xpt and the SEND datasets, with the information specified in the FDA Business Rules. Any information needed to establish traceability should be presented in the nSDRG. The Tumor Findings dataset (tf.xpt) is necessary if the SEND datasets are the basis for creation of the tumor.xpt dataset. If sponsors choose to not submit Tumor Finding dataset (tf.xpt) with the SEND submission, the algorithm used to calculate 'Time in days to detection of tumor' should be included in the nSDRG.

Body Weight Gain (BG) Domain
It is not necessary to include a BG domain in CDER submissions.

Comments (CO) Domain
Comments submitted in the CO domain should be relevant to study interpretation. To reduce ambiguity, abbreviations in any free text field should be avoided or outlined in the nSDRG.

4.1.3.4 Scope of SEND

4.1.3.4.1 Scope of SEND for SENDIGs v3.0 and v3.1

The following is the Agency's current thinking of the scope of SEND for studies listed in the SENDIG version 3.0 and version 3.1, as supported in the FDA Data Standards Catalog. The intent is to provide clarification on the expectation of SEND for studies listed in these SENDIGs, specifically addressing the following language:

> ". . . SENDIG is designed to support data typically found in single-dose general toxicology, repeat-dose general toxicology, and carcinogenicity studies, as well as respiratory and cardiovascular testing done during safety pharmacology studies. . . ."[35]

It is acknowledged that some of these study types can encompass a broad range of study designs (e.g., number of animals per group, number of endpoints tested) and have different drug development purposes (e.g., exploratory or tolerability studies versus standard toxicity studies designed to assess clinical safety). Given the variability of study design and intent of a nonclinical study, the Agency is providing clarification on what studies are subject to the SEND requirement. Study types outlined in the FDA-supported SENDIGs that are out of scope for this discussion include those described in the SENDIG-AR v1.0 (the scope of SEND for nonclinical natural history and efficacy studies in Animal Rule submissions are discussed in 4.1.3.4.2). The Agency's current

[35] See CDISC SENDIG v3.1 (Section 1.1) available at www.cdisc.org.

interpretation of the scope of SEND is subject to change as new SENDIGs are supported and required by the Agency and will be updated in this document as needed.

Overall, the expectation of SEND datasets for nonclinical studies is linked to the pharmacological and toxicological information required to provide FDA with the data needed to assess and support the safety of the proposed clinical investigations.[36] These data form the basis of the rationale on how the sponsor concluded that it is reasonably safe to conduct the proposed clinical trial. If the nonclinical pharmacology or toxicology study is required to support a regulatory decision by the Agency, such that the absence of this study would result in a determination that there is insufficient information to assess the risks to human subjects, then the nonclinical study would require SEND. Further clarification on specific topics is outlined below.

A. SEND is required[37] for single-dose and repeat-dose general toxicology studies that are submitted by the sponsor to support the safety of a proposed clinical trial under commercial IND development or for the support of marketing authorization and/or labeling. These nonclinical studies generally identify potential safety concerns, support the dose and duration of human clinical trials, and characterize the toxicologic profile of the test article and proposed clinical product. Study design incorporates endpoints that can sufficiently inform the potential for clinical adverse events by identifying any nonclinical target organ toxicity and dose or exposure dependency. These studies may be conducted at any point in development ranging from support of an initial safe starting dose for a first-in-human trial to those that support longer duration clinical trials. Submission of these studies may occur at any time during development even if the proposed clinical investigation protocol, that the study supports, has not yet been submitted.

B. When general toxicity studies incorporate other study types (e.g., cardiovascular safety pharmacology, genetic toxicity), SEND datasets for these additional study types would also be expected only when they can be modeled in an FDA-supported SENDIG. For example, if a cardiovascular safety pharmacology study was incorporated into a repeat-dose toxicity study, then SEND would be required for both study types.

C. The age of the animal at study start does not impact whether the SEND requirement applies. Dedicated juvenile animal studies that typically include multiple phases cannot currently be modelled in FDA-supported SENDIGs

[36] 21 CFR 312.23(a)(8).

[37] See the Data Standards Catalog for the latest version of SEND required and the relevant requirement dates for specific submission types. Please note all references to SEND being required in this Guide refer to this standardized format being required for an electronic submission of clinical or nonclinical study data under section 745A(a) of the FD&C Act.

and therefore would not require SEND. However, when general toxicology studies (single- or repeat-dose) are conducted with juvenile animals (e.g., young, post-weaning animals), SEND is required as outlined above.

D. Carcinogenicity studies and repeat-dose toxicity studies that support a carcinogenicity Special Protocol Assessment (SPA) require SEND when they are initiated after an FDA-supported SENDIG requirement date as described in the Data Standards Catalog. These studies are used to inform regulatory decisions related to the risk to human subjects and ultimately impact labeling.

E. The requirement for SEND is not limited to the drug substance. Nonclinical studies that are modeled in an FDA-supported SENDIG version (e.g., repeat-dose toxicology) and are conducted to assess the safety of any component or metabolite of the proposed clinical therapeutic product, require SEND. Examples of such components include but are not limited to the active moiety (API), impurities, excipients, leachables, extractables, pro-drugs, combination products, vaccine adjuvants, and drug/device combinations.

F. The study report status or the finalization of the study report (i.e., draft, interim or final) does not impact whether the SEND requirement applies.

G. The requirement for SEND is not limited to GLP studies. As both GLP and non-GLP toxicity studies may be submitted to the FDA to support clinical safety, the decision for inclusion of SEND is independent of GLP status. In cases where non-GLP toxicity studies are submitted to support a determination of safety, as outlined above, such studies must include SEND.

H. SEND is not required for study types that are not listed in an FDA-supported SENDIG (e.g., primary pharmacology) even if one or more endpoints are able to be modeled in SEND.

I. If SEND datasets are generated by the sponsor for nonclinical studies that are not intended to support clinical safety, the Agency would accept these; however, submission of these datasets would not be required.

Technical Specifications

A. Sponsors are encouraged to use the Study Data Standardization Plan (SDSP) to communicate the intent to submit SEND datasets during product development, and to allow for discussion with the review division when there is any ambiguity on the SEND requirement for a study (See sections 8.2.2 and 2.1 of the Guide).

B. When SEND is not submitted for reasons outlined under Section 4.1.3.4.1 (Scope of SEND for SENDIGs v3.0 and v3.1), use of a simplified ts.xpt file may be needed where the value "NA" (Not Applicable) should be populated in the TSVALNF field (See section 8.2.2 of the Guide).

C. For further information on nonclinical Weight of Evidence documents, refer to Section 7.1 of the Guide.

4.1.3.4.2 Scope of SEND for SENDIG-Animal Rule v1.0

SEND datasets will be required for any nonclinical natural history or efficacy study initiated after March 15, 2022, for NDAs, ANDAs, and BLAs and any nonclinical natural history or efficacy study initiated after March 15, 2023, for certain INDs[38] that are submitted to CDER and for which the CDER review division expects a full tabulation of data (i.e., line listings of the results for each individual animal) to support detailed review. Although not required, FDA also recommends that sponsors submit SEND datasets for such studies that are initiated before March 15, 2022, and March 15, 2023, as applicable. In addition, SEND datasets are recommended for such studies that are submitted to pre-INDs and FDA's Animal Model Qualification Program.

Application-specific questions about which natural history and efficacy studies should include full tabulations of data and datasets should be discussed with the CDER review division as early as possible during product development. Similarly, questions about natural history studies that will be submitted to an animal model qualification package should be discussed with the Animal Model Qualification Program (CDERAnimalModelQualification@fda.hhs.gov).

4.1.4 General Considerations: SDTM, SEND, and/or ADaM

4.1.4.1 Variables in SDTM and SEND: Required, Expected, and Permissible

For the purposes of SDTM and SEND submissions, all required, expected, and permissible variables that were collected, plus any variables that are used to compute derivations, should be submitted.[39]

FDA recognizes that SDTM contains certain operationally derived variables that have standard derivations across all studies (e.g., --STDY, EPOCH). If the data needed to derive these variables are missing, then these variables cannot be derived and the values should be null. The following are examples of some of the permissible and expected variables in SDTM and SEND that should be included, if available:

1. Clinical baseline flags (e.g., last non-missing value prior to first dose) for laboratory results, vital signs, ECG, pharmacokinetic concentrations, and microbiology results. Nonclinical baseline flags (e.g., last non-missing value prior to first dose in parallel design studies) for laboratory results, vital signs, and ECG

[38] On March 11, 2020, FDA published a *Federal Register* notice (85 FR 14205) announcing the dates that FDA's support began and requirements become effective for specific Animal Rule data standards. That document omitted the 36-month implementation period for certain INDs as required by FDA's guidance for industry *Providing Regulatory Submissions in Electronic Format--Standardized Study Data*. On June 10, 2021, FDA published a *Federal Register* notice (86 FR 30960) that corrected that error.

[39] See CDISC SDTM Implementation Guides and the SEND Implementation Guides at www.cdisc.org for additional information on variables referenced throughout this Guide.

results. Currently for SDTM and SEND, baseline flags should be submitted if the data were collected or can be derived.

2. EPOCH designators in SDTM. Please follow CDISC guidance for terminology.[40] The variable EPOCH should be included for clinical subject-level observation (e.g., adverse events, laboratory, concomitant medications, exposure, and vital signs). This will allow the reviewer to easily determine during which phase of the study the observation occurred (e.g., screening, on-therapy, follow-up), as well as the actual intervention the subject experienced during that phase.

3. Whenever --DTC, --STDTC or --ENDTC, which have the role of timing variables, are included in a general observation class domain, the matching study day variables (--DY, --STDY, or --ENDY, respectively) should be submitted. For example, in most findings domains, --DTC is expected, which means that --DY should also be submitted. In the SDTM subject visits domain, SVSTDTC is required and SVENDTC is expected; therefore, both SVSTDY and SVENDY should be submitted.

As mentioned in section 4.1.3.3, in certain GLP nonclinical studies submitted in SEND, PCDTC and PCDY may be imputed.

4.1.4.2 Dates in SDTM and SEND

Dates in SDTM and SEND domains should conform to the ISO 8601 format. Examples of how to implement dates are included in the SDTMIGs and SENDIGs.[41]

4.1.4.3 Naming Conventions in SDTM and SEND

Naming conventions (variable name and label) and variable formats should be followed as specified in the SDTMIGs and SENDIGs.

