# Data Submission

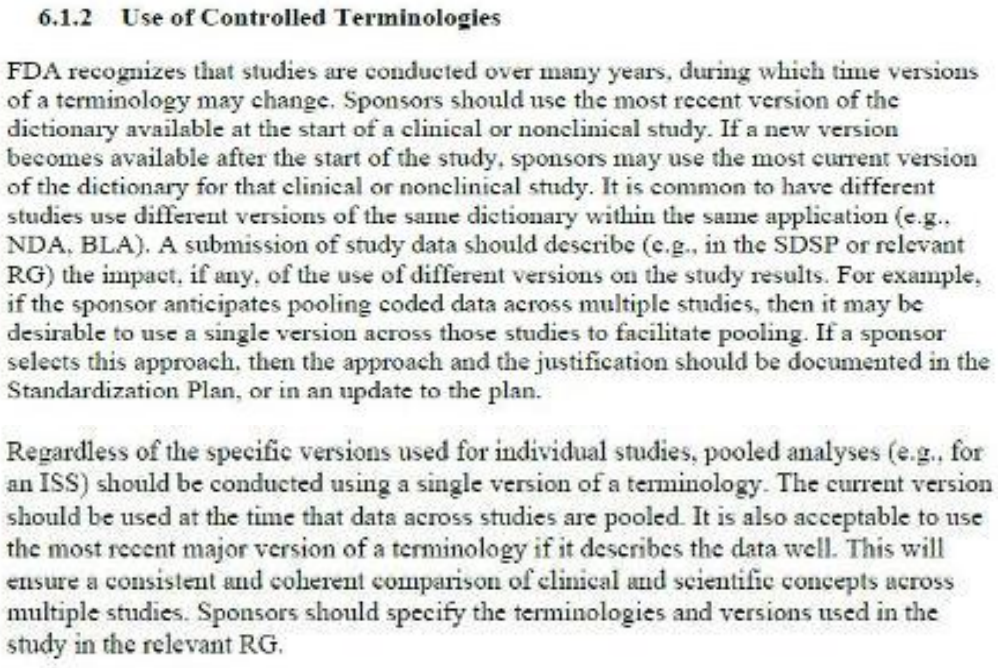
***Question***

Should any upversioning be done when submitting studies that use older versions of SDTM/ADaM IGs?

***PHUSE Team: 25 February 2025***

The sponsor should discuss any requirements for upversioning the submission datasets with the regulatory agency review team early on. In general, Define-XML may need to be upversioned, but not to the latest SDTM/ADaM IGs since the study start date in the TS domain is before 2017. Known issues/non-compliance resulting from running datasets using older versions of the SDTM/ADaM IG on the current compliance rules should be documented into the appropriate sections of the data reviewer's guides.

Upversioning to a common SDTM/ADaM IG and harmonising controlled terminologies may be required for integrated analyses purposes. Methods of upversioning to a common IG and harmonising controlled terminologies should be documented into the appropriate sections of the data reviewer's guides for integration datasets or in the Study Data Standardization Plan (SDSP), as indicated in the FDA Study Data Technical Conformance Guide v5.9, section 6.1.2 (dated October 2024, https://www.fda.gov/media/153632/download): **Click on image to enlarge.**



**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. It is also acceptable to use the most recent major version of a terminology if it describes the data well. 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.

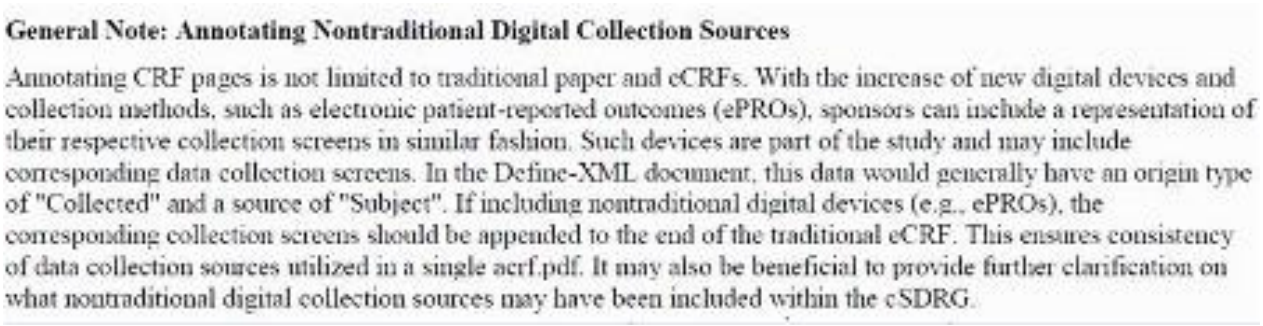
# Data Submission

***Question***

Should eCOA forms be included in the acrf.pdf, in addition to the EDC forms, and submitted as part of eCTD M5? What about central lab and ECG data?

