CHANGES IN FDA TECHNICAL CONFORMANCE GUIDE V3.1,V3.2 AND V3.3

05-Seb-2017, Ingelheim , Boehringer Ingelheim Pharma GmbH & Co. KG
Marion Friebel

PAREXEL®
YOUR JOURNEY. OUR MISSION.®

DEFINITION

- **This Study Data Technical Conformance Guide provides specification, recommendations and general considerations on how to submit standardized study data using FDA–support**

- **The guide is separated in the same sections as before:**

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

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.

DEFINITION - CONTINUE

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 FDA-supported therapeutic area standards (TA).

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

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

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

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 1.4

V3.3

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

- more specific : nonclinical study data added

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 2.2

THIS POINTS AND LINKS WERE ADDED IN V3.2:

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).

---

[12] A specific template for a Study Data Reviewer's Guide for clinical studies is not specified. However, an example of a Study Data Reviewer's Guide (e.g., template, completion guidelines and examples) can be found at http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide.

[13] A specific template for a Study Data Reviewer's Guide for nonclinical studies is not specified. However, an example of a Study Data Reviewers Guide (e.g., template, recommendations and examples) can be found at http://www.phusewiki.org/wiki/index.php?title=Nonclinical_Study_Data_Reviewers_Guide.

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 2.3

THIS POINT WAS ADDED IN V3.2:

- Additional information about ADRG . This follow the new naming convention

- An ADRG for clinical data should be called an ADRG and the document should be a PDF file 'adrg.pdf' upon submission.

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 3.3

v3.1

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. 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.

v3.3

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.

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 4.1

4.1.1.2 SDTM General Considerations

v3.1

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.

v3.3

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 is provided it should be clearly identified so that the reviewer can distinguish the results of adjudication from data as originally collected.

CONTINUE-UPDATES IN SECTION 4.1.

4.1.1.3 SDTM Domain Specifications

v3.1

DS Domain (Disposition)
When there is more than one disposition event, the EPOCH variable should be used to aid in distinguishing between them. This will allow identification of the EPOCH in which each event occurred. 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.

v3.3

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.

CONTINUE - UPDATES IN SECTION 4.1.

4.1.2.2 General Considerations

v3.1

Generally, ADaM facilitates FDA review. 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).

v3.3

Generally, ADaM facilitates 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).

CONTINUE - UPDATES IN SECTION 4.1.

4.1.2.10 Software Programs

v3.1

Sponsors should provide the software programs used to create all ADaM datasets along with the tables and figures associated with primary and secondary efficacy analyses in order to help reviewers to better understand how the datasets, tables and figures were created. The specific software utilized should be specified in the ADRG. 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. Therefore, it is not necessary to submit the programs in a format or content that allows the FDA to directly run the program under its given environment. Any submitted programs (scripts) generated by an analysis tool should be provided as ASCII text files or PDF files, e.g., adsl.sas should be submitted as either adsl.txt or adsl.pdf.

v3.3

Sponsors should provide the software programs used to create all ADaM datasets along with the tables and figures associated with primary and secondary efficacy analyses in order to help reviewers to better understand how the datasets, tables and figures were created. The specific software utilized should be specified in the ADRG. 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. Sponsors should not submit software programs with executable file extensions. Sponsors should submit in ASCII text format.

CONTINUE - UPDATES IN SECTION 4.1.

4.1.3.2 General Considerations

v3.1

.....

v3.3

.....

Sponsors should use the VISITDY variable if findings, which were intended to be analyzed together, were collected across multiple study days. For postmortem findings in MA, MI, and OM, indicate groupings of grace day data collections using the VISITDY variable in the DS domain. For in-life findings domains like LB or EG, add VISITDY to the domain to indicate grouping of measurements across grace days when measurements are grouped in the Study Report. For example, an ECG might be collected on Day 20, determined to be uninterpretable, and repeated on Day 21. If those ECG findings are grouped for analysis in the Study Report, VISITDY should be provided and set to Day 20 for both ECG collections to provide traceability in the SEND dataset.

CONTINUE - UPDATES IN SECTION 4.1.

4.1.3.3 SEND Domain Specification

v3.1

Currently, SUPPQUAL should be used to capture some collected information (e.g., pathology modifiers) until the SEND is further refined to adequately represent such information.

Microscopic Findings (MI) Domain

Sponsors should ensure that the transformation of findings from MIORRES to MISTRESC closely adheres to the instructions in the SENDIG. Modifiers for which there are variables available (e.g. MISEV, MILAT, etc.) should be placed appropriately. There should be no severities (e.g., minimal, mild, etc.) included in MISTRESC. Sponsors should use the VISITDY variable if postmortem findings which were intended to be analyzed together were collected across multiple study days.

Macroscopic Findings (MA) Domain

Sponsors should use the VISITDY variable if postmortem findings which were intended to be analyzed together were collected across multiple study days.

v3.3

Currently, SUPPMA and SUPPMI 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. Non-neoplastic findings in MISTRESC, where controlled terminology is not required, 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. Details and severities for which there are variables available (e.g. MISEV, MILAT, MIANTREG, etc.) should be placed appropriately and not duplicated in MISTRESC or SUPPMI.

CONTINUE - UPDATES IN SECTION 4.1.

4.1.3.3 SEND Domain Specification

v3.1

v3.3

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 contain sufficient information to ensure traceability between counts in tables, listings, and figures to the unique terms in CLSTRESC. For example, if "vomitus, food" and "vomitus, clear" are tabulated separately in the study report, CLSTRESC should be standardized to "vomitus, food" and "vomitus, clear" rather than "vomitus". Differences between the representation in CL and the presentation of Clinical Observations in the Study Report should be mentioned in the NSDRG.