4.1.4.4 SDTM and SEND Versions

When submitting clinical or nonclinical data, sponsors should not mix versions within a study. As noted above, the Catalog lists the versions that are supported by FDA.

4.1.4.5 Data Definition Files for SDTM, SEND, and ADaM

The data definition file describes the metadata of the submitted electronic datasets, and is considered arguably the most important part of the electronic dataset submission for regulatory review. This data definition specification for submitted datasets defines the metadata structures that should be used to describe the datasets, variables, possible values of variables when appropriate, and controlled terminologies and codes. An insufficiently documented data definition file is a common deficiency that reviewers have noted. Consequently, the sponsor needs to provide complete detail in this file, especially for the specifications pertaining to derived variables. In addition, sponsors should also make certain that the code list and origin for each variable are clearly and easily accessible from the data definition file. The version of any external dictionary should be clearly stated both in the data definition file and in the full TS domain when it is submitted. The

[40] See http://www.cancer.gov/cancertopics/terminologyresources/page6.
[41] See http://www.cdisc.org

internal dataset label should also clearly describe the contents of the dataset. For example, the dataset label for an efficacy dataset might be 'Time to Relapse (Efficacy).'

Separate data definition files should be included for each type of electronic dataset submission, i.e., a separate data definition file for the SDTM datasets of a given clinical study, a separate data definition file for the SEND datasets of a given nonclinical study, and a separate data definition file for the ADaM datasets of a given clinical study. The data definition file should be submitted in XML format, i.e., a properly functioning define.xml[42]. In addition to the define.xml, a printable define.pdf should be provided if the define.xml cannot be printed.[43] To confirm that a define.xml is printable within the CDER IT environment, it is recommended that the sponsor submit a test version to cder-edata@fda.hhs.gov prior to application submission. The Catalog lists the currently supported version(s) of define.xml. It should be noted that define.xml version 2.0 is the preferred version. Sponsors should include a reference to the style sheet as defined in the specification (as listed in the Catalog) and place the corresponding style sheet in the same submission folder as the define.xml file. Within the eCTD study tagging file (STF), valid file-tags for define.xml are 'data-tabulation-data-definition' for SEND or SDTM datasets or 'analysis-data-definition' for ADaM datasets.

4.1.4.6 Annotated Case Report Form (aCRF) for SDTM

An annotated case report form (aCRF) is a PDF document that maps the clinical data collection fields used to capture subject data (electronic or paper) to the corresponding variables or discrete variable values contained within the SDTM datasets. Regardless of whether the clinical database is in a format supported by the Catalog, an aCRF should be submitted preferably at the time a protocol is submitted. The aCRF should be provided as a PDF with the file name 'acrf.pdf.'[44]

The aCRF should include treatment assignment forms, when applicable, and should map each variable on the CRF to the corresponding variables in the datasets (or database). The aCRF should include the variable names and coding for each CRF item.

When data are recorded on the CRF but are not submitted, the CRF should be annotated with the text 'NOT SUBMITTED.' There should be an explanation in the relevant RG stating why these data have not been submitted.

4.1.4.7 Requirements During Specific Public Health Emergencies Declared by the Secretary of HHS

4.1.4.7.1 SEND Requirements During the COVID-19 Public Health Emergency

HHS Declared Public Health Emergency Reference:

[42] See https://www.cdisc.org/standards/data-exchange/define-xml
[43] Detailed FDA PDF specifications are located on FDA's Electronic Common Technical Document Web site, http://www.fda.gov/ectd
[44] Previously acrf.pdf was called blankcrf.pdf.

There is currently an outbreak of respiratory disease caused by a novel coronavirus. The virus has been named "SARS-CoV-2" and the disease it causes has been named "Coronavirus Disease 2019" (COVID-19). On January 31, 2020, the Department of Health and Human Services (HHS) issued a declaration of a public health emergency related to COVID-19 and mobilized the Operating Divisions of HHS.[45]

Impacted Electronic Data Standard(s) and submission type(s):
The Standard for Exchange of Nonclinical Data (SEND) for commercial INDs submitted to CDER.

Rationale and Data Standards Requirement

Datasets for nonclinical studies that can be modeled in an FDA-supported Standard for Exchange of Nonclinical Data (SEND) Implementation Guide (SENDIG) version and were initiated after an applicable SEND implementation date outlined in the FDA Data Standards Catalog are required to be submitted in SEND format. However, for the duration of the COVID-19 public health emergency, to help prevent delays in the initiation of clinical trials for products with a proposed indication to diagnose, cure, mitigate, treat, or prevent COVID-19 (COVID-19 specific indications), FDA will not require these datasets in SEND format until the time of submission of a marketing application for products with COVID-19 specific indications. For further information and resources including the guidance for industry, *Providing Regulatory Submissions In Electronic Format – Standardized Study Data*, refer to the following website: https://www.fda.gov/industry/study-data-standards-resources/study-data-submission-cder-and-cber.

To help simplify submissions for products with COVID-19 specific indications under commercial IND development that currently do not have SEND datasets available for a nonclinical study, FDA recommends that a simplified ts.xpt file be submitted with each nonclinical study requiring SEND, as outlined in the FDA Data Standards Catalog (https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources). The simplified ts.xpt file will help facilitate acceptance of the IND submission at the electronic gateway. The ts.xpt file should include the use of the null value (i.e., "NA") to populate the TSVALNF field. Further instructions for creation of the simplified ts.xpt can be found in this Study Data Technical Conformance Guide under Section 8.2.2 and in the FDA "Simplified ts.xpt creation Guide" (https://www.fda.gov/industry/study-data-standards-resources/study-data-submission-cder-and-cber). Additional questions may be directed to edata@fda.hhs.gov.

[45] Secretary of Health and Human Services, Determination that a Public Health Emergency Exists (originally issued Jan. 31, 2020, and subsequently renewed), *available at* https://www.phe.gov/emergency/news/healthactions/phe/Pages/default.aspx.

5. Therapeutic Area Topics

5.1 General

Generally, when a data standard is released by a Standards Development Organization for public use, it is not supported by FDA until it completes a testing and acceptance process and is announced in the *Federal Register*. Testing and acceptance is conducted to assess the impact of the new standard on FDA medical science review and the consistency and usability of the standard with FDA review tools.

Therapeutic area (TA) standards are not data standards, but rather extend the CDISC foundational standards (e.g., SDTM and ADaM) to represent data that pertain to specific disease areas. CDISC publishes a TA User Guide (TAUG) for each therapeutic area which includes the extensions as disease-specific metadata, examples and recommendations for use (https://www.cdisc.org/standards/therapeutic-areas). The CDISC TAUGs should not be interpreted as FDA guidance.

Questionnaires, Ratings and Scales are often used as outcome measures in clinical studies. The instruments listed in the TAUGs should not be viewed as FDA recommended instruments. Sponsors should consult with the appropriate FDA review division on the best approach for each specific study.

5.2 Supported Therapeutic Areas

Sponsors may use new TA extensions of a CDISC standard, but are not required to until the extensions have been incorporated into a SDTMIG version supported by FDA (the supported SDTMIGs are listed in the Catalog). Sponsors should explain the rationale in the cSDRG for using TA extensions that are not currently listed in this document.

If the study data submitted follows a Therapeutic Area User Guide (TAUG), include the values for TSPARM/TSPARMCD and TSVAL indicated in the table from section 4.1.1.3 in the TS domain.

The TA extensions that are currently incorporated into FDA supported CDISC foundational standards include:

5.2.1 Dyslipidemia Therapeutic Area User Guide v1

5.2.2 Chronic Hepatitis C Therapeutic Area Data Standard User Guide v1

5.2.3 QT Studies Therapeutic Area User Guide v1

5.2.4 Diabetes Therapeutic Area User Guide v1.0 – Supplement for ADaM

5.2.5 Tuberculosis Therapeutic Area User Guide v2.0

5.2.6 Diabetic Kidney Disease Therapeutic Area User Guide v1.0

5.2.7 Ebola Therapeutic Area User Guide v1.0

The Ebola Virus Disease (EVD) Therapeutic Area User Guide (TAUG) identified the ISARIC[46] EVD CORE Clinical Dataset as input; however, only one of the two sets of source data is represented in the TAUG. The Survivor forms are not included because they contain primarily standard data seen in many studies. Sponsors should be aware of both components of the ISARIC CORE Dataset when conducting EVD clinical trials.

5.2.8 Rheumatoid Arthritis Therapeutic Area User Guide v1.0

Standardization for Radiologic Score variables is not available in the Rheumatoid Arthritis TAUG. Sponsors should refer to Radiographic Scoring methods as outcome measures in rheumatoid arthritis for additional guidance. Additionally, while the Controlled Terminology for the HAQ-DI Questionnaire is being finalized by CDISC, sponsors should refer to the Stanford HAQ-DI instrument. It is advised to consult with the review division for further guidance regarding a specific study.

5.2.9 Malaria Therapeutic Area User Guide v1.0

For Transmission Intensity:
Description and implementation examples demonstrating how malaria transmission intensity is calculated at the site are currently not available in the TAUG. Sponsors should consult with the appropriate FDA review division on the best approach for each specific study.

For Meal Data:
Implementation examples demonstrating how the types of meals (i.e., fatty meals or drinks) are currently not available in the TAUG. Sponsors should consult with the appropriate FDA review division on the best approach for each specific study.

[46] International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC)

5.2.10 Kidney Transplant Therapeutic Area User Guide v1.0

The Kidney Transplant TAUG does not address two important data elements. First, the date of the request for a biopsy is important for review, not just the date the biopsy was performed. Second, evidence of C4d staining status in renal allografts (+ or -) is important in the Banff classification criteria for the diagnosis of acute and chronic antibody-mediate rejection. Sponsors should discuss these two data elements with the appropriate review division.

5.2.11 TAUG-Influenza v1.1

5.2.12 Virology Therapeutic Area User Guide v2.1

5.2.13 Prostate Cancer Therapeutic Area User Guide v1.0

The TAUG v1.0 does not include a guidance on where to capture "Reason Not Done" information for the tumor lesions that were Inevaluable (this is a known issue). In addition, the Agency considers it more accurate use the phrase 'tumor lesions' rather than 'tumors'.