***PHUSE Team: 02 February 2025***

According to the SDTM Metadata Submission Guidelines v2.0, section 3.1.1 (pg 7, https://www.cdisc.org/standards/foundational/sdtm/sdtm-metadata-submission-guidelines-v2-0): **Click on image to enlarge:**



**General Note: Annotating Nontraditional Digital Collection Sources**

Annotating CRF pages is not limited to traditional paper and eCRFs. With the increase of new digital devices and collection methods, such as electronic patient-reported outcomes (ePROs), sponsors can include a representation of their respective collection screens in similar fashion. Such devices are part of the study and may include corresponding data collection screens. In the Define-XML document, this data would generally have an origin type of "Collected" and a source of "Subject". If including nontraditional digital devices (e.g., ePROs), the corresponding collection screens should be appended to the end of the traditional eCRF. This ensures consistency of data collection sources utilized in a single acrf.pdf. It may also be beneficial to provide further clarification on what nontraditional digital collection sources may have been included within the cSDRG.

eCOA forms should be annotated and appended to the end of the traditional eCRF pages. However, central lab and ECG collection information does not need to be formatted on a physical page and annotated. The list of such tests performed in the protocol provides sufficient information on what was collected.

***Question***

I have the following questions related to submitting Define-XML for ADaM datasets to the FDA. Metadata of ADaM datasets is submitted separately from the metadata of SDTM datasets. However, several SDTM variables are copied over to the ADaM datasets for traceability, which initiates following additional questions for these variables:

1. Should we apply the FDA rule for same-named variables per dataset submission package (i.e. separately for ADaM vs SDTM datasets)? The FDA business rule v1.5 is “a variable’s length across a study should be no longer than the maximum length of the actual data (except for SUPPQUAL)”, following the FDA’s Study Data Technical Conformance Guide v4.9 section 3.3.3: “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.”
2. Should variable length in ADaM datasets be the same as the variable length in SDTM datasets, also in the case of different actual data in ADaM datasets (e.g. because only a subset of records was used)?
3. Should the complete SDTM codelist be copied over to the ADaM metadata, even when some values are no longer applicable in the ADaM dataset (e.g. for a --TEST that is not retained in ADaM)?

# Data Submission

**4. Should the order of values in the codelist of SDTM variables in the ADaM Define-XML be identical to that in the SDTM Define-XML?**

***PHUSE Team: 29 January 2025***

The following guidelines have been published by CDISC to help answer these questions:

- SDTM Metadata Submission Guidelines v2.0 (https://www.cdisc.org/standards/foundational/sdtm/sdtm-metadata-submission-guidelines-v2-0)
- ADaM Metadata Submission Guidelines v1.0 (https://www.cdisc.org/standards/foundational/adam/adam-metadata-submission-guidelines-v1-0) Note that the FDA's Study Data Technical Conformance Guide v5.9, published in October 2024, itemises this guidance in Appendix D to "have been evaluated by CBER and CDER and are not considered to align with their current business needs." The conformance guide further states: "Consult with your division for more specific instructions."
- Define-XML Office Hours
  Webinar recording: https://www.cdisc.org/events/webinar/define-xml-office-hours
- PHUSE Define-XML Version 2.0 Completion Guidelines (https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Define-XML+Version+2.0+Completion+Guidelines.pdf)

Valid strategies may be adopted for a successful submission package depending on the uniqueness of the submission. The sponsor should directly contact the FDA's eData team for any specific questions about interpreting the FDA Study Data Technical Conformance Guide (email CBER at mailto:cber-edata@fda.hhs.gov or CDER at cder-edata@fda.hhs.gov).

The following are answers to the itemised questions:

1. There have not been any requirements suggested from CDISC ADaM to harmonise the variable lengths for same-named variables in ADaM. At most, ADSL variables carried into other ADaM datasets may need to be the same variable lengths to avoid merging warnings or setting the attribute from the first dataset.
2. When retaining SDTM variables in ADaM datasets for traceability, the sponsor should keep all attributes – except length – the same. The length can be different depending on any sub-setting being done within the ADaM datasets to satisfy the analysis needs.
3. The CDISC ADaM Metadata Guidelines v1.0, section 2, pg. 5, states that: "No definitive guidance exists as to whether codelists in Define-XML should contain all of the possible values for a given study. For example, suppose only mild adverse events have been reported for a study where the data collection instrument collected AE severity as Mild, Moderate or

Severe. The sponsor could opt to include all three possible terms in the codelist or limit it to only the single term present in the data. In the sample submission, we have chosen to include all possible values in the codelists."

4. There is no specific guidance on the order of the codelist terms/values in Define-XML. Some variables may not have an inherent order and may not add any more value to what is being conveyed in the Define-XML document.

***Question***

I have the following questions on the CDISC Define-XML v2.1.7 guidance, as the examples do not provide full clarity:

1. Can you confirm the dot (.) should not be taken into account when determining the length of datatype=float variables?
2. If so, when including a def:displayFormat, should we apply this as (length+1).significant digits (e.g. length=4, significant digits=2 -> displayFormat=5.2)?
3. Although Define-XML does not limit the number of characters used, is there a recommended maximum length for comments/derivation algorithms beyond which a separate supporting document with 'additional derivations' is recommended? Does the stylesheet have limitations on the maximum length for presentation?
4. Per FDA Study Data Technical Conformance Guide v4.9, section 3.1.5 – "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" – multi-byte characters are allowed (either in derivation algorithms and/or comments but also in variable values).
1. Do you have more insight into submission of data containing multi-byte characters? Even though we have been trying to eliminate any non-basic ASCII characters from source data, this is not always easy. So, if there are (clearly defined) situations where it is allowed, this would be helpful.
2. Should the length of such variables be determined based on the number of bytes or on the number of characters?

