PHUSE EU Connect 2024

Paper SA06

Harmonising FDA and PMDA requirements

Luca Semeraro, Alira Health, Bologna, Italy

ABSTRACT

Ensuring compliance with the electronic submission guidelines set forth by the Food and Drug Administration (FDA) and the Pharmaceuticals and Medical Devices Agency (PMDA) for data submitted in a standardized CDISC format has become increasingly crucial. Currently, the task of submitting identical electronic data packages to both agencies poses certain challenges.

However, understanding these regulatory requirements is essential for effectively planning, developing, and submitting study data deliverables. To shed light on this complex landscape, the paper highlights both the similarities and differences between the FDA and PMDA guidelines for CDISC-standardized data submission. Following this comparative overview, the paper offers a detailed analysis of specific PMDA requirements.

Additionally, the paper includes examples from two case studies that demonstrate the practical application of these guidelines and the challenges faced during package preparation. This approach aims to equip stakeholders with a comprehensive understanding of the regulatory landscape, facilitating smoother and more compliant submission to both regulatory agencies.

INTRODUCTION

In the end-to-end drug development process, understanding regulatory requirements is essential for planning, developing, and submitting study data deliverables. The FDA's requirement to submit data in the Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) formats became mandatory on December 17th, 2016, for all studies initiated after that date. Similarly, the PMDA's mandate for standardized data submission took effect on April 1st, 2020, regardless of the study's initiation date.

Although both agencies require electronic study data submissions in CDISC SDTM and ADaM formats, submitting the same electronic data package to both presents significant challenges. Notably, there are key differences in the data standards that must be addressed before submission to each agency. This paper focuses on comparing the FDA and PMDA guidelines and requirements for submitting standardized CDISC data.

In April 2023, while we were finalizing the submission packages of a study for FDA submission and approaching the database lock, the sponsor requested the adaptation of these packages to comply with PMDA guidelines for submission to that agency as well. A few months later, the same sponsor made the same request for another study. This paper provides examples from these two case studies to illustrate the differences and challenges encountered when adapting the above packages, highlighting how specific issues were addressed. The paper aims to provide an approach for sponsors who plan to submit a study to both regulatory authorities.

ABBREVIATIONS, ACRONYMS AND DEFINITIONS

Term	Description
aCRF	Annotated Case Report Form
ADaM	Analysis Data Model
ADRG	Analysis Data Reviewer's Guide
ANDA	Abbreviated New Drug Application
ARM	Analysis Results Metadata
ASCII	American Standard Code for Information Interchange
BLA	Biologics License Applications
CBER	Center for Biologics Evaluation and Research
CDER	Center for Drug Evaluation and Research
CDISC	Clinical Data Interchange Standards Consortium
cSDRG	Clinical Study Data Reviewer's Guide
FDA	United States Food and Drug Administration
IND	Investigational New Drug applications
LOINC	Logical Observation Identifiers Names and Codes
MedDRA	Medical Dictionary for Regulatory Activities
NDA	New Drug Application
PMDA	Japan Pharmaceuticals and Medical Devices Agency
SDSP	Study Data Standardization Plan
SDTCG	Study Data Technical Conformance Guide
SDTM	Study Data Tabulation Model

Term	Description
SNOMED CT	Systematized Nomenclature of Medicine – Clinical Terms
UNII	Unique Ingredient Identifiers
XML	eXtensible Markup Language

ABOUT PMDA

The PMDA was established and began operations on April 1st, 2004, as part of the Reorganization and Rationalization Plan for Special Public Corporations.[4]

The mission of the PMDA is to enhance public health in Japan, and it is focused on three activities:

- Relief Services for Adverse Health Effects: providing relief to those affected by adverse drug reactions or infections from biological products, and health benefits for patients such as SMON, HIV-positive and AIDS patients
- Product Reviews: offering guidance and conducting reviews on the quality, efficacy and safety of drugs and medical devices through an integrated system that spans from pre-clinical research to approval, and developing standards
- Post-marketing Safety Measures: collecting, analyzing, and disseminating post-market safety information.