Based on datasets previously submitted to the Agency, about 10% of scans are not readable in identifying bone lesions. FDA recommends capturing Image Readability flag for all scans, but the current TAUG does not address this. Sponsors should consult with the appropriate FDA review division on the best approach for each specific study.

For the Disease Assessments and Response for Metastatic Disease, in the proposed Non-Standard Variables (NSV) comparison reference variable CMPREF, FDA recommends providing a value of 'First Post Treatment Scan' instead of 'Flare' to make it more inclusive, as not all subjects will have a flare in the 12 week scans.

FDA recommends submitting patient-level aggregated data if an Independent Review Committee is part of a study and should include the overall assessment of disease status (e.g., disease progression) on bone scans and soft tissue scans (CT or MRI). Sponsors should consult with the appropriate FDA review division on the best approach for each specific study.

5.2.14 Schizophrenia Therapeutic Area User Guide v1.1

The Schizophrenia TAUG does not address two important data elements. First, the subjects daily living situation for the past 12 months. Second, when a protocol violation prompts study termination, sponsors should use the existing Disposition domain as appropriate and provide a referential link to any detailed information regarding the protocol violation. Sponsors should consult with the appropriate FDA review division on the best approach for each specific study.

5.2.15 Major Depressive Disorder Therapeutic Area User Guide v1.0

When reviewing the Major Depressive Disorder TAUG, please also reference the FDA's Guidance for Industry document for MDD. Additionally, please consult the Division of Psychiatry Products when planning the submission.

5.2.16 Traumatic Brain Injury Therapeutic Area User Guide v1.0

5.2.17 Duchenne Muscular Dystrophy Therapeutic Area User Guide v1.0

5.2.18 Vaccines Therapeutic Area User Guide v1.1

The Vaccine TAUG should be used in conjunction with the FDA Guidance for Industry "Submitting Study Datasets for Vaccines to the Office of Vaccines Research and Review." Investigator determined reactogenicity reporting should follow the "Interim User Guide for COVID-19" examples on page 32 with the following revisions:

- inclusion of the Investigator date/time of collection of the event in CE;
- inclusion of additional language in example 2 description first sentence to read "In the study in this example, subjects kept a diary for 3 days assessing the severity of symptoms.";
- change of date/day of investigator assessment in FACE to 2020-04-02 (day 2)
- addition of rows in FACE to report data obtained from the subjects diary from study day 2 (moderate vomiting) and 3 (no vomiting).

The Vaccine TAUG represents the concept of maximum in the NSV, COLSRT (Collected Summary Result Type). We assume that a daily value/result will be a maximum value for the day. The protocol should clarify that a maximum value should be recorded for each day. If you will be reporting more than one value per day, please consult with your review team on how the data should be reported.

5.2.19 Chronic Obstructive Pulmonary Disease Therapeutic Area User Guide v1

5.2.20 Colorectal Cancer Therapeutic Area User Guide v1.0

Issue about Primary Tumor: The TAUG V1.0 does not provide guidance about the identification, location, or laterality of the primary tumor. Even though this is noted as a Known Issue, the importance of primary tumor for colorectal cancer is well established and impacts interpretation of trial results. The FDA recommends that data related to the primary tumor be provided.

Issue about Prior Therapies: The TAUG does not provide guidance about the importance of documenting prior therapies and this is considered an oversight given the importance of these data. The FDA recommends that data related to prior therapies be included in clinical trial data.

Issue about Non-Target Lesions: The TAUG does not discuss the importance of providing data to document the change in size of non-target lesions. This information is required when using certain criteria (e.g., iRECIST). If these data are not provided in the clinical data base, then the response criteria cannot be confirmed by the Agency. Therefore, these data on non-target lesions are necessary if criteria, like iRECIST, is used for trials in colorectal cancer.

5.2.21 Huntington's Disease Therapeutic Area User Guide v1.0

5.2.22 Post Traumatic Stress Disorder Therapeutic Area User Guide v1.0

5.2.23 Clostridium Difficile Associated Diarrhea Therapeutic Area User Guide v1.0

5.2.24 Acute Kidney Injury v1.0

6. Terminology

6.1 General

Common dictionaries should be used across all clinical studies and throughout the submission for each of the following: adverse events, concomitant medications, procedures, indications, study drug names, and medical history. FDA recommends that sponsors use, where appropriate, the terminologies supported and listed in the Catalog. It is important that coding standards, if they exist, be followed (e.g., ICH Medical Dictionary for Regulatory Activities (MedDRA) Term Selection: Points-to-Consider document). Frequently, sponsors submit data that do not conform to terminology standards, for example, misspelling of MedDRA or WHODrug Global terms, lack of conformance to upper / lower case, or the use of hyphens. All controlled terms submitted in datasets should conform to the exact case and spelling used by the terminology maintenance organization (e.g., MedDRA, CDISC controlled terminology). These conformance issues make it difficult to use or develop automated review and analysis tools. The use of a dictionary that is sponsor-defined or an extension of a standard dictionary should be avoided if possible, but, if essential, its use should be documented in the define.xml file and the relevant RGs.

6.1.1 Controlled Terminologies

Controlled terminology standards are an important component of study data standardization and are a critical component of achieving semantically interoperable data exchange (See Appendix A). Generally, controlled terminology standards specify the key concepts that are represented as definitions, preferred terms, synonyms, codes, and code system.

The analysis of study data is greatly facilitated by the use of controlled terms for clinical or scientific concepts that have standard, predefined meanings and representations. In

electronic study data submissions, sponsors should provide the actual verbatim terms that were collected (e.g., on the CRF), as well as the coded term.

Controlled terminology is also useful when consistently applied across studies to facilitate integrated analyses (that are stratified by study) and cross-study comparative analyses (e.g., when greater statistical power is needed to detect important safety signals). Cross-study comparisons and pooled integrated analyses occasionally provide critical information for regulatory decisions, such as statistical results that support effectiveness,[47] as well as important information on exposure-response relationships[48] and population pharmacokinetics[49].

6.1.2 Use of Controlled Terminologies

FDA recognizes that studies are conducted over many years, during which time versions of a terminology may change. Sponsors should use the most recent version of the dictionary available at the start of a clinical or nonclinical study. If a new version becomes available after the start of the study, sponsors may use the most current version of the dictionary for that clinical or nonclinical study. It is common to have different studies use different versions of the same dictionary within the same application (e.g., NDA, BLA). A submission of study data should describe (e.g., in the SDSP or relevant RG) the impact, if any, of the use of different versions on the study results. For example, if the sponsor anticipates pooling coded data across multiple studies, then it may be desirable to use a single version across those studies to facilitate pooling. If a sponsor selects this approach, then the approach and the justification should be documented in the Standardization Plan, or in an update to the plan.

Regardless of the specific versions used for individual studies, pooled analyses (e.g., for an ISS) should be conducted using a single version of a terminology. The current version should be used at the time that data across studies are pooled. This will ensure a consistent and coherent comparison of clinical and scientific concepts across multiple studies. Sponsors should specify the terminologies and versions used in the study in the relevant RG.

---

[47] See the guidance for industry *Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products*, available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072008.pdf. We update guidance periodically. To make sure you have the most recent version of guidance, check the FDA Drugs guidance Web page at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

[48] See the guidance for industry *Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications*, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072109.pdf.

[49] See the guidance for industry *Population Pharmacokinetics*, available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072137.pdf.

6.1.2.1 Use of the Specific Controlled Term 'OTHER'

It is understood that the expansion of controlled terminology may lag behind scientific advancement, and that sometimes there may not be a relevant term within a controlled terminology's value set to describe a clinical trial event, finding, or observation. However, it is not recommended to map a collected value to 'OTHER' when there is a controlled term available to match the collected value – even when the terminology allows for sponsor expansion. Each unique value in a --TERM field mapped to a --DECODE value of 'OTHER' should have a clear rationale outlined in the relevant RGs.

6.1.3 Maintenance of Controlled Terminologies

The use of supported controlled terminologies is recommended wherever available. If a sponsor identifies a concept for which no standard term exists, FDA recommends that the sponsor submit the concept to the appropriate terminology maintenance organization as early as possible to have a new term added to the standard dictionary. FDA considers this *good terminology management practice*. The creation of custom terms (i.e., so-called *extensible* code lists) for a submission is discouraged, because this does not support semantically interoperable study data exchange. Furthermore, the use of custom or extensible code lists should not be interpreted to mean that sponsors may substitute their own nonstandard terms in place of existing equivalent standardized terms. Sponsors should allow sufficient time for a proposed term to be reviewed and included in the terminology, as it is desirable to have the term incorporated into the standard terminology before the data are submitted. If custom terms cannot be avoided, the submitter should clearly identify and define them within the submission, reference them in the relevant RGs, and use them consistently throughout the application.

If a sponsor identifies an entire information domain[50] for which FDA has not accepted a specific standard terminology, the sponsor may select a standard terminology to use, if one exists. FDA recommends that sponsors include this selection in the *Standardization Plan* (See section 2.1) or in an update to the existing plan, and reference it in the relevant RG. If no controlled terminology exists, the sponsor may define custom terms. For clinical studies, the non-FDA supported terms (whether from a non-supported standard terminology or sponsor-defined custom terms) should then be used consistently throughout all relevant studies within the application. Although the consistent use of non-FDA supported terms across all nonclinical studies within an application is recommended, it is understood that that this may not always be possible.

6.2 CDISC Controlled Terminology

Sponsors should use the terminologies and code lists in the CDISC Controlled Terminology, which can be found at the NCI (National Cancer Institute) Enterprise Vocabulary Services.[51] For variables for which no standard terms exists, or if the available terminology is insufficient, the sponsor should propose its own terms. The sponsor should provide this information in the define.xml file and in the relevant RGs.

[50] By *information domain*, we mean a logical grouping of clinical or scientific concepts that are amenable to standardization (e.g., adverse event data, laboratory data, and histopathology data, imaging data).
[51] See http://www.cancer.gov/cancertopics/terminologyresources/page6.