***PHUSE Team: 29 January 2025***

Valid strategies may be adopted for a successful submission package depending on the uniqueness of the submission. The sponsor should directly contact the FDA's eData team for any specific questions about interpreting the FDA Study Data Technical Conformance Guide (email CBER at cber-edata@fda.hhs.gov or CDER at cder-edata@fda.hhs.gov).

The following are answers to the itemised questions:

# Data Submission

1. Please refer to the description of the length in the table in the Define-XML v2.1 document, section 4.3.1: "The largest allowable whole number width plus the maximum number of decimal digits."
2. Sponsors should note that the length pertains to the requirement that the receiving system be able to process and store values, whereas the displayFormat pertains to the presentation.
3. There is no recommended maximum length, as this would be sponsor-defined. The sponsor should consider the user experience of the reviewer to ensure readability of the Define-XML document when setting this length.
4. Most sponsor organisations have a mapping file that will map special characters (e.g. foreign languages other than Japanese or Chinese) to the closest ASCII printable values equivalent or, if not found, leave as blank.

a) The sponsor should discuss with the FDA, as it is not recommended to use characters beyond ASCII printable values below 128, per the FDA's Study Data Technical Conformance Guide v4.9, section 3.1.5. As stated in the PHUSE white paper Transport for the Next Generation, the SAS v5 XPT file "only supports US ASCII Character encoding. No multibyte characters are possible; this requires translation and/or transcription from the source data." (Pg 4 of reference: https://www.cdisc.org/sites/default/files/2023-05/Transport-for-the-Next-Generation-Version-1.0.pdf)

b) The definition of text length in the Define-XML v2.1 document is mentioned in section 4.3.1 to be the maximum allowable length of the characters.

***Question***

When is it a good idea to submit an eCTD or standardised data sample to the FDA, since 'they are not reviewed by FDA reviewers at any time', per FDA TCG section 7.3 [eCTD Sample Submission] (https://www.fda.gov/media/153632/download)?

Here are some possible situations:

1. A submission for a new therapeutic area
2. A submission for a new indication
3. A submission for a new product
4. A submission for a product where the data structure has changed

***PHUSE Team: 15 August 2024***

The purpose of submitting a standardised data sample to the FDA is to determine if the gateway submission is successful. As per Submit an eCTD or Standardized Data Sample to FDA | FDA – the document that is footnoted in FDA TCG section 7.3 [eCTD Sample Submission] (https://www.fda.gov/media/153632/download) – it is indicated that: 'The validation

of sample submissions does not involve scientific review of the content and is only intended to address conformance to FDA supported electronic submission and data standards.'
It is probably best to interpret the TCG statement of they are not reviewed by FDA reviewers at any time' as referring to 'scientific review of content', and, without confirming with the FDA directly, a good faith understanding that the FDA could possibly review from the perspective of 'conformance to FDA supported electronic submission and data standards'.

***Question***

How do I interpret the FDA Study Data Technical Conformance Guide section 3.3.2 (Dataset Size) with respect to documentation of submitted datasets in Define-XML? Should I only consider documenting the non-split datasets/domain in the Define-XML; that is, only documenting the datasets present in the datasets folder in Define-XML, while the split datasets that are submitted under the 'split' sub-folder should only be explained in the Data Reviewer's Guide?

***PHUSE Team: 29 August 2023***

There are generally two situations for splitting datasets: due to dataset size constraints or for illustrative purposes.
When splitting due to dataset size constraints, the split is performed after the entire dataset has been created. Section 7.2 of the Study Data Technical Conformance Guide states: "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." There is no need to document the split datasets in Define-XML for submission. These split datasets only need to be documented in the Data Reviewer's Guide.
When splitting data for illustrative purposes, such as the FA or QS domain, the FDA's eData team recommends submitting the split datasets without including the entire dataset. Document only the split datasets in the Define-XML. It is recommended to discuss this illustrative split with the regulatory agency before submission.
Note that the SDTM Metadata Submission Guidance discusses this as well: "The split datasets do not require additional Define-XML documentation." The ADaM Metadata Submission Guidance does not mention split.

***Question***

Suppose a sponsor needs to submit additional datasets (e.g., ADaM) to the FDA for an already submitted study. Should the sponsor submit a Define-XML that includes all the datasets (new and old) or a Define-XML with just the new datasets? Will these new datasets be placed in the same folder as the original datasets? Or will they reside in a separate folder? Should the Define file be

called the same name (i.e. Define-XML) to avoid conflict with the original Define file with old datasets if only new datasets are submitted later?

PHUSE Team Response: 10 June 2023

The Study Data Technical Conformance Guide (TCG) [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document], currently v5.2, updated in May 2023, has two relevant sections that assist in addressing this question.

Section 4.1.1.2 (SDTM General Considerations) indicates the following:
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). When updated datasets (e.g. 'ae.xpt', 'lb.xpt') are submitted, updated and complete Define-XML and cSDRG covering all datasets should be submitted using the "replace" life cycle operator to update the original file.