PMDA has worked to accelerate its operations by improving its review system and consultation activities. In recent years, PMDA has achieved the shortest review times for new drugs among regulatory agencies in developed countries, including those in Europe and the United States.[4]

COMPARISON OF FDA AND PMDA REQUIREMENTS

Electronic submissions of standardized clinical and nonclinical study data adhering to CDISC format is required from FDA for all studies starting on or after December 17th, 2016 for New Drug Applications (NDAs), Abbreviated New Drug Application (ANDAs) and for Biologics License Applications (BLAs), and for those starting on or after December 17th, 2017 for commercial Investigational New Drug applications (INDs). This guideline is not applicable to noncommercial INDs and to INDs for devices that are regulated by Center for Biologics Evaluation and Research (CBER) as biological products. This requirement extends to all subsequent submissions, including amendments, supplements and reports associated with any of the submission types mentioned above. Waivers may be granted for studies that began before December 17th, 2016. [2, 3]

The same requirement became effective for PMDA as of April 1st, 2020, irrespective of study start date. However, PMDA allows sponsors to submit e-study data in legacy format for studies in orphan drug started before April 1st, 2020.[9] Other exceptions may be possible and may consist of submitting the electronic study data in legacy format or not submitting the electronic study data itself. So, it is recommended to consult with the PMDA in advance for:

- Anti-HIV drugs which are eligible for prior assessments based on the Pharmaceutical and Medical Safety Bureau (PMSB)/ Evaluation and Licensing Division (ELD) Notification No. 1015 (dated 12-November-1998)
- Phase I studies, with the exception of those of oncology drugs, started before April 1st, 2020
- Studies in “public knowledge-based applications”
- Old studies that are considered extremely difficult to convert to CDISC format.

Electronic submission for studies other than clinical studies (e.g. nonclinical studies) is currently being discussed. For that reason, this paper does not cover CDISC Standard for Exchange of Nonclinical Data (SEND).

In this section the paper seeks to outline the key similarities and differences between the requirements of the two regulatory agencies. It begins by examining the planning and standards applied, then explores the submitted datasets and their documentation, and concludes with an analysis of the submission file preparation, with particular emphasis on the validation phase.

STUDY DATA PLANNING

According to the FDA Study Data Technical Conformance Guide (SDTCG)[1], for both clinical and nonclinical studies, sponsors should provide the Study Data Standardization Plan (SDSP) during the early stages of the product development under the IND. It is an important communication tool between the applicant and the FDA that should be provided with pre-NDA and pre-BLA meetings, and it is advisable to initiate it at the pre-IND stage of product development. The sponsor should include the study details and the version of the standards used. It helps the FDA identify potential data standardization challenges early in the development process. As the development program progresses and new studies are planned, the SDSP should be revised and included in ongoing communications with the FDA.

Differently from FDA, PMDA requires to submit “Explanation of Electronic Study Data Form A (Attachment 8-1)” and “Explanation of Electronic Study Data Form B (attachment 8-2)” for consultations related to submission of electronic study data for new drug applications. These forms describe the contents of electronic study data planned to be submitted to the PMDA, and the finalized documents are required for the pre-NDA meeting.

Form A includes an “Overview of clinical data package” (Section 3), which provides a summary of individual clinical studies (study regions, population, design, treatment, endpoints, etc.), and “Information on CDISC-conformant clinical studies, integrated analyses and clinical pharmacology analyses for which electronic data are planned to be submitted” (Section 4), detailing the CDISC standards used, as well as the datasets, programs and documents included in the submission. It should be included with the items for “Consultation on data format of submission of electronic study data” and “Consultation on preparation of submission of electronic study data”.[7]

Form B, instead, gathers information regarding studies for which exemptions are being requested from the agency. Specifically, it includes details on the "Clinical studies and analyses for which the applicant desire to be exempt to submit electronic study data" (Section 3.2.3), on the process from data acquisition to the current status of holding data, and the content of data including compliance with data standards (Section 4), and "Information on the clinical studies, integrated analyses, and clinical pharmacology analyses for which electronic data are expected to be submitted in a format other than CDISC Standards" (Section 5). It should be attached to the item of "Consultation on exemption of submission of electronic study data".[8]

Both documents should be filled with information either completed or planned at the timing of consultation.