6.3 Adverse Events

6.3.1 MedDRA

6.3.1.1 General Considerations

MedDRA is used for coding adverse events.[52] Generally, the studies included in an application are conducted over many years and may have used different MedDRA versions. The expectation is that sponsors or applicants will use the most current version of MedDRA at the time of study start. However, there is no requirement to recode earlier studies

The spelling and capitalization of MedDRA terms should match the way the terms are presented in the MedDRA dictionary (e.g., spelling and case). Common errors that have been observed include the incorrect spelling of a System Organ Class (SOC) and other MedDRA terms.

To avoid potential confusion or incorrect results, the preparation of the adverse event dataset for the ISS should include MedDRA terms from the most current version of MedDRA at the time that data across studies are pooled. The reason for an ISS based on a single version of MedDRA is that reviewers often analyze adverse events across studies, including the use of Standardized MedDRA Queries.[53] In addition, sponsors should use the MedDRA-specified hierarchy of terms. The SDTM variables for the different hierarchy levels should represent MedDRA-specified primary SOC-coded terms.

6.4 Medications

6.4.1 FDA Unique Ingredient Identifier

6.4.1.1 General Considerations

The Unique Ingredient Identifier (UNII)[54] should be used to identify active ingredients (specifically, active moieties) that are administered to investigational subjects in a study (either clinical or nonclinical). This information should be provided in the SDTM TS domain. UNIIs should be included for all active moieties of investigational products (TSPARMCD= TRT or TRTUNII), active comparators (TSPARMCD= COMPTRT), and any protocol-specified background treatments (TSPARMCD= CURTRT).

If a medicinal product has more than one active moiety, then multiple records in the full TS should be provided, one for each active moiety. For example, if the investigational product is Bactrim (a combination of sulfamethoxazole and trimethoprim), then TS will contain two records for TSPARMCD= TRT: one for sulfamethoxazole and one for trimethoprim.

[52] See https://www.meddra.org/
[53] See http://www.meddra.org/standardised-meddra-queries.
[54] See http://www.fda.gov/ForIndustry/DataStandards/SubstanceRegistrationSystem-UniqueIngredientIdentifierUNII/

The preferred substance names and UNII codes can be found by searching FDA's Substance Registration System, hosted by the National Library of Medicine.[55] We recognize that unapproved substances may not yet have registered UNII codes. We recommend that sponsors obtain UNII codes for unapproved substances as early in drug development as possible, so that relevant information, such as study data, can be unambiguously linked to those substances.

6.4.2 WHODrug Global

6.4.2.1 General Considerations

World Health Organization (WHO) Drug Global[56] is a dictionary maintained and updated by Uppsala Monitoring Centre. WHODrug Global contains unique product codes for identifying drug names and listing medicinal product information, including active ingredients and therapeutic uses.

Typically, WHODrug Global is used to code concomitant medications. The variable --DECOD should be populated with the active substances from the WHODrug Global Dictionary, and --CLAS populated with the drug class.

When using WHODrug Global, --CLAS is recommended to be populated with the Anatomic Therapeutic Chemical (ATC) class most suitable per intended use, and the remainder of the ATC classes, if any, placed in SUPPCM. Alternately, the use of the SUPPCM or FACM domains to populate all ATC Classes associated with the --DECOD value is acceptable. ATC classes should be submitted at the fourth level or most specific available as defined within WHODrug Global.

Generally, studies included in a submission are conducted over many years and may have used different WHODrug Global versions to code concomitant medications. The expectation is the most current B3-format annual version of WHODrug Global at the time of study start will be used to code concomitant medications. There is no requirement to recode earlier studies to align with the WHODrug Global version of later studies.

6.5 Pharmacologic Class

6.5.1 Medication Reference Terminology

6.5.1.1 General Considerations

The Veterans Administration's Medication Reference Terminology (MED-RT)[57] should be used to identify the pharmacologic class(es) of all active investigational substances that are used in a study (either clinical or nonclinical). This information should be provided in the SDTM TS domain when a full TS is indicated. The information should be provided as one or more records in TS, where TSPARMCD= PCLAS.

[55] The Substance Registration System can be accessed at https://fdasis.nlm.nih.gov/srs/
[56] See http://www.who-umc.org/
[57] See https://rxnav.nlm.nih.gov/MED-RT_Documentation.pdf

Pharmacologic class is a complex concept that is made up of one or more component concepts: mechanism of action (MOA), physiologic effect (PE), and chemical structure (CS).[58] The established pharmacologic class is generally the MOA, PE, or CS term that is considered the most scientifically valid and clinically meaningful. Sponsors should include in TS (the full TS) the established pharmacologic class of all active moieties of investigational products used in a study. FDA maintains a list of established pharmacologic classes of approved moieties.[59] If the established pharmacologic class is not available for an active moiety, then the sponsor should discuss the appropriate MOA, PE, and CS terms with the review division. For unapproved investigational active moieties where the pharmacologic class is unknown, the PCLAS record may not be available.

6.6 Indication

6.6.1 SNOMED CT

6.6.1.1 General Considerations

The International Health Terminology Standards Organization's (IHTSDO) Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT)[60] should be used to identify the medical condition or problem that the investigational product in a study is intended to affect (treat, diagnose or prevent, i.e., the indication). This information should be provided in the SDTM TS domain (the full TS) as a record where TSPARMCD= INDIC and TSPARMCD= TDIGRP. SNOMED CT was chosen to harmonize with Indication information in Structured Product Labeling (SPL)[61]. Because the granted indication may include important qualifiers to fulfill the need for adequate directions for use (e.g., descriptors of the population to be treated, adjunctive or concomitant therapy, or specific tests needed for patient selection), the indication section in a label may not be fully represented by available SNOMED CT codes.

6.7 Laboratory Tests

6.7.1 LOINC

6.7.1.1 General Considerations

The Logical Observation Identifiers Names and Codes (LOINC) is a clinical terminology housed by the Regenstrief Institute LOINC codes are universal identifiers for laboratory and other clinical observations that enable semantically interoperable clinical data exchange. The laboratory portion of the LOINC database contains the categories of

[58] See the guidance for industry and review staff *Labeling for Human Prescription Drug and Biologic Products—Determining Established Pharmacologic Class for Use in the Highlights of Prescribing Information*, available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm186607.pdf.

[59] Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StructuredProductLabeling/UCM346147.zip

[60] http://www.ihtsdo.org/snomed-ct/.

[61] See https://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/default.htm

chemistry, hematology, serology, microbiology (including parasitology and virology), toxicology, and more. The SDTM standard supports LOINC codes using the LBLOINC variable. LOINC codes should not be added to SEND datasets.

When submitting LOINC codes you should:

1) Continue submitting laboratory data in the CDISC SDTM format using CDISC laboratory terminology alongside the LOINC code for a given laboratory test.
2) Enter LOINC codes in the LBLOINC field of the SDTM LB domain and populate LBMETHOD when available. When LOINC codes are unavailable, leave the field blank.
3) Submit LOINC codes only when they are available from the clinical laboratories as a pass-through only, i.e. reporting the codes as received from the laboratories with no modifications. FDA understands that there may be inconsistencies in the specification and interpretation of LOINC codes submitted across tests, studies, and subjects.
4) Provide in-vitro diagnostic (IVD) device information in the SDTM Device Identifiers (DI) domain, when available. This information will help inform further FDA guidance on the consistency of LOINC codes associated with laboratory devices.

7. Electronic Submission Format

7.1 eCTD Specifications

Study datasets and their supportive files should be organized into a specific file directory structure when submitted in the eCTD[62] format (See Figure 1 and Table 2 below). Note that this structure is distinct from the eCTD headings and hierarchy folder structure, and does not affect it. Submission of files within the appropriate folders allows automated systems to detect and prepare datasets for review, and minimizes the need for manual processing.

The study identifier (STUDYID in trial summary (TS) and [study-id] in the study tagging file (STF)) should be identical wherever possible.[63] For studies where alignment of the study identifier across TS and STF is not feasible, the value for [study-id] used in the STF should be included in TS using the parameter SPREFID. Though SPREFID is not in the SDTM controlled terminology for TSPARMCD, please use SPREFID to reconcile study identifiers where necessary for SEND or SDTM studies. FDA will use SPREFID to

[62] See http://www.ich.org/products/ctd.html.
[63] ICH M2 EWG: The eCTD Backbone File Specification for Study Tagging Files (June 2008) https://www.ich.org/page/study-tagging-file-specification-and-related-files and CDISC Submission Metadata Model https://www.cdisc.org/system/files/all/reference_material_category/application/pdf/submissionmetadatamodelv2.pdf.

match study identifiers across STF and TS to establish the study start date where necessary for evaluation against the eCTD validation criteria.

Do not use the eCTD 'append' lifecycle operator when submitting updated or changed content within study data files that were previously submitted. Updated files should be submitted using the 'replace' operator.

If you need to split a file that exceeds file size limits (See section 3.3.2), you should submit the smaller split files in the 'split' sub-folder in addition to the larger non-split file in the original data folder. There is no need for a second define.xml file to be submitted within the split subfolder.

For rodent carcinogenicity studies submitted in 4.2.3.4, the tumor.xpt file and its associated define.pdf should be placed in analysis\legacy\datasets subfolder under the study datasets folder.

For information on how to incorporate datasets into the eCTD, please reference the Guidance to Industry *Providing Regulatory Submissions in Electronic Format: Certain Human Pharmaceutical Product Applications and Related Submissions Using the Electronic Common Technical Document Specifications*.[64] The file folder structure for study datasets is summarized in Figure 1. Table 2 provides the study dataset and file folder structure and associated description. For more detailed examples of file folder structures for clinical and non-clinical datasets in both standardized and legacy formats, please see Appendix E: Example Study Data Folder Structures.

When nonclinical Weight of Evidence (WOE) documents are submitted to the Agency as assessments for particular topics or as justification of why a toxicity study is not needed, it is recommended that these toxicity risk assessments are submitted to the nonclinical eCTD Modules relevant to the topic. Examples are listed below:

Rodent Carcinogenicity: Module 4.2.3.4
Reproductive and Developmental Toxicity: Module 4.2.3.5
Juvenile Animal Toxicity: Module 4.2.3.5

Cross-reference to these WOE documents may also be included within eCTD Module 2.4 (Nonclinical Overview summaries). Supporting literature references submitted with any WOE document should be submitted to eCTD Module 4.3 (Literature References). When a WOE document is submitted to an eCTD module that is subject to the Technical Rejection Criteria (e.g., carcinogenicity risk assessment submitted to Module 4.2.3.4), a simplified ts.xpt file must accompany this document. The TSVALNF field of the simplified ts.xpt file should be populated with the null value "NA" (Not Applicable) as further described under Section 8.2.2 (Support on Data Validation Rules) of this Technical Conformance Guide.