Section 4.1.2.2 (ADaM General Considerations) indicates the following:
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).
Unlike SDTM, where it is always required to maintain the same SDTM dataset names, ADaM datasets that are resubmitted can have new dataset names that differ from the original submission.
Regarding the Define-XML specifically, the TCG wording can be interpreted in at least two ways:
1) Each time new information is submitted, a new Define-XML is needed.
2) Only the first Define-XML is considered in the submission, so anything further requested is just supplementing what is already there.

Sponsor experience with resubmitting ADaM datasets has varied. In some cases, after ADaM datasets were initially accepted by the FDA and then more data were requested, a separate Define-XML, along with the new ADaM datasets, was provided. In the experience of some sponsors, new ADaM datasets have been provided to the FDA, with or without Define-XML, depending on the timeframe and the speed of the response. If the FDA has accepted ADaM datasets and requests additional data, a Define-XML may be prepared with new data only in a separate directory. The Define-XML may not need to be renamed. Sometimes, the same ADaM dataset names can be used. If there are challenges in set-up or other considerations, different ADaM dataset names can be used as long as they are clearly explained in the response letter. Sometimes, the FDA accepts additional ADaM datasets without a Define-XML; however, this is an exception. When resubmitting subsets of ADaM data, some sponsors have found success by only including updated datasets in Define-XML to avoid broken links.

# Data Submission

There is no one-size-fits-all solution for resubmitting an ADaM dataset, as it depends on the specific circumstances, so discussion with the FDA review division is essential.
The location of resubmissions within m5 of the eCTD is the responsibility of the regulatory publishing team members.

***Question***

Do sponsors who are considering submissions for oncology studies under RTOR (Real Time Oncology Review) need to adhere to the OOD Safety Data Specifications that are mentioned in this RTOR guidance? It seems that these specifications were still under review by CDISC. What is the current status of the review? Even if it is under review, what is the FDA's position on adhering to these specifications for upcoming submissions under RTOR?

***PHUSE Team Response: 25 April 2023***

According to the FDA (email dated April 11, 2023), OCE/OOD safety data specifications are intended to serve as a good practice recommendation and are not considered mandatory at this time. However, the FDA has sought feedback and is planning to release an updated version shortly. Flexibility is provided for sponsors on how the ADaM datasets are submitted under RTOR, while still adhering to ADaMIG v1.1 CDISC guidance. For formal advice on RTOR data submission for a study under consideration, the sponsor organisation should submit the questions in a meeting package to the corresponding review division.

***Question***

How should a sponsor handle requests from the FDA related to the creation of a table containing outputs in the pivotal trials of a submission, including hyperlinks to the code used to create the table and the primary ADaMs used? (See the example below.) This request is a) different from the ARM deliverable embedded in the Define-XML that we create as part of a submission and b) not specified in any FDA guidance to date. **Click image to enlarge:**

| Output Number | Title | CSR Location | Program and/or Macro | Input Dataset(s) |
|---|---|---|---|---|
| 15.2.1-1.1 | Duration of efficacy observation period | | T_EXP_02.sas | ADCM |
| 15.2.1-2.1 | Study treatment compliance (based on eDiary) | | T_EXP_03.sas | ADCM |
| 15.2.1-2.2 | Study treatment compliance (based on eCRF) | | T_EXP_03.sas | ADCM |
| 15.2.2-1.1 | Hypotheses testing of the primary and secondary efficacy endpoints at Month 6 | | T_TESTST.sas | ADTTE |
| 15.2.2-10.1 | Empirical probability density function (ePDF) of absolute change from baseline to Month 6 in neuropathic pain monthly score by PGIC-PS score at Month 6 | | MF_PLOT_EPDF.sas | ADQSD |
| 15.2.2-10.2 | Empirical probability density function (ePDF) of percent change from baseline to Month 6 in neuropathic pain monthly score by PGIC-PS score at Month 6 | | MF_PLOT_EPDF.sas | ADQSD |
| 15.2.2-10.3 | Empirical probability density function (ePDF) of absolute change from baseline to Month 6 in neuropathic pain monthly score by PGIC-DS score at Month 6 | | MF_PLOT_EPDF.sas | ADQSD |
| 15.2.2-10.4 | Empirical probability density function (ePDF) of percent change from baseline to Month 6 in neuropathic pain monthly score by PGIC-DS score at Month 6 | | MF_PLOT_EPDF.sas | ADQSD |
| 15.2.2-10.5 | Empirical probability density function (ePDF) of absolute change from baseline to Month 6 in neuropathic pain monthly score by PGIS-P response categories | | MF_PLOT_EPDF.sas | ADQSD |
| 15.2.2-10.6 | Empirical probability density function (ePDF) of percent change from baseline to Month 6 in neuropathic pain monthly score by PGIS-P response categories | | MF_PLOT_EPDF.sas | ADQSD |
| 15.2.2-10.7 | Empirical probability density function (ePDF) of absolute change from baseline to Month 6 in neuropathic pain monthly score by PGIS-D response categories | | MF_PLOT_EPDF.sas | ADQSD |
| 15.2.2-10.8 | Empirical probability density function (ePDF) of percent change from baseline to Month 6 in neuropathic pain monthly score by PGIS-D response categories | | MF_PLOT_EPDF.sas | ADQSD |

**PHUSE Team Response: 28 February 2023**

Although this has not been a commonly experienced situation, the response and method of providing the information may be tailored to when it is requested by the regulatory agency.