Planning is therefore a key factor in submission preparation, and CROs should guide their clients in developing this plan, for example by assisting them in completing Form A for PMDA before starting SDTM/ADaM programming activities and discussing it with the agency during initial meetings. Currently, the sponsors are sometimes not aware that they should do this document, and just complete it prior the submission, which can lead to additional re-work, costs, and approval delays, if PMDA doesn't agree with the data structure and ask for changes. In our experience, the sponsor asked us to fill these forms but we had to inform the client that this should be done by the regulatory team which prepares the submission, as only a few sections concern programming standards and activities.

DATA STANDARD CATALOGS

Both the FDA and PMDA require all electronic submissions to comply with the Data Standard Catalogs issued by each agency at the time of study start. Applicants are advised to use the latest version listed in the catalog when preparing the data, but they can select the version to use for their study when multiple versions of a standard are listed. The FDA may reject NDAs and BLAs or decline to accept ANDAs if the electronic submissions' study data fail to meet the required standards outlined in its data standards catalog.

The FDA updates its Data Standard Catalog at least every six months, whereas PMDA does not update them regularly. The latest versions of the catalogs were released on the following dates:

- FDA: September 16th, 2024
- PMDA: March 29th, 2024

The following table compares these standards, highlighting that the PMDA currently requires fewer standard terminologies than the FDA.

Standard Terminology	Use	FDA [2]	PMDA [6]
CDISC Controlled Terminology	General Clinical Data	▪ Versions 2011-06-10 or later are currently required. ▪ Previous versions cannot be used for studies initiated after June 13th, 2011.	▪ Versions 2011-06-10 or later are currently supported. ▪ Previous versions no longer supported since June 30th, 2017: PMDA need to be consulted when using these versions.
MedDRA (Medical Dictionary for Regulatory Activities)	Adverse Events and Medical history	▪ The current version is required for study starts on or after March 15th, 2019, and for certain INDs as of March 15th, 2020. ▪ Version 8.0 or earlier are acceptable for other cases.	Version 8.0 and later supported from October 1st, 2016.
WHODrug Global (World Health Organization Drug Dictionary)	Medications	Current version B3 format supported from March 15th, 2019.	▪ WHO Drug Dictionary version 2008:4 (2008-12-01) or later are currently supported. ▪ B3 format is acceptable, even though it is not listed clearly in the catalogs.
LOINC (Logical Observation Identifiers Names and Codes)	Laboratory Test Name	The latest version must be used for the LB domain from March 15th, 2020 for NDAs, ANDAs and certain BLAs, and from March 15th, 2021 for certain INDs.	No requirement for LB codes.
UNII (Unique Ingredient Identifiers)	Substances, including active ingredients, active moieties, inactive ingredients	The latest version must be used for the LB domain from December 17th, 2016 for NDAs, ANDAs and certain BLAs, and from December 17th, 2017 for certain INDs.	For parameters using UNII, it is permissible to store only code values accessible to the applicant.
MED-RT (Medication Reference Terminology)	Pharmacological Class	The latest version must be used for the LB domain from December 17th, 2016 for NDAs, ANDAs and certain	For parameters using MED-RT, it is permissible to store only code values accessible to the applicant.

Standard Terminology	Use	FDA [2]	PMDA [6]
		BLAs, and from December 17th, 2017 for certain INDs.	
SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms)	Indication and Usage	It is required for studies started on or after December 17th, 2016 for NDAs, ANDAs and certain BLAs, and from December 17th, 2017 for certain INDs.	For parameters using SNOMED CT, it is permissible to store only code values accessible to the applicant.