[64] See "eCTD Technical Conformance Guide" (Electronic Common Technical Document Technical Conformance Guide (PDF – 160KB)) for further details.

Figure 1: Folder Structure for Study Datasets

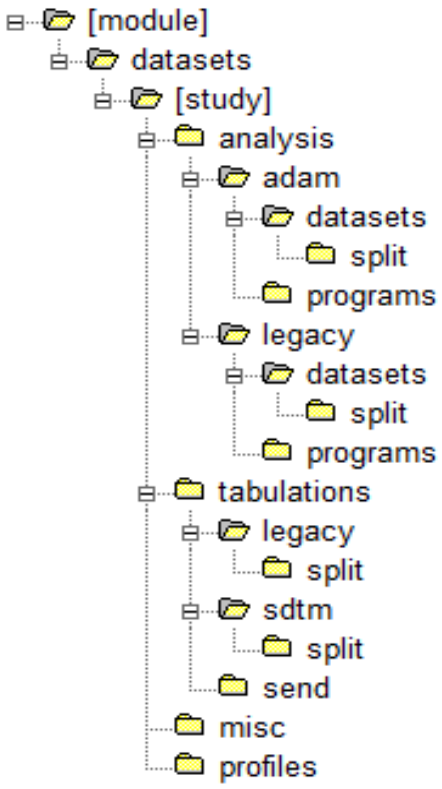


Table 2: Study Dataset and File Folder Structure and Description

Folder Name	Folder Level	Description/Contents
[module]	1	Refers to the eCTD module in which study data are being submitted. Name this folder m4 for nonclinical data and m5 for clinical data. Do not place files at this level.
datasets	2	Resides within the module folder as the top-level folder for study data (nonclinical or clinical) being submitted for the specified module (m4 or m5). Do not place files at this level.
[study]	3	Name this folder with the study identifier or analysis type performed (e.g., study123, iss, ise). Do not place files at this level.
analysis	4	Contains folders for analysis datasets and software programs; arrange in designated level 6 subfolders. Do not place files at this level.
adam	5	Contains subfolders for ADaM datasets and corresponding software programs. Do not place files at this level.
datasets	6	Place ADaM datasets in this subfolder.
split	7	Place any split ADaM datasets in this subfolder.
programs	6	Place software programs for ADaM datasets, tables and figures in this subfolder.
legacy	5	Contains legacy formatted analysis datasets and corresponding software programs. Do not place files at this level.
datasets	6	Place legacy analysis datasets in this subfolder. In m4 place tumor.xpt and its associated define.pdf in this folder.
split	7	Place split legacy analysis datasets in this subfolder.
programs	6	Place software programs for legacy analysis datasets, tables and figures in this subfolder.
misc	4	Place miscellaneous datasets that don't qualify as analysis, profile, or tabulation datasets in this subfolder. This subfolder was formerly named "listings".
profiles	4	Place patient profiles in this subfolder.
tabulations	4	Contains subfolders for tabulation datasets. Do not place files at this level.
legacy	5	Place legacy (non-standardized) tabulation datasets in this folder.
split	6	Place any split legacy tabulations datasets in this subfolder.
sdtm	5	Place SDTM tabulation datasets in this subfolder. Should only be used in m5 for clinical data.
split	6	Place any split SDTM files in this subfolder.
send	5	Place SEND tabulation datasets in this subfolder. Should only be used in m4 for animal data.

7.2 eCTD Sample Submission

The FDA would like to work closely with people who plan to provide a submission using the eCTD specifications and offer to help smooth the process. The Agency also offers a process for submitting sample standardized datasets for validation. Sample submissions are tests only and not considered official submissions. They are not reviewed by FDA reviewers at any time. The Electronic Submissions page provides more information regarding the test submission process.[65]

8. Study Data Validation and Traceability

8.1 Definition of Study Data Validation

Study data validation helps to ensure that the study data are compliant, useful, and will support meaningful review and analysis. Validation activities occur at different times during submission and review of study data, including submission receipt and at the beginning of the regulatory review. Validation of study data that occurs upon receipt of a submission follows the process for Technical Rejection Criteria for Study Data.

8.2 Types of Study Data Validation Rules

1. Standards Development Organizations (e.g., CDISC) provide rules that assess conformance to its published standards (See www.CDISC.org).

2. FDA eCTD Technical Rejection Criteria for Study Data that assess conformance to the standards listed in the Catalog (See above).

3. FDA Business and Validator rules to assess that the data support regulatory review and analysis.

8.2.1 FDA Business and Validator Rules

FDA Business Rules describe the business requirements for regulatory review to help ensure that study data are compliant and useful and support meaningful review and analysis. The list of business rules will grow and change with experience and cross-center collaborations. All business rules should be followed where applicable. The business rules are accompanied with validator rules which provide details regarding FDA's assessment of study data for purposes of review and analysis. The FDA Validator Rules also represent the latest understanding of what best supports regulatory review. The Study Data Standards Resources webpage page provides links to the currently available FDA Business and Validator rules.[66]

[65] See http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm174459.htm

[66] See http://www.fda.gov/eStudyResources

8.2.2 Support on Data Validation Rules

Sponsors should evaluate their study data before submission against the conformance rules published by an SDO, the eCTD Technical Rejection Criteria for Study Data, and the FDA Business Rules. Sponsors may also wish to use the FDA Validator Rules to understand what is available to the FDA reviewer. Sponsors should either correct any discrepancies between study data and the standard or the business rules or explain meaningful discrepancies in the relevant Reviewer Guide (RG). Additional information about conformance to the standard, FDA Business Rules, or FDA Validator Rules that could facilitate review of the submitted data, or establish consistency and traceability between the study data and the Study Report, should also be provided in the relevant RG.

Technical Rejection Criteria and Use of a Simplified ts.xpt for Clinical Studies

Compliance of clinical study reports with applicable standards associated with SDTM and ADaM is ensured by applying Technical Rejection Criteria for Study Data (TRC) when xpt formatted datasets are submitted to FDA in TRC applicable sections within Module 5.

When a xpt formatted dataset is submitted, the STF for the study is then checked for the presence of a trial summary (TS) file (full or simplified). A full ts.xpt file would be expected when the study type and study initiation date meet the criteria for requiring SDTM and ADaM datasets as described in the current FDA Data Standards Catalog.

There are cases in which a xpt formatted dataset submitted to TRC applicable sections within eCTD Module 5 using one of the STFs (see section 7.1) is not required to include accompanying SDTM and ADaM datasets. In such cases, a simplified ts.xpt file should be included with the xpt formatted dataset. A simplified ts.xpt file serves to provide limited machine-readable information such that any submitted xpt formatted dataset not requiring SDTM and ADaM datasets will be appropriately identified by the Center's processing system[67,68].

There may also be cases where SDTM and ADaM are not required even though the study started after December 17, 2016. The list below comprises possible examples (not an exhaustive list):

- pilot studies submitted to an ANDA application
- failed studies submitted to an ANDA application

When SDTM and ADaM are not applicable in a study started after December 17, 2016, the following format of a simplified ts.xpt file should be used, where the TSVALNF field is to be populated with the null value "NA" (Not Applicable):

[67] See Technical Rejection Criteria for Study Data Validation at: https://www.fda.gov/industry/study-data-standards-resources/study-data-submission-cder-and-cber.

[68] See **eCTD Submission Standards** located at https://www.fda.gov/eCTD for further information on the validation tool FDA is currently using and all eCTD validation criteria and rules.

STUDYID	TSPARMCD	TSVAL	TSVALNF
Use Study ID in STF	SSTDTC	*(Leave blank)*	NA

Technical Rejection Criteria and Use of a Simplified ts.xpt for Nonclinical Studies (eCTD Modules 4.2.3.1, 4.2.3.2, and 4.2.3.4) for CDER

Compliance of nonclinical study reports with applicable standards associated with SEND is ensured by applying Technical Rejection Criteria (TRC) to any nonclinical study report submitted to CDER under eCTD modules 4.2.3.1, 4.2.3.2, or 4.2.3.4 that includes one of the following three file tags: 'pre-clinical-study-report', 'legacy-clinical-study-report', or 'study-report-body'.

When a nonclinical study report is submitted using the file tag 'pre-clinical-study-report', 'legacy-clinical study-report', or 'study-report-body' in the study tagging file (STF) or the study is submitted with an xpt formatted dataset, the STF for the study is then checked for the presence of a trial summary (TS) file (full or simplified). A full ts.xpt file would be expected when the study type and study initiation date meet the criteria for requiring SEND datasets as described in the current FDA Data Standards Catalog (e.g., a single dose toxicity study initiated after December 17, 2017 for INDs).

There are cases in which a study report submitted to eCTD Modules 4.2.3.1, 4.2.3.2 or 4.2.3.4 using one of the STFs listed above is not required to include accompanying SEND datasets. In such cases, a simplified ts.xpt file should be included with the study report. A simplified ts.xpt file serves to provide limited machine-readable information such that any submitted study report not requiring SEND will be appropriately identified by the Center's processing system[69,70].

A simplified ts.xpt file would be expected when the study type could be modeled in an applicable SEND Implementation Guide (SENDIG) version (e.g., repeat dose toxicity) but the study initiation date is prior to the implementation of the requirement (e.g., before or on Dec. 17, 2016 for NDAs). When this is the case, the following format of a simplified ts.xpt file may be used:

STUDYID	TSPARMCD	TSVAL	TSVALNF
Use Study ID in STF	STSTDTC	yyyy-mm-dd	*(Leave blank)*

[69] See Technical Rejection Criteria for Study Data Validation at: https://www.fda.gov/industry/study-data-standards-resources/study-data-submission-cder-and-cber

[70] See **eCTD Submission Standards** located at https://www.fda.gov/eCTD for further information on the validation tool FDA is currently using and all eCTD validation criteria and rules.