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 numeric or ISO 8601 format
- Grouping of each different set of time series measurements used to calculate a rekated 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.

When a measurement is identified as being above or below a limit or quantitation threshold in PCSTRESC and/or PCLLOQ, standardized units for the threshold should be provided in PCSTRESU.

CONTINUE - UPDATES IN SECTION 4.1.

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

v3.1

3. Whenever --DTC, --STDTC or --ENDTC, which have the role of timing variables, are included, the matching Study Day variables (--DY, --STDY, or --ENDY, respectively) should be included. For example, in most Findings domains, --DTC is Expected, which means that --DY should also be included. In the Subject Visits domain, SVSTDTC is Required and SVENDTC is Expected; therefore, both SVSTDY and SVENDY should be included.

v3.3

3. Whenever --DTC, --STDTC or --ENDTC, which have the role of timing variables, are included, the matching Study Day variables (--DY, --STDY, or --ENDY, respectively) should be included. For example, in most Findings domains, --DTC is Expected, which means that --DY should also be included. In the SDTM Subject Visits domain, SVSTDTC is Required and SVENDTC is Expected; therefore, both SVSTDY and SVENDY should be included.

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 5

5.1 General

v3.1

CDISC Therapeutic Area (TA) standards are comprised of existing data elements, but may introduce new data elements (e.g. domains, variables, terminologies). These data elements are components of current CDISC implementation guides or will be integrated into future implementation guides. CDISC publishes a user guide for each therapeutic area use case which describes the most common data elements for clinical studies (http://www.cdisc.org/therapeutic).

v3.3

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 pertains 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 (http://www.cdisc.org/therapeutic). The CDISC TAUGs should not be interpreted as FDA guidance.

CONTINUE - UPDATES IN SECTION 5

5.2 Supported Therapeutic Areas

v3.1

Generally, when a data standard is released for public use by the SDO, it is not supported by FDA and is not listed in the FDA Data Standards Catalog. FDA performs acceptance testing to determine its ability to support new TA data elements.[32] The CDISC data elements associated with following therapeutic areas are supported by FDA:

5.2.1 Chronic Hepatitis C

5.2.2 Dyslipidemia

5.2.3 Diabetes

5.2.4 QT Studies

5.2.5 Tuberculosis

v3.3

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 Data Standards Catalog). Sponsors should explain the rationale in the cSDRG for using TA extensions that are not currently listed in the Guide.

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

5.2.1 Chronic Hepatitis C

5.2.2 Dyslipidemia

5.2.3 Diabetes

5.2.4 Diabetic Kidney Disease

5.2.5 Ebola

The Ebola Virus Disease (EVD) Therapeutic Area User Guide (TAUG) identified the ISARIC[33] 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 trials. Sponsors should be aware of both components of the ISARIC CORE Dataset when conducting EVD clinical trials.

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 7

7.1 eCTD File Directory Structure

v3.1

All datasets should be referenced in the eCTD XML backbone. Datasets included within the eCTD should be accurately tagged within a study tagging file to ensure proper identification and organization.[49] 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.

v3.3

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.*"[50] 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.

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

CHANGES FROM GUIDE V3.1 TO GUIDE V3.3 (RELEASED MARCH 2017) –UPDATES IN SECTION 8

8.2 Types of Study Data Validation Rules

v3.1

8.2.1 Types of Data Validation Rules

Generally, FDA recognizes two types of validation rules – Conformance and Quality.

8.2.2 Support on Data Validation Rules

The Standards Web page[51] provides links to the currently available validation rules, i.e. both conformance rules and quality checks.

Sponsors should validate their study data before submission using the most recently published validation rules and either correct any validation errors or explain in the Reviewer's Guide (SDRG or ADRG) why certain validation errors could not be corrected. The recommended pre-submission validation step is intended to minimize the presence of validation errors at the time of submission.

v3.3

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 FDA Data Standards 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 is compliant and useful and supports 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 detail regarding FDA's assessment of study data for purposes of review and analysis. The Standards Web page provides links to the currently available business rules and FDA validator rules.[52]

[52] See https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm2005545.htm

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. The FDA validator rules also represent the latest understanding of what bests supports regulatory review. Sponsors should either correct any discrepencies between study data and the standard or the business rules or explain meaningful discrepencies in the Reviewer Guide (i.e., nSDRG, cSDRG or ADRG).

CONTINUE - UPDATES IN SECTION 8

8.3 Study Data Traceability

v3.1

8.3.1 Overview

An important component of a regulatory review is an understanding of the provenance of the data (i.e., traceability of the sponsor's results back to the CRF data). Traceability permits an understanding of the relationships between the analysis results, 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

As noted in section 1.1, the submission of standardized study data will be required according to the timetable specified in the eStudy Data guidance. During the transition period to required study data standards, FDA recognizes that some study data (i.e., legacy data) submissions may not conform to FDA-supported study data standards and may need to be converted.

v3.3

8.3.1 Overview

An important component of a regulatory review is an understanding of the provenance of the data (i.e., 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

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, line 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.

Based upon reviewer experience, establishing traceability …..

....submission to the Agency.

CONTINUE - UPDATES IN SECTION 8

8.3.2 Legacy Study Data Conversion to Standardized Study Data

v3.1

Sponsors should use processes .....

…..subsequently converted to a standard format.

v3.3

Although not strictly a legacy conversion, for nonclinical studies where data is 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, no non-standardized data should be submitted.

THANK YOU