Suppose the request is made by the regulatory agency ahead of the submission. In that case, the sponsor may utilise the Analysis Data Reviewer's Guide (ADRG) section 7.2 and include the additional columns and hyperlinks for Input Data and Program File. For requests by the regulatory agency after submission, the sponsor may consider providing a document containing the list of program files, the output number and title produced by each program file, and a description of each, where necessary. Refer to the Study Data Technical Conformance Guide for a list of valid file types that can be included in the submission to the FDA.
Sponsors should discuss the need to provide executable code with the regulatory agency ahead of providing the program files.

***Question***

When you submit a custom ADaM dataset like ADAE2 (in addition to ADAE), how do you validate it before your submission?

***PHUSE Team Response: 25 August 2022***

Sponsors can include multiple ADAE datasets in a study, as there is no guidance dissuading the use of multiple analysis datasets for adverse events. The use of multiple ADAE datasets will also not result in any Pinnacle 21 finding. However, it is possible for a reviewer from a given regulatory agency to expect a single ADAE dataset; in this case, the sponsor may be required to prepare an overall dataset.
One method of combining all adverse events into a single dataset may be the use of ACATy or the use of multiple TEAE flags corresponding to each treatment drug in addition to the primary TEAE flag (refer to the ADaM Structure for Occurrence Data (OCCDS) Implementation Guide v1.1). Additionally, AOCCFL can be utilised in combination with TRTEMFL, and TRTEMwFL and AOCCzzFL can be created for the drug w.

***Question***

What type of information about a subject with multiple screenings must be submitted to the FDA?
The FDA Study Data Technical Conformance Guide (v4.8.1, October 2021) includes the following description regarding a subject with multiple screenings:
4.1.1.2:
“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 survey 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."

4.1.1.3:

"DM Domain (Demographics)

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

Screen failures, when provided, should be recorded in DM with the ARM, ARMCD, ACTARM, and ACTARMCD fields left blank. For subjects randomised to a treatment group but not treated, the planned arm variables (ARM and ARMCD) should be populated; however, the actual treatment arm variables (ACTARM and ACTARMCD) should be left blank.

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 the DM, with additional screenings in a custom domain that has a structure similar to the 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."

***PHUSE Team Response: 11 March 2022***

While the FDA's Study Data Technical Conformance Guide (v4.8.1, October 2021) sections 4.1.1.2 and 4.1.1.3 specifically mention how to store data related to subjects with multiple screenings or enrollments in the SDTM domains, sponsors may vary in their approach to identifying records associated with each screening or enrollment attempt.

The Multiple Subject Instances Team at CDISC is currently working on creating the new DC domain that will address this issue. For each USUBJID, this domain will contain multiple entries for each time the subject was screened or enrolled in the study. The SUBJID value will reflect the subject identifier for that time of participation. The Multiple Subject Instances Team at CDISC also recommends including the SUBJID in the parent domain as a permissible variable, following the FDA Study Data Technical Conformance Guide, Section 4.1.1.2. The new DC domain and this recommendation will be included in the future release of SDTM IG v4.0.

The SDTM ADaM Implementation FAQ provided the following examples of how some sponsors are capturing multiple subject instances using the current SDTM domains and variables. Your company may choose to represent this data differently. It is recommended that any concerns about the data capture process be discussed with the FDA before submission to ensure that they will not result in a denial at the time of submission.

# Data Submission

Some sponsors may set the value of SUBJID to be consistent within each domain, matching it to some portion of the value of USUBJID, and use the --REFID variables to map each screening or enrolment subject identifier. The use of --REFID for this purpose should be mentioned in the cSDRG.
Instead of generating a custom domain to capture additional screening identifiers or enrollment identifiers (mentioned in Section 4.1.1.3 of the FDA Study Data Technical Conformance Guide), sponsors may capture the additional identifiers in the supplemental qualifier of the DM domain (SUPPDM). The extra data stored in SUPPDM should be described in the cSDRG.
Some sponsors have already added the SUBJID into the parent SDTM domains, following the FDA Technical Conformance Guide section 4.1.1.2. Some sponsors have defined the SVSTDTC and SVENDTC variables in the SV domain to represent the start and end dates of the first attempt at screening or enrollment, respectively, and the last effort at screening or enrollment. Similarly, the SESTDTC variable in the SE domain can be defined as the date of the first informed consent.

***Question***

Does clinical trial data need to be submitted at the time of an IND submission? If necessary, does a full data package need to be submitted, including all SDTM domains, ADaM datasets, reviewer's guides, and Define files with complete validation?

***PHUSE Team Response: 11 March 2022***

While clinical trial data and the complete data package are not required to be submitted at the time of an IND, it would provide the FDA with an early look at the clinical data and the ADaM datasets that the sponsor is considering for the study. The data submitted may include any previous experience with the study drug in humans (often from foreign use). It is always good to discuss the need for submitting data at the time of the IND with the review division at the agency at the Pre-IND meeting. The review division can then decide whether to receive and review the data.