At the end of April 2023, while finalizing the SDTM and ADaM packages for submission to the FDA, the sponsor requested that these packages be adapted for submission to the PMDA in Japan. Special attention was given to the versions of the standards used in the study to ensure they were also acceptable for the PMDA. Here are some of the standards used:

- CDISC CT: 2021-12-17
- SNOMED CT: 2021-07-31
- MED-RT: 2022-07-05
- UNII: 2022-02-13
- WHO Drug Dictionary: Global B3 2021Q1
- MedDRA: version 24.1

After thorough research on the PMDA's requirements, it was decided, in agreement with the sponsor, to retain the same versions used for the FDA submission in the packages for the PMDA, as these standards met the requirements of both regulatory agencies.

THERAPEUTIC AREA STANDARDS

Both the FDA and PMDA use CDISC standards for therapeutic areas, which define the structure and format of clinical trial data. CDISC creates Therapeutic Area User Guides (TAUGs) for various conditions, which outline standard datasets and variables to capture clinical data for specific diseases.

The FDA has a more extensive adoption of TAUGs, and in some cases, mandates the use of specific therapeutic area standards in submissions. Sponsors are encouraged to use TAUGs that align with their therapeutic area, particularly for new drugs or biologics. For studies following a TAUG, the FDA requires the inclusion of the TSVAL value in the TS (Trial Summary) domain for TSPARMCD="CTAUG" (CDISC Therapeutic Area User Guide) for tracking purposes, with all supported therapeutic areas listed in Section 5.2 of the SDTCG.[1]

The PMDA also adhere to CDISC standards for therapeutic areas but is generally less prescriptive in enforcing specific TAUGs. Its primary requirement is compliance with SDTM and ADaM standards, which are consistent with global CDISC guidelines. In practice, sponsors often align PMDA submissions with those for the FDA, provided the data meet both agencies' requirements.

Both agencies collaborate with CDISC to harmonize therapeutic area standards, though the FDA tends to adopt new TA guidelines more quickly, while the PMDA takes a more cautious approach in implementing them.

SDTM AND ADAM DATASETS

The SDTM and ADaM datasets that adhere to CDISC standards should be submitted to both FDA and PMDA in the SAS® XPORT file transport format Version 5.0 (referred to as SAS XPORT format), with each dataset provided as a separate file. For PMDA only, the same requirement holds for datasets containing variables in Japanese and the character sets or encoding schemes used to create these datasets should be detailed in the reviewer's guide. Furthermore, the FDA, in partnership with CDISC and PHUSE, has tested the CDISC Dataset JSON message exchange standard, with initial results suggesting it could replace XPT v5.0. CBER and CDER will conduct further evaluations to assess its suitability for regulatory study data submissions and will share findings and seek input from stakeholders during the process.

The PMDA SDTCG specifies that for clinical pharmacology studies both the PP domain, which stores pharmacokinetic parameters, and the PC domain, which contains drug concentration data, should be submitted, as pharmacokinetic parameters are considered essential data for capturing the drug's characteristics. Additionally, it is recommended to create a RELREC dataset based on the SDTM-IG to explain the relationship between the PC and PP domains. However, if creating the RELREC dataset is challenging, it is acceptable to describe these relationships in a document such as the reviewer's guide. This requirement also applies to the FDA, even though it is not explicitly stated in the SDTCG. In the two studies under review, RELREC was not included. However, the sponsor chose to document the relationship between the two domains in the materials to be submitted to the regulatory agencies.

According to CDISC standards, the FDA's SDTCG states that some permissible or expected variables should be included in SDTM even if no data are collected. Examples include:

- Baseline flags for laboratory results, vital signs, ECG readings, pharmacokinetic concentrations, and microbiology results.
- The EPOCH variable, which helps reviewers easily identify the study phase in which an observation or event was recorded.
- When timing variables such as --DTC, --STDTC, or --ENDTC are included in a general observation class domain, the corresponding study day variables (--DY, --STDY, or --ENDY) should also be provided.

The above rules are requirement in PMDA SDTCG. In our studies, we ensured that the above guidelines were followed subsequently removing the permissible variables with all missing values.