There may also be cases where a study initiation date is not relevant. When nonclinical submissions are primarily text based, do not have tabulated data or line listings, are specifically sent to or requested by the Agency due to emergent safety concerns (with prior agreement), or only contain data that are not modeled in an applicable SENDIG, a simplified ts.xpt file should be used. The list below comprises possible examples of the types of submissions that meet these criteria (not an exhaustive list):

- Expert pathologist's report (Working Group Report) or Veterinarian report (e.g., Veterinary Cardiologist)
- Nonclinical safety report
- Carcinogenicity protocol amendments or Carcinogenicity risk assessments
- Exploratory or tolerability toxicology study summaries (e.g., text based, limited animals used with few endpoints tested). Does not include those studies that would be submitted to the Agency to support the adequacy of dose selection for subsequent nonclinical studies (e.g., dose range finding studies to support dosing for rodent carcinogenicity studies).
- Literature study reports specifically used as nonclinical support for safety
- Nonclinical study protocols
- Study types not currently modeled in an applicable SENDIG
- Specialized toxicity studies conducted where there are no study parameters modeled in an applicable SENDIG (e.g., a single-dose toxicity study conducted to only assess otic endpoints)
- The Agency, at its discretion, could allow for use of a simplified ts.xpt file with submission of a study report (e.g., for reasons of safety or significant clinical concern)

When a study initiation date is not applicable, the following format of a simplified ts.xpt file should be used, where the TSVALNF field is to be populated with the null value "NA" (Not Applicable):

STUDYID	TSPARMCD	TSVAL	TSVALNF
Use Study ID in STF	STSTDTC	*(Leave blank)*	NA

It is recommended that the Study Data Standardization Plan (SDSP) should be used during development (See section 2.1) to communicate the intent to submit SEND datasets. The SDSP can be updated so that all historical, current, and planned use of study data standards is included. When appropriate, the SDSP may also be used to further explain the intended use of simplified ts.xpt files. SDSP instructions are available (https://www.phuse.eu/css-deliverables) and allow flexibility to accommodate any type of submission. Use of the SDSP will allow for identification of potential data standardization issues and timely discussion with the review division, if needed.

Information on the Technical Rejection Criteria and the FDA Data Standards Catalog may be found at: https://www.fda.gov/industry/study-data-standards-resources/studydata-submission-cder-and-cber

The CDER resource 'Creating Simplified ts.xpt Files', using free and open-source software may be found at https://www.fda.gov/industry/study-data-standardsresources/study-data-submission-cder-and-cber

If there are any questions as to the appropriate use of the simplified ts.xpt file, contact the CDER eDATA Team at cder-edata@fda.hhs.gov or CBER eData Team at cber-edata@fda.hhs.gov.

8.3 Study Data Traceability

8.3.1 Overview

An important component of a regulatory review is an understanding of the provenance of the data (e.g., traceability of the sponsor's results back to the CRF data). Traceability permits an understanding of the relationships between the analysis results (tables, listings and figures in the study report), analysis datasets, tabulation datasets, and source data. Traceability enables the reviewer to accomplish the following:

- Understand the construction of analysis datasets
- Determine the observations and algorithm(s) used to derive variables
- Understand how the confidence interval or the p-value was calculated in a particular analysis
- Relate counts from tables, listings, and figures in a study report to the underlying data

Based upon reviewer experience, establishing traceability is one of the most problematic issues associated with any data conversion. If the reviewer is unable to trace study data from the data collection of subjects participating in a study to the analysis of the overall study data, then the regulatory review of a submission may be compromised. Traceability can be enhanced when studies are prospectively designed to collect data using a standardized CRF, e.g., CDASH. Traceability can be further enhanced when a flow diagram is submitted showing how data move from collection through preparation and submission to the Agency.

Reviewers evaluating nonclinical studies have similar needs to the above list, though in the case of nonclinical studies traceability allows the reviewer to understand and trace relationships between analysis results, single animal listings in the Study Report, and the tabulation data sets. Traceability between the Study Report and tabulation data can be enhanced when data in collection systems has a well-defined relationship to the SEND standard.

8.3.2 Legacy Study Data Conversion to Standardized Study Data

Legacy study data are study data in a non-standardized format, not supported by FDA, and not ever listed in the Catalog. Sponsors should use processes for legacy data conversion that account for traceability. Generally, a conversion to a standard format will map every data element as originally collected to a corresponding data element described

in a standard. Some study data conversions will be straightforward and will result in all data converted to a standardized format. In some instances, it may not be possible to represent a collected data element as a standardized data element. In these cases, there should be an explanation in the RG as to why certain data elements could not be fully standardized or were otherwise not included in the standardized data submission. The legacy data (i.e., aCRF, legacy tabulation data, and legacy analysis data) may be needed in addition to the submission of converted data.

In cases where the data were collected on a Case Report Form (CRF) or electronic CRF but were not included in the converted datasets, the omitted data should be apparent on the annotated CRF and described in the RG. The tabular list of studies in the *Standardization Plan* should indicate which studies contained previously collected non-standard data that were subsequently converted to a standard format.

For nonclinical studies where data are converted to SEND from a previously established collection system, instances may arise where it is not possible to represent a collected data element as a standardized data element. In these cases, there should be an explanation in the nSDRG as to why certain data elements could not be fully standardized or were otherwise not included in the standardized data submission. As the Study Report should contain a complete representation of the study data in the individual animal listings, no non-standardized electronic study data should be submitted.

8.3.2.1 Traceability Issues with Legacy Data Conversion

FDA does not recommend a particular approach to legacy clinical study data conversion, but rather explains the issues that should be addressed so that the converted data are traceable and adequate to support review.

Table 3 presents some of the issues that can be observed during a review when legacy study data are converted to SDTM and submitted with legacy analysis datasets.

Table 3: Traceability Issues: Legacy Data Conversion to SDTM Only

1. Limited ability to determine location of collected CRF variables in the converted SDTM data unless the legacy aCRF is re-annotated.
2. Limited traceable path from SDTM to the legacy analysis data.
3. Limited ability to replicate/confirm legacy analysis datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
4. Limited ability to confirm derivation of intermediate analysis datasets or custom domains.
5. Difficulty in understanding the source or derivation methods for imputed or derived variables in integrated/pooled data, supplemental qualifiers, and related records.

Table 4 presents the issues when legacy study data and legacy analysis data are independently converted to SDTM and ADaM formats, respectively, rather than ADaM datasets being created directly from the SDTM datasets (converted from legacy study data).

Table 4: Traceability Issues: Independent Legacy Data Conversion to SDTM and ADaM

Issues
1. Limited ability to determine location of collected CRF variables in the converted SDTM data unless the legacy aCRF is re-annotated.
2. Limited traceable path from SDTM to the legacy analysis data.
3. Limited ability to replicate/confirm legacy analysis datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
4. Limited ability to confirm derivation of intermediate analysis datasets or custom domains.
5. Limited traceable path from SDTM to the ADaM datasets.
6. Limited ability to replicate ADaM datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
7. Limited traceable path from ADaM to the Tables, Figures and the Clinical Study Report (CSR).
8. Difficulty in understanding the source or derivation methods for imputed or derived variables in integrated/pooled data, supplemental qualifiers, and related records.

Table 5 presents the issues when legacy data are converted to SDTM and ADaM formats in sequence (i.e., converting legacy study data to SDTM and then creating ADaM from the SDTM). The key concern is the traceability from ADaM to the Tables, Figures and CSR.

Table 5: Traceability Issues: Legacy Data Conversion to SDTM and ADaM in Sequence

1. Limited ability to determine location of collected CRF variables in the converted SDTM data unless the legacy aCRF is re-annotated.
2. Limited traceable path from SDTM to the legacy analysis data.
3. Limited ability to replicate/confirm legacy analysis datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
4. Limited ability to confirm derivation of intermediate analysis datasets or custom domains.
5. Limited traceable path from ADaM to the Tables, Figures and the CSR.
6. Difficulty in understanding the source or derivation methods for imputed or derived variables in integrated/pooled data, supplemental qualifiers, and related records.

8.3.2.2 Legacy Data Conversion Plan and Report

Sponsors should evaluate the decision involved in converting previously collected non-standardized data (i.e., legacy study data) to standardized data (i.e., SDTM, and ADaM). Sponsors should provide the explanation and rationale for the study data conversion in the RG. To mitigate traceability issues when converting legacy data, FDA recommends the following procedures:

1. Prepare and submit a legacy data conversion plan and report.
   - The plan should describe the legacy data and the process intended for the conversion.
   - The report should present the results of the conversions, issues encountered and resolved, and outstanding issues.
   - The plan and report should be provided in the SDRG.
2. Provide an aCRF, for clinical data, that maps the legacy data elements.
   1. Sponsors should provide two separate CRF annotations, one based on the original legacy data, and the other based on the converted data (i.e., SDTM) when legacy datasets are submitted. The legacy CRF tabulation data should include all versions and all forms used in the study.
3. Record significant data issues, clarifications, explanations of traceability, and adjudications in the RG. For example, data were not collected or were collected using different/incompatible terminologies, or were collected but will not fit into, for example, SDTM format.
4. Legacy data (i.e., legacy aCRF, legacy tabulation data, and legacy analysis data) may be needed in addition to the converted data.

Submission of a Legacy Data Conversion Plan and Report is not expected for nonclinical studies where data were collected in a previously established data collection system.

Appendix A: Data Standards and Interoperable Data Exchange

This appendix provides some of the guiding principles for the Agency's long-term study data standards management strategies. An important goal of standardizing study data submissions is to achieve an acceptable degree of *semantic interoperability* (discussed below). This appendix describes different types of interoperability and how data standards can support interoperable data exchange now and in the future.

At the most fundamental level, study data can be considered a collection of data elements and their relationships. A data element is the smallest (or *atomic*) piece of information that is useful for analysis (e.g., a systolic blood pressure measurement, a lab test result, a response to a question on a questionnaire).

A data value is by itself meaningless without additional information about the data (so called *metadata*). Metadata is often described as *data about data*. Metadata is structured information that describes, explains, or otherwise makes it easier to retrieve, use, or manage data.[71] For example, the number *44* itself is meaningless without an association with Hematocrit and the unit of measurement (e.g. "%"). Hematocrit in this example is metadata that further describes the data.