Question

At one point, a joint CDISC/FDA team worked on defining locations in SDTM/ADaM for BIMO components, allowing the CLINSITE information to be extracted from the submitted data instead of a separate dataset. This joint effort has been put on hold for the time being. However, at this point, the team recommends continuing to reference the current BIORESEARCH MONITORING TECHNICAL CONFORMANCE GUIDE.
(https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioresearch-monitoring-technical-conformance-guide https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioresearch-monitoring-technical-conformance-guide).
And Standardised Format for Electronic Submission of NDA and BLA Content for the Planning of Bioresearch Monitoring (BIMO) Inspections for CDER Submissions.

# Data Submission

(https://www.fda.gov/regulatory-information/search-fda-guidance-documents/standardized-format-electronic-submission-nda-and-bla-content-planning-bioresearch-monitoring-bimo) for details.

***PHUSE Team Response: 13 August 2020***

The team has previously provided its response to the question on "Requirements for Clinical Site Data and Subject Level Data Listings for FDA CDER's Inspection Process (also called BIMO submission or OSI Pre-NDA request)".

BIMO

***Question***

These questions are primarily going out to the sub-team that worked on the Best Practices for Submission of Event Adjudication Data White Paper. The White Paper provided handy tips on how to map adjudicated data to the new custom SDTM domain of EA. The following are the follow-up questions to this White Paper.

Are there any plans to include the EA domain in future CDISC SDTM IG releases? If so, which IG is this being targeted for? Is it ok to assume that sponsors can submit this as a custom domain to regulatory agencies until then?

***PHUSE Team Response: 14 July 2020***

CDISC SDS informed that the adjudication project is under consideration and may be in the future SDTMIG (beyond SDTMIG v3.4). Submitting EA as a custom domain is allowed by the current SDTMIG. The proposed domain in the White Paper is based on previous submission experiences and can be used for submission until the CDISC publishes a new domain.

The White Paper did not get into any suggestions on how to map this into ADaM. This may be intentional, as it may depend on the nature of the analysis surrounding adjudicated data or even the type of adjudicated data itself. Is there any general recommendation you can make?

For ADaM, a statistical/reporting analysis plan determines which data should be included in the analysis datasets and how the data are used for reporting and associated analyses. An example is not included in the White Paper because, in general, only the final adjudication assessment is included in the ADaM dataset. However, an example of how to capture final evaluations in the EA domain is provided in the White Paper.

***Question***

Is exposure data from the parent study required to be in the SDTM data of a follow-up study (no treatment given in follow-up study)? Is it needed for the FDA and PMDA? Can the exposure data be carried over from the parent study SDTM into the follow-up study SDTM data, or does it need to be re-collected on the CRF of the follow-up study?

# Data Submission

***PHUSE Team Response: 9 January 2020***

In general, if data is not collected on the CRF for the follow-up study, it is not recommended to report it into SDTM datasets. In this example, we recommend not carrying it over to SDTM for the follow-up study. Instead, this information can be presented in the analysis dataset.

***Question***

Is there a Standard in the industry for determining the study start date for clinical studies? Is it the finalised protocol date, first subject in date, or first initiation date?

***PHUSE Team Response: 22 November 2018***

As per the guidance from the FDA - Providing Regulatory Submission in Electronic Format - Standardised Study Data, 'the study start date for clinical studies is the earliest date of informed consent among any subject that enrolled in the study'. For example, see Study Start Date in the SDTM Trial Summary Domain (TSPARMCD = SSTDTC). For nonclinical studies, the study start date is the date on which the study protocol or plan is approved (signed) by the Study Director, also known as the study initiation date. For example, see Study Start Date in the SEND Trial Summary Domain (TSPARMCD = STSTDTC). This definition is consistent with the Study Data Standardised Plan (SDSP) PHUSE template, which has been reviewed and authorised for use.

Additional References:

FDA Guidance, "Providing Regulatory Submissions In Electronic Format —Standardised Study Data" https://www.fda.gov/downloads/Drugs/Guidances/UCM292334.pdf

Study Data Standardised Plan PHUSE template

***Question:***

What goes in the 'misc' folder with an m5 eCTD folder structure? For example, a lookup file containing the SMQ assignment. The file is used during the creation of pooled ADaM to support an ISS. We want to provide this dataset to the reviewer. This does not contain the subject's data, and it is not SDTM or ADaM. Should this be placed in the 'misc' folder? Or to the analysis folder and described in the define.xml, and classified as non-ADaM? Or is it enough to explain its structure in the ADRG?

***PHUSE Team Response: 12 April 2018***

According to the FDA Study Data Technical Conformance Guide (version 4.0), Section 7, which describes the Electronic Submission Format, the misc folder should "contain datasets that do not qualify as analysis, profile, or tabulation datasets in this subfolder." These datasets should be in SAS Transport Format (.xpt). Since these datasets do not qualify as analysis, profile, or tabulation, they

do not need to be included in the define.xml; however, information about the use of these datasets should be included in the reviewer's guide.
If you have other documents/files that support the creation of your datasets, be they analysis or tabulation, or your TLGs, such as a spreadsheet for CTC Toxicity Grade or SMQ assignment, you can convert them to an acceptable format (e.g. PDF, TXT, or XPT) and place these in the "misc" folder. The file name must be in all lowercase letters or numbers with no spaces or special characters; only a hyphen is allowed in the name. Conventions for file names can be found in the Technical Requirements for the Registration of Pharmaceuticals for Human Use.
Name, page 11-12. Information about these additional files and their use in creating the datasets should be included in the reviewer's guides.