ADaM datasets serve as a support tool for regulatory reviewers, but it's not mandatory to submit ADaM for every analysis defined in the SAP. The PMDA requires datasets for analyses that are performed to obtain significant results on efficacy and safety, as well as clinical study outcomes that provide the rationale for dosage and administration settings. This includes primary efficacy analysis, secondary efficacy analyses, primary safety analyses, basic analyses of adverse events, and analyses investigating the effect of major factors on efficacy and safety. It is advisable to consult with the PMDA before proceeding with the submission. On the other hand, the FDA requires the submission of ADaM datasets to support efficacy and safety analyses, including the analysis of immunogenicity.

The definition documents should be created separately for SDTM and ADaM datasets as eXtensible Markup Language (XML) files, and named "define.xml". These files contain references to style sheets that allow the content to be displayed and should be stored in the same folder as the corresponding datasets, along with the style sheets. The definition document is used to describe the datasets, variables, possible variable values, and controlled terminologies and codes. Additionally, the document should specify the versions of the controlled terminologies and dictionaries used, that are stored also in the TS domain.

Versions of Implementation Guides for SDTM and ADaM, and define-XML supported by agencies are summarized in the below table.

Data Exchange Standard	FDA [2]	PMDA [6]
SDTM	▪ SDTM-IG v3.3 (SDTM v1.7) ▪ SDTM-IG v3.2 (SDTM v1.5) support ends on the December 13th, 2023 ▪ SDTM-IG v3.4 (SDTM v2.0) requirement will begin on March 15th, 2025	SDTM-IG v3.1.2 (SDTM v1.2) or later
ADaM	▪ ADaM-IG v1.1 or later (ADaM v2.1) ▪ ADaM-IG v1.0 (ADaM v2.1) support ends on March 15th, 2019 for NDAs, ANDAs and certain BLAs, and one year later for certain INDs	ADaM-IG v1.0 or later (ADaM v2.1)
Define-XML	▪ Define.xml v2.0 and v2.1 ▪ Define.xml v1.0 support ended on March 15th, 2018	▪ Define.xml v2.0 and v.2.1 ▪ Define.xml v1.0 support will end on March 31st, 2025

PMDA strongly recommend in its SDTCG the inclusion of the Analysis Results Metadata (ARM) in the definition documents of the ADaM datasets, but submission as a separate PDF file is allowed too. FDA doesn't specifically require it, but reviewers find it useful for identifying the critical analyses, providing links between results, datasets, and variables, and documenting the analyses performed to obtain the main results of efficacy and safety, as well as clinical study results providing the rationales for setting of the dosage and administration. ARM is also related to the analysis programs for submission, and it's advisable to agree with the PMDA in advance during a Clinical Consultation meeting on which ADaM datasets and analysis programs will be submitted.

In the case studies presented in this paper, the same define-XML for SDTM datasets was included in the submission packages for both regulatory agencies. However, for the ADaM definition file, the sponsor required an ARM with the main figures and tables exclusively for submission to the PMDA. Consequently, we generated the define-XML file based on the ADaM specifications for FDA package, while simultaneously developing the ARM in an Excel file. This ARM was then integrated into the existing define-XML for the PMDA package using a SAS program.[12] Both define-XML files successfully passed all compliance checks using the Pinnacle 21® validation tool.

HANDLING OF UNITS IN FINDINGS DOMAINS

The handling of units in Findings domains, such as Laboratory (LB) and Vital Signs (VS), differs between the FDA and PMDA, with each regulator having distinct preferences and requirements for clinical trial data submissions.

The FDA mandates the use of standardized units across all clinical data submissions, particularly for laboratory and vital signs data. The agency recommends that results be consistently converted to standardized units, even if different units are used across sites, to ensure uniformity in the --STRESU field.[1]

In contrast, the PMDA does not impose a strict requirement for converting laboratory and vital signs data into standardized units. While the use of SI units is encouraged in principle, the PMDA often accepts results in their original units as collected during the study. However, if units vary across clinical sites or regions, the PMDA expects this to be clearly documented in the submission.[9]

For submissions to both the FDA and PMDA, it is common practice to include both the original units (via --ORRESU) and the standardized units (via --STRESU) to meet the expectations of both agencies, as demonstrated in the two case studies referenced in this paper. Providing clear documentation of conversion factors and unit definitions is crucial to avoid confusion or inconsistencies, particularly when multiple units are used across different sites. This ensures transparency and accuracy in data interpretation across submissions.