Just as it is important to standardize the representation of data (e.g., M and F for male and female, respectively), it is equally important to standardize the metadata. The expressions Hematocrit = 44; Hct = 44, or Hct Lab Test = 44 all convey the same information to a human, but an information system or analysis program will fail to recognize that they are equivalent because the metadata is not standardized. It is also important to standardize the definition of the metadata, so that the meaning of a hematocrit value is constant across studies and submissions.

In addition to standardizing the data and metadata, it is important to capture and represent relationships (also called associations) between data elements in a standard way. Relationships between data elements are critical to understand or interpret the data. Consider the following information collected on the same day for one subject in a study:

Systolic Blood Pressure = 90 mmHg
Position = standing
Systolic Blood Pressure = 110 mmHg
Time = 10:23 a.m.
Time = 10:20 a.m.
Position = lying

[71] Metadata is said to "give meaning to data" or to put data "in context." Although the term is now frequently used to refer to XML (extensible markup language) tags, there is nothing new about the concept of metadata. Data about a library book such as author, type of book, and the Library of Congress number, are metadata and were once maintained on index cards. SAS labels and formats are a rudimentary form of metadata, although they have not historically been referred to as metadata.

When presented as a series of unrelated data elements, they cannot reliably be interpreted. Once the relationships are captured, as shown below using arrows, the interpretation of a drop in systolic blood pressure of 20 mmHg while standing, and therefore the presence of clinical orthostatic hypotension, is possible. Standardizing study data therefore involves standardizing the data, metadata, and the representation of relationships.

Time = 10:20 a.m. ←→ Position = lying ←→ Systolic Blood Pressure = 110 mmHg
Time = 10:23 a.m. ←→ Position = standing ←→ Systolic Blood Pressure = 90 mmHg

With these fundamental concepts of data standardization in mind, data standards can be considered in the context of interoperable data exchange.

Interoperability

Much has been written about interoperability, with many available definitions and interpretations within the health care informatics community. In August 2006, the President signed an Executive Order mandating that the Federal Government use interoperable data standards for health information exchange.[72] Although this order was directed at Federal agencies that administer health care programs (and therefore not the FDA), it is relevant to this guidance because it defined interoperability for use by Federal agencies:

"Interoperability" means the ability to communicate and exchange data accurately, effectively, securely, and consistently with different information technology systems, software applications, and networks in various settings, and exchange data such that clinical or operational purpose and meaning of the data are preserved and unaltered.

Achieving interoperable study data exchange between sponsors, applicants and FDA is not an all-or-nothing proposition. Interoperability represents a continuum, with higher degrees of data standardization resulting in greater interoperability, which in turn makes the data more useful and increasingly capable of supporting efficient processes and analyses by the data recipient. It is therefore useful to understand the degree of interoperability that is desirable for standardized study data submissions.

In 2007, the Electronic Health Record Interoperability Work Group within Health Level Seven issued a white paper that characterized the different types of interoperability based on an analysis of how the term was being defined and used in actual practice.[73] Three types of interoperability were identified: technical, semantic, and process interoperability. A review of these three types provides insight into the desired level of interoperability for standardized study data submissions.

Technical interoperability describes the lowest level of interoperability whereby two different systems or organizations exchange data so that the data are useful. The focus of

[72] See http://www.cga.ct.gov/2006/rpt/2006-R-0603.htm.
[73] See Coming to Terms: Scoping Interoperability for Health Care http://www.hln.com/assets/pdf/Coming-to-Terms-February-2007.pdf.

technical interoperability is on the conveyance of data, not on its meaning. Technical interoperability supports the exchange of information that can be used by a person but not necessarily processed further. When applied to study data, a simple exchange of non-standardized data using an agreed-upon file format for data exchange (e.g., SAS transport file) is an example of technical interoperability.

Semantic interoperability describes the ability of information shared by systems to be understood, so that nonnumeric data can be processed by the receiving system. Semantic interoperability is a multi-level concept with the degree of semantic interoperability dependent on the level of agreement on data content terminology and other factors. With greater degrees of semantic interoperability, less human manual processing is required, thereby decreasing errors and inefficiencies in data analysis. The use of controlled terminologies and consistently defined metadata support semantic interoperability.

Process interoperability is an emerging concept that has been identified as a requirement for successful system implementation into actual work settings. Simply put, it involves the ability of systems to exchange data with sufficient meaning that the receiving system can automatically provide the right data at the right point in a business process.

An example of process interoperability in a regulatory setting is the ability to quickly and automatically identify and provide all the necessary information to produce an expedited adverse event report in a clinical trial upon the occurrence of a serious and unexpected adverse event. The timely submission of this information is required by regulation to support FDA's mandate to safeguard patient safety during a clinical trial. Process interoperability becomes important when particular data are necessary to support time-dependent processes.

Because the vast majority of study data are submitted after the study is complete, achieving process interoperability for study data submissions in a regulatory setting is relatively unimportant, at least for the foreseeable future. It is reasonable to conclude that it is most desirable to achieve *semantic interoperability* in standardized study data submissions.

In summary, the goal of standardizing study data is to make the data more useful and to support semantically interoperable data exchange between sponsors, applicants, and the FDA such that it is commonly understood by all parties.

Appendix B: Trial Summary (TS) Parameters for Submission – Clinical

FDA Desired - Clinical	TSPARMCD	TSPARM	FDA Notes
Y	ACTSUB	Actual Number of Subjects	
Y	ADAPT	Adaptive Design	
Y	ADDON	Added on to Existing Treatments	
Y	AGEMAX	Planned Maximum Age of Subjects	
Y	AGEMIN	Planned Minimum Age of Subjects	
Y	COMPTRT	Comparative Treatment Name	
Conditional	CRMDUR	Confirmed Response Minimum Duration	If applicable.
Conditional	CTAUG	CDISC Therapeutic Area User Guide	If applicable. the value should be the exact listing as in section 5.2 of the Technical Conformance Guide. Use as many rows as needed.
Conditional	CURTRT	Current Therapy or Treatment	Where ADDON = 'Y'. Use as many rows as needed.
Y	DCUTDESC	Data Cutoff Description	GRPID relates DCUTDTC to DCUTDESC.
Y	DCUTDTC	Data Cutoff Date	GRPID relates DCUTDTC to DCUTDESC.
Conditional	EGBLIND	ECG Reading Blinded	For QT submissions.
Conditional	EGCTMON	ECG Continuous Monitoring	For QT submissions.
Conditional	EGLEADPR	ECG Planned Primary Lead	For QT submissions.
Conditional	EGLEADSM	ECG Used Same Lead	For QT submissions.
Conditional	EGRDMETH	ECG Read Method	For QT submissions.
Conditional	EGREPLBL	ECG Replicates at Baseline	For QT submissions.
Conditional	EGREPLTR	ECG Replicates On-Treatment	For QT submissions.
Conditional	EGTWVALG	ECG Twave Algorithm	For QT submissions.
Y	EXTTIND	Extension Trial Indicator	
Y	FCNTRY	Planned Country of Investigational Sites	Use as many rows as needed.
Conditional	FDATCHSP	FDA Technical Specification	If applicable. the value should be the exact listing as in the appendix of the Technical Conformance

FDA Desired - Clinical	TSPARMCD	TSPARM	FDA Notes
			Guide. Use as many rows as needed.
Y	HLTSUBJI	Healthy Subject Indicator	
Conditional	INDIC	Trial Disease/Condition Indication	For a healthy volunteer study, TSVALNF = 'NA'.
Conditional	INTMODEL	Intervention Model	Where STYPE = 'INTERVENTIONAL'.
Conditional	INTTYPE	Intervention Type	Where STYPE = 'INTERVENTIONAL'.
Y	LENGTH	Trial Length	
Y	NARMS	Planned Number of Arms	
Y	NCOHORT	Number of Groups/Cohorts	
Y	OBJPRIM	Trial Primary Objective	Use as many rows as needed.
Y	OBJSEC	Trial Secondary Objective	Use as many rows as needed.
Conditional	OUTMSEXP	Exploratory Outcome Measure	If applicable. Use as many rows as needed.
Y	OUTMSPRI	Primary Outcome Measure	Use as many rows as needed.
Conditional	OUTMSSEC	Secondary Outcome Measure	Use as many rows as needed.
Conditional	PCLAS	Pharmacologic Class	If STYPE = 'INTERVENTIONAL' and where applicable for INTTYPE.
Y	PDPSTIND	Pediatric Postmarket Study Indicator	
Y	PDSTIND	Pediatric Study Indicator	
Y	PIPIND	Pediatric Investigation Plan Indicator	
Y	PLANSUB	Planned Number of Subjects	
Conditional	RANDQT	Randomization Quotient	Where '1' denotes all subjects randomized to the investigational treatment.
Y	RDIND	Rare Disease Indicator	
Y	REGID	Registry Identifier	Use as many rows as needed.
Conditional	RLPSCRIT	Relapse Criteria	If applicable.

FDA Desired - Clinical	TSPARMCD	TSPARM	FDA Notes
Conditional	SDMDUR	Stable Disease Minimum Duration	If applicable.
Y	SENDTC	Study End Date	
Y	SEXPOP	Sex of Participants	
Y	SPONSOR	Clinical Study Sponsor	
Y	SDTMVER	SDTM Version	The value should be the exact term listed in the FDA Data Standards Catalog in Column E. If multiple SDTM Versions are used for a study the every version should be listed on each row.
Y	SDTIGVER	SDTM IG Version	The value should be the exact term listed in the FDA Data Standards Catalog in Column F. If multiple SDTM IG Versions are used for a study the every version should be listed on each row.
Y	STOPRULE	Study Stop Rules	If no stopping rule, STOPRULE = 'NONE'.
Conditional	STRATFCT	Stratification Factor	If applicable. Use as many rows as needed.
Y	SSTDTC	Study Start Date	
Y	STYPE	Study Type	
Y	TBLIND	Trial Blinding Schema	
Y	TCNTRL	Control Type	
Conditional	TDIGRP	Diagnosis Group	Where HLTSUBJI = 'N'.
Y	THERAREA	Therapeutic Area	
Y	TITLE	Trial Title	Use as many rows as needed.
Y	TPHASE	Trial Phase Classification	
Conditional	TRT	Investigational Therapy or Treatment	If STYPE = 'INTERVENTIONAL'.
Y	TTYPE	Trial Type	Use as many rows as needed.