Additional References:
Electronic Common Technical Document Specification
Technical Conformance Guide 2018

***Question***

Does the legacy data in non-CDISC format need to be converted to SDTM for all studies included in FDA or PMDA submissions? If a sponsor has one pivotal study in non-CDISC format and the other pivotal study in CDISC format, do I need to convert both to CDISC format before submission?

***PHUSE Team Response: 7 June 2017***

**FDA:**

The study submitted electronically must be in CDISC format if the study start date is after December 17, 2016.
If ALL studies included in the NDA started after the mandate date and data are collected in legacy format, then yes, the conversion from legacy to SDTM is required.
If all studies included in the NDA started before and do not meet the CDISC mandate date of December 17, 2016, then it is still acceptable to submit the data in legacy (non-CDISC) format.
In addition to the CDISC mandate above, if there is no consistent data format across all studies, e.g., the pivotal studies are in SDTM format. Still, the supporting studies are in legacy format. The data contents and formats are proposed in the briefing package (BP) before the meeting with the FDA. The reviewers either agree with your proposal or request changes to its content and format.
The Study Data Standardisation Plan, which can be shared as early as the pre-IND stage and is recommended before the EOP2, is a means of communicating proposed study standards for nonclinical and clinical studies within an IND/indication to the FDA. This is the opportunity to agree upon study standards early in the development of a compound.

# Data Submission

**The SDSP:**

It is used to establish and document a plan for describing the data standardisation approach for planned studies within a specific submission in the development program.

Contains information about the intended and/or current state of data standards that are being used for studies within a compound.

It is used as a communication tool with the FDA or other Health Authorities to ensure that the reviewers understand the data standards that the sponsor is using for each study.

It is recommended to be included as part of a regulatory submission to Health Authorities.

**PMDA:**

Since October 2016, the PMDA has accepted submissions in CDISC format. Until March 2020, a transition period will be in effect during which the PMDA accepts legacy submissions and partial data submissions (hybrid submissions). Sponsors need to have a special consultation meeting (consultation on data format of submission of electronic study data) with PMDA when deciding to submit the electronic datasets (approximately one year before the submission - it will be the timing of the decision of the submission package) to agree on the electronic datasets format for NDA submission.

From April 2020, all required study data must be submitted in CDISC format, regardless of when the study started. Studies in legacy format will need to be converted; No waivers are allowed on this point. Closed or completed studies will require data conversion if the survey meets eStudy data submission criteria as described in the Basic Principles on Electronic Submission of Study Data for NDA's (binding document):

Target studies (Phase I and CP studies, Phase 2-3 studies) will be those classified as evaluation studies in the submission package.

If a Phase I or CP study is used as an evaluation study and falls into one of the following categories, then electronic data is always required.

Phase I studies of oncology drugs.

Phase I studies were conducted on both Japanese and non-Japanese subjects (e.g., global clinical trials and bridging studies).

QT/QTc studies based on the ICH E14 guideline.

For other Phase I and CP studies that don't meet the above criteria, electronic data are required when PMDA needs them for their review. The study types will include those where standard PK analysis is conducted, as well as Population PK and PBPK.

Additional References:

# Data Submission

**FDA Binding Documents:**

Providing Regulatory Submissions in Electronic Format

Section 745A(a) of the Federal Food, Drug, and Cosmetic Act, Guidance for Industry

Standardised Study Data, Guidance for Industry

FDA Non-binding documents and other resources can also be found on the FDA webpage for Study Data Standard Resources.

**PMDA Binding Documents:**

Basic Principles on Electronic Submission of Study Data for New Drug Applications

Q&A Guide "Basic Principles on Electronic Submission of Study Data for NDA's"

PMDA Non-Binding documents and other resources can also be found on the PMDA website for Advanced Review with the Electronic Data Promotion Group.

***Question***

How do I make a test submission to the FDA?

**PHUSE Team Response: 7 June 2017**

The FDA provides a dedicated website page on how to submit an eCTD or standardised data Sample to the FDA – see reference below. The page provides recommendations and steps to submit a sample.

**Additional References:**

Submit an eCTD or Standardised Data Sample to the FDA.

***Question***

Will JumpStart (DataFit) services be available for Pharma clients before submission? What kind of checks are included in JumpStart?

***PHUSE Team Response: 7 June 2017***

JumpStart as a Service is specific to the FDA. There are multiple versions of open-source validator tools available for use that are similar to the version of DataFit that the FDA uses. The use of a validator to check for compliance issues and the inclusion of a Study Data Reviewer's Guide will enable a Sponsor to provide close to all the information the FDA considers during the data fitness portion of a JumpStart service. Standard demographic analysis panels are available via the GitHub code repository. The FDA will be sharing more scripts in the near future.

Additional References:

https://github.com/phuse-org/phuse-scripts

# Data Submission

***Question***

When will CDISC (SDTM/ADaM) data standards be mandatory for data submission, and how does this differ from each regulatory agency?

*PHUSE Team Response: 12 September 2017*

The data standards requirements may differ from country to country, and each regulatory body will have its own set of requirements. Below you will find basic available information from the US (FDA), Japan (PMDA), and other countries that may or may not require dataset submission at this point.