ANNOTATED CRF

An annotated Case Report Form (aCRF) is required by regulatory authorities. It is a PDF document titled "acrf.pdf" that links the clinical data collection fields used to capture subject data (whether electronic or paper) to the corresponding variables or categorical variable values in the SDTM datasets.[1, 9] The document is annotated by following the SDTM Metadata Submission Guidelines (SDTM-MSG) provided by CDISC, with v2.0 released on March 30th, 2021. Furthermore, the FDA SDTCG clearly states that an aCRF should be preferably submitted together with the protocol, regardless of whether the clinical database format is supported by the Catalog.[1] Thanks to the consistency of the guidelines, in our studies it was possible to use the same aCRF for submission to both regulatory agencies.

REVIEWER'S GUIDES

The creation of relevant Reviewer Guides (RGs) is strongly recommended by both agencies as part of a standards-compliant study data submission. The clinical Study Data Reviewer's Guide (cSDRG) and the Analysis Data Reviewer's Guide (ADRG) should outline any special considerations, directions, or conformance issues that could assist a reviewer in understanding and utilizing the submitted data during the review process, as well as clarifying the relationship between the study report and data, and explaining the utilization and adherence to data standards in the dataset creation.

Besides protocol and clinical study design information, the cSDRG includes a list and details of all SDTM domains to be submitted and of major supplemental qualifiers. It should be stored as a PDF file named "csdrg.pdf", but names like "study-data-reviewers-guide" are also acceptable for PMDA. Similarly, ADRG should be created as a PDF file titled "adrg.pdf" to describe all ADaM datasets and the programs and variables affecting the most critical analyses. Names like "analysis-data-reviewers-guide.pdf" are also acceptable file names for PMDA.[1,9]

For PMDA submission, information on Japanese data must be included in both cSDRG and ADRG, and both reviewer's guides may be written in Japanese.[9] The PMDA also accepts RGs written in English.

Unlike the FDA, the PMDA requires a clear explanation of any validation criteria violations that cannot be resolved prior to submission. As mentioned earlier, the PMDA mandates the preparation of e-study data before the Consultation on data format, along with the inclusion of the aforementioned Form A. The RGs could be attached to Form A. This highlights the importance of planning, alignment, and collaboration among the teams involved in the preparation and submission activities.

Creating a single Reviewer Guide for both agencies, PMDA and FDA, is possible, but it presents some challenges. Each agency has specific requirements and different guidelines for data submission. However, standard templates, such as those published by Pharmaceutical Users Software Exchange (PHUSE), can be used, aiming to include rules that meet the needs of both agencies. This could help reduce the differences between the required documents, but adjustments will still be necessary to meet the specific requirements of each regulatory body.

In the case studies described in this paper, the use of PHUSE standard templates allowed us to keep the RGs identical for both regulatory agencies, apart from Section 4 of the cSDRG and Section 6 of the ADRG, "Data Conformance Summary", due to differences in validation rules (which will be discussed later in this paper). Additionally, Section 7 of the ADRG, "Submission of Programs", also differed between the two agencies. In our studies, it was agreed with the sponsor to have separate documents for the two regulatory authorities. However, we could have also unified the documents by including the version of the validation engines and the data validation results from both agencies, and submitting the same program, as we will discuss in the following paragraph.

SOFTWARE PROGRAMS

PMDA and FDA have similar requisites for the programs to be submitted: sponsors should provide FDA the software programs used to create all ADaM datasets and generate tables and figures associated with primary and secondary efficacy analyses, whereas the PMDA claim cover programs to create ADaM datasets and used for analyses on the important results on efficacy and safety and clinical study results that provide the rationales for setting of the dosage and administration in order to understand the process by which the variables for the analyses were created and to confirm the analysis algorithms and results.[1, 9] Information on the specific software used for creating the programs must be specified in the RGs.