Appendix C: Trial Summary (TS) Parameters for Submission – Nonclinical

The term "Conditional" means a parameter might not be relevant to a trial or study design. If the TS Parameter is relevant to the study design and is listed as "Conditional" in the table below, it should be included in the SEND dataset submitted to the FDA.

FDA Desired - Nonclinical	TSPARMCD	TSPARM	FDA Notes
See Notes	AGE	Age	Age of subjects planned for the study population as an integer. Either AGE or AGETXT should be populated (not both). If the planned age is a range, then use AGETXT
See Notes	AGETXT	Age Text	Age of subjects planned for the study population expressed as a range. Either AGE or AGETXT should be populated (not both). If an age integer value is available, populate the AGE variable instead
Y	AGEU	Age Unit	
Conditional	ASOCSTDY	Associated Study	If applicable.
Y	DOSDUR	Dosing Duration	
Y	DOSENDTC	End Date/Time of Dose Interval	
Y	DOSSTDTC	Start Date/Time of Dose Interval	
Y	EXPENDTC	Experimental End Date	
Y	EXPSTDTC	Experimental Start Date	
Y	GLPFL	GLP Flag	
Y	GLPTYP	Good Laboratory Practice Type	
Conditional	INTSAC	Time to Interim Sacrifice	Include when the study has an interim sacrifice
Y	PCLASS	Pharmacologic Class	We recognize that pharmacologic class can change throughout the drug development timeline. Refer to the FDA Established Pharmacologic Class (EPC) Text Phrase Document (https://www.fda.gov/media/144963/download).
Conditional	PDOSFRQ	Planned Dose Frequency	The planned number of doses administered per a specific interval, as defined in the SENDIG Animal Rule v1.0. Use of PDOSFRQ is recommended for all study types modelled in FDA-supported SENDIG versions when relevant.
Conditional	RECSAC	Recovery Period	Include when the study has a recovery sacrifice

FDA Desired - Nonclinical	TSPARMCD	TSPARM	FDA Notes
Y	ROUTE	Route of Administration	
Conditional	SBSTRAIN	Strain/Substrain Details	If applicable.
Y	SDESIGN	Study Design	
Y	SEXPOP	Sex of Participants	
Y	SLENGTH	Study Length	
Y	SNDCTVER	SEND Controlled Terminology Version	
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.
Y	STITLE	Study Title	
Y	STRAIN	Strain/Substrain	
Y	STRPSTAT	Study Report Status	The status of the study report associated with the dataset, as defined in the SENDIG Animal Rule v1.0. Use of STRPSTAT is recommended for all study types modelled in FDA-supported SENDIG versions.
Y	STSTDTC	Study Start Date	

FDA Desired - Nonclinical	TSPARMCD	TSPARM	FDA Notes
Y	TFCNTRY	Test Facility Country	
Y	TRMSAC	Time to Terminal Sacrifice	
Y	TRT	Investigational Therapy or Treatment	
Y	TRTCAS	Primary Treatment CAS Registry Number	We recognize that the CAS number may not be immediately available, especially at the opening IND submission.
Y	TRTUNII	Primary Treatment Unique Ingredient ID	We recognize that the UNII code may not be immediately available, especially at the opening IND submission.
Y	TRTV	Treatment Vehicle	
Y	TSTFLOC	Test Facility Location	
Y	TSTFNAM	Test Facility Name	

Appendix D: Additional Documents Evaluated By FDA

The Agency recognizes that there are may be additional documents beyond Therapeutic Area User Guides (TAUGs), Implementation Guides (IGs), and Models that provide technical information about how to implement a CDISC standard and that these documents fall outside the scope of the FDA Data Standards Catalog. Use of the documents listed here is encouraged. For documents not yet listed here, please consult with your division.

1. **CDISC Document: Interim User Guide for COVID-19**

2. **CDISC Document: Guidance for Ongoing Studies Disrupted by COVID-19 Pandemic**

   It is the current preference of the Agency that for all clinical studies, not limited to those impacted by COVID-19, subject visit data for scheduled (whether or not they occurred), and unscheduled visits be submitted in one single dataset structured as the current CDISC Subject Visits (SV) domain. It is also Agency preference that three non-standard variables (NSVs) for missed visits, --REASOC (Reason for Occur Value), --EPCHGI (Epi/Pandemic Related Change Indicator), and --CNTMOD (Contact Mode), outlined in the CDISC document "Guidance for Ongoing Studies Disrupted by COVID-19 Pandemic" be included within the SV domain and not within the supplemental SUPPSV domain or in other SDTM datasets. Submitting subject visits information in one single structured dataset allows both the human and technology consumer of this information to operate efficiently and with confidence that all visit data are considered during regulatory review.

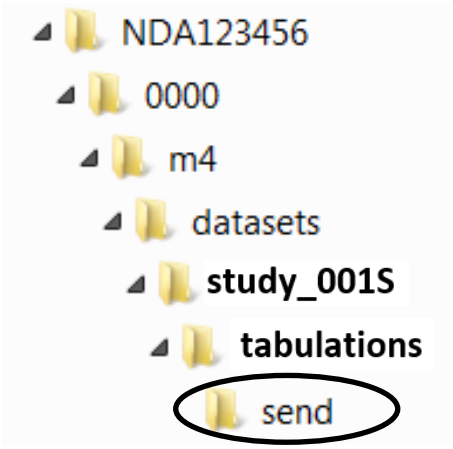
   As always, consult with the relevant FDA review division for the best approach in a specific application. Further updates to Agency thinking regarding how to submit data for studies that may have been impacted by the COVID-19 pandemic will be posted in updates to the Study Data Technical Conformance Guide.

3. **Occurrence Dataset Structure (OCCDS) v1.0**

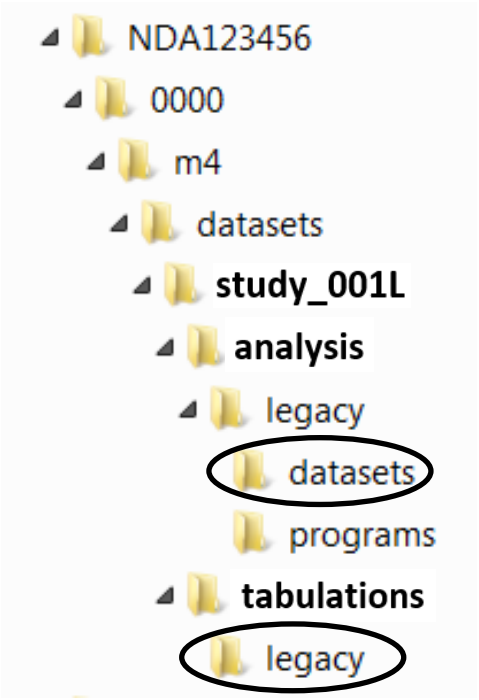
Appendix E: Example Study Data Folder Structure

M4 (Non-Clinical)

Study_001S (Standardized Data Tabulation Datasets):

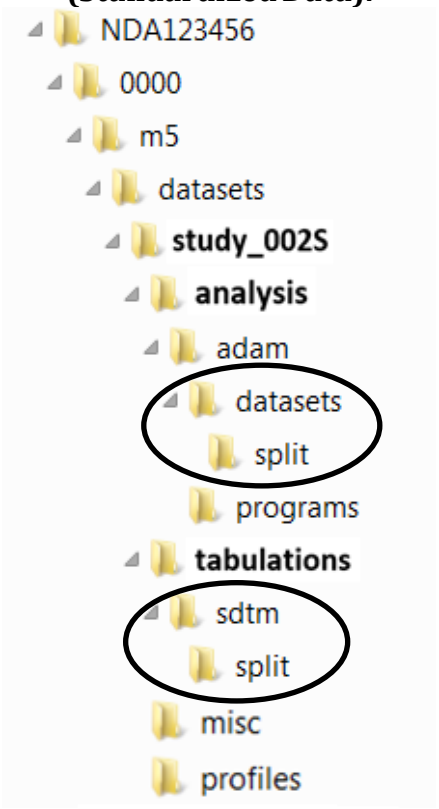


Study_001L (Legacy Data):



M5 (Clinical)

Study_002S (Standardized Data):



Study_002L (Legacy Data):

Glossary

The following is a list of acronyms and terms used in this Guide:

aCRF:	Annotated Case Report Form
ANDA:	Abbreviated New Drug Application
ADaM:	Analysis Data Model
ADRG:	Analysis Data Reviewer's Guide
ADSL:	Subject-Level Analysis Data
ASCII:	American Standard Code for Information Interchange
CBER:	Center for Biologics Evaluation and Research
CDASH:	Clinical Data Acquisition Standards Harmonization
CDER:	Center for Drug Evaluation and Research
CDISC:	Clinical Data Interchange Standards Consortium
CS:	Chemical Structure
CSR:	Clinical Study Report
eCTD:	Electronic Common Technical Document
GLP:	Good Laboratory Practice
ICH:	International Council for Harmonisation
IND:	Investigational New Drug
ISE:	Integrated Summary of Efficacy
ISO:	International Organization for Standardization
ISO 8601:	ISO character representation of dates, date/times, intervals, and durations of time
ISS:	Integrated Summary of Safety
ITT:	Intent-to-Treat
LOINC:	Logical Observation Identifiers and Codes
MedDRA:	Medical Dictionary for Regulatory Activities
MOA:	Mechanism of Action
NDA:	New Drug Application
NDF-RT:	National Drug File – Reference Terminology
PDF:	Portable Document Format
PE:	Physiologic Effect
RG:	Reviewer Guides (e.g., cSDRG, nSDRG, ADRG located in eCTD m4 and m5)
SDRG:	Study Data Reviewer Guide (original term, replaced by cSDRG and nSDRG)
cSDRG:	SDRG used for clinical data
nSDRG:	SDRG used for nonclinical data
SDTM:	Study Data Tabulation Model
SEND:	Standard for Exchange of Nonclinical Data
SNOMED:	Systematized Nomenclature of Medicine
UNII:	Unique Ingredient Identifier
XML:	eXtensible Markup Language
XPORT:	SAS Transport Version 5