US (FDA): The CDER and CBER strongly encourage IND sponsors and NDA applicants to consider implementing and using study data standards as early as possible in the product development life cycle, ensuring that data standards are accounted for in the design, conduct, and analysis of studies.

Sponsors whose studies start after December 17, 2016, must submit data in the data formats supported by the FDA and listed in the FDA Data Standards Catalogue. This applies to NDAs, BLAs, ANDAs, and subsequent submissions to these types of applications.

For INDs, the requirement applies to studies that start after December 17, 2017.

Beginning after the dates specified above, the FDA may refuse to file for NDAs and BLAs or refuse to receive for ANDAs any electronic submission whose study data do not conform to the required standards specified in the FDA Data Standards Catalogue.

See the Technical Rejection Criteria for Study Data (PDF - 87 KB).

FDA Submission Type & Timing

NDA, ANDA, and specific BLA submissions - Studies which start after 2016-12-17 (December 17, 2016).

Commercial INDs and amendments, except for submissions described in section 561 of the Federal Food, Drug, and Cosmetic Act - Studies which start after 2017-12-17 (December 17, 2017).

For the definition of "study start date," see the Providing Regulatory Submissions in Electronic Format - Standardised Study Data (PDF - 131 KB).

**Source for FDA:**

-*US FDA Website* on Study Data for Submission to CDER and CBER

-*US FDA Website* on Study Data Standards Resource

Additional reference documents/webinar for FDA:

Study Data Standards in eCTD: What You Need to Know About the New Technical Rejection Criteria, October 12, 2016: eCTD Study Data Standards Webinar.

Japan (PMDA): Electronic data submission begins on 01-Oct-2016, with a transition period as noted below.

PMDA Submission Type & Timing

# Data Submission

NDAs (*eStudy Data submission criteria**) - Transition Period - Submission on or after 2016-10-01 (October 1 2016) until 2020-03-31 (March 31, 2020).
All NDAs (*eStudy Data submission criteria**) - Submission on or after 2020-04-01 (April 1, 2020).
During the transition period, the PMDA accepts legacy submissions and partial data submissions (hybrid submissions).
The Sponsor needs to have a special consultation meeting (consultation on the data format for the submission of electronic study data) with PMDA one year before submission (this will be the timing of the decision on the submission package).
During that meeting, the Sponsor must have an agreement with PMDA regarding the electronic datasets for NDA submission.
All required study data need to be submitted in CDISC format after April 1, 2020.
No waivers are allowed after this date. All clinical study data meeting the eStudy submission criteria* must be compliant with the CDISC standard format for submissions on or after April 11, 2020. Therefore, closed or completed studies in legacy format need to be converted.
**eStudy Data submission criteria*: electronic data in CDISC format needed for studies meeting the following criteria.
Target studies (Phase I and Clinical Pharmacology studies, Phase 2-3 studies) will be those classified as an evaluation study in the submission package.
If a Phase I or Clinical Pharmacology study is used as an evaluation study and falls into one of the following categories, then electronic data is always required.
Phase I studies of oncology drugs.
Phase I studies conducted on both Japanese and non-Japanese subjects (e.g, global clinical trials and bridging studies).
QT/QTc studies based on the ICH E14 guideline.
For other phase I and Clinical Pharmacology studies that don't meet the above criteria, electronic data is required when PMDA needs it for their review. The study types will be those where standard PK analysis is conducted, Population PK, and PBPK.
Supported standard and versions data standard catalogue and validation rules, including rejection criteria, are available together with applicable guidance on the PMDA Advance Review with Electronic Data promotion group website.

**Source for PMDA:**

-PMDA Technical Notification for Electronic Data Submissions.
-PMDA Website for Advanced Review with the Electronic Data Promotion Group.

# Data Submission

The response below for other regulatory agencies was compiled on June 28, 2017; the regulation may have been updated since then. We strongly suggest checking each regulatory website for the most current information.

-Other countries like Europe and China recommend the use of CDISC data standards and define.xml following the FDA requirements, but they do not mandate it yet.

As far as the European Medicines Agency (EMA) goes, the Clinical Trial Advisory Group on clinical trial data formats (CTAG2) is working on advising the EMA on clinical data formats, where it is leaning toward CDISC standards (although if it accepts, it would likely follow a similar progression as the FDA, with a 2-3 year pilot. CTAG2 provided the EMA with recommendations to utilise CDISC (SDTM/ADaM) and define.xml in accordance with FDA requirements.

Advice to the European Medicines Agency on Clinical Trial Data Formats. (30APR2013)

Source from EMA:

EMA Website page on Documents from advisory groups on clinical-trial-data

China Food and Drug Administration (CDFA) has endorsed CDISC standards in their Clinical Trial Data Management Technology Guide* (July 2016): they mention "CDISC standards have seen more and more recognition and are widely used in the industry, and have become an international clinical trial data "common language".

Although not yet mandatory in every country, CDISC data standards have operational use, such as transferring data between organisations and sponsor warehousing, making it a good idea to produce CDISC-compliant datasets, even if not technically required for submission. This also allows the creation of a single, unique package with minimal or minor updates for submission in various countries.

*English translation of the Clinical Trial Data Management Technology Guide is not available on the CFDA website. The CDISC website has its own English translation of the document.

**Source for CFDA:**

CFDA Website (Chinese)

CFDA Website (English)