FDA requirements allow the sponsor to use reliable software programs with appropriate testing procedures and submit these programs in American Standard Code for Information Interchange (ASCII) text format.[1] This requirement is often mistakenly interpreted as the need to rename SAS programs with a .txt extension. However, SAS programs are in ASCII format by default, regardless of the file extension. Therefore, files can be submitted with the .sas extension without any renaming. On the other hand, the PMDA SDTCG states that the submitted programs are not restricted to specific software or versions. However, the file name must include the extension added by the analysis software.[9]

File names should only contain lowercase letters, numbers, or hyphen-minus ("-"). The underscore ("_") is accepted by the PMDA but not by the FDA, and all other special characters are prohibited by both agencies. Additionally, file names may be up to 32 characters long for datasets and 64 for other programs, including the extension, when submitting to the PMDA. For FDA submissions, a maximum of 64 characters is allowed for all file names. Neither agency expects sponsor to submit executable programs, although the FDA may request them in certain cases.[1,9]

The FDA's SDTCG does not include specific requirements for submitting operational company macros. However, macros used to perform essential analyses must still be submitted, as they are crucial for the review process. In contrast, PMDA guidelines recommend that macros preferably be submitted when they are used in the programs being submitted. If submitting macros is challenging or impractical, providing detailed specifications that describe the analysis algorithms is considered sufficient for reviewers.[9]

In the two studies under review, datasets and outputs were all generated using SAS v9.4, and the programs were renamed to meet the naming conventions required by the eCTD guidelines. Accordingly, SAS programs were provided to the sponsor for submission to the PMDA, while .txt files containing the string "-sas" before the file extension were prepared for the FDA submission. However, we could have included the same .sas files in both packages, with names that comply with the rules outlined earlier. All macros used for ADaM and analyses, except for company-specific ones, were supplied to the sponsor in dedicated catalogs and as SAS programs.

DATASETS AND VARIABLES REQUIREMENTS

Each dataset should be submitted as a single transport file. The FDA's ability to process datasets depends on multiple factors, and datasets exceeding 5 gigabytes (GB) should be divided into smaller segments, with each one no larger than 5 GB. Sponsor should submit these smaller, split datasets along with the original, unsegmented dataset to facilitate the review process. Definition files should be linked to the main dataset. The split datasets should be stored in a separate subdirectory labeled "split". A detailed explanation of the dataset splitting process should be included in the Reviewer's Guides.[1] For PMDA, there is no specific limitation in terms of dataset size, but it is recommended to consult PMDA beforehand if the dataset size is greater than or equal to 5 GB.[9]

Section 4.1.5 of the PMDA SDTCG outlines the procedures for handling Japanese data within datasets and the submission process when a domain includes Japanese items. Generally, for domains and datasets containing Japanese characters, two datasets must be provided during submission: one dataset with Japanese characters and another dataset with the same data represented using ASCII letter sets.[9] Further details are provided later in this paper.

Datasets and variables names, as well as labels, must adhere to CDISC standards for both regulatory agencies, with names limited to a maximum of 8 characters and labels to 40 characters. Names must be descriptive and consistent across the submission, with clear labels provided for each variable. Datasets names should consist only of lowercase letters and numbers, starting with a letter. For legacy studies initiated on or before December 17th, 2016, FDA allows the use of the underscore character "_" in variable and dataset names. Similarly, uppercase letters in variable names are permitted by FDA under these conditions instead of lowercase.

Both variable names and labels, as well as dataset labels should use ASCII text codes exclusively. While UTF-8 encoding can be used to extend character sets; however, the use of extended mappings is generally not recommended. The length allocated for each column containing character (text) data should be set to the maximum length of the variable used across all datasets in the study. Furthermore, only ASCII and punctuation characters can be used in variables.

ECTD FOLDER FOR SUBMISSION

Study datasets and their supporting files must be organized into a specific and predefined directory structure when submitted in the eCTD format (Figure 1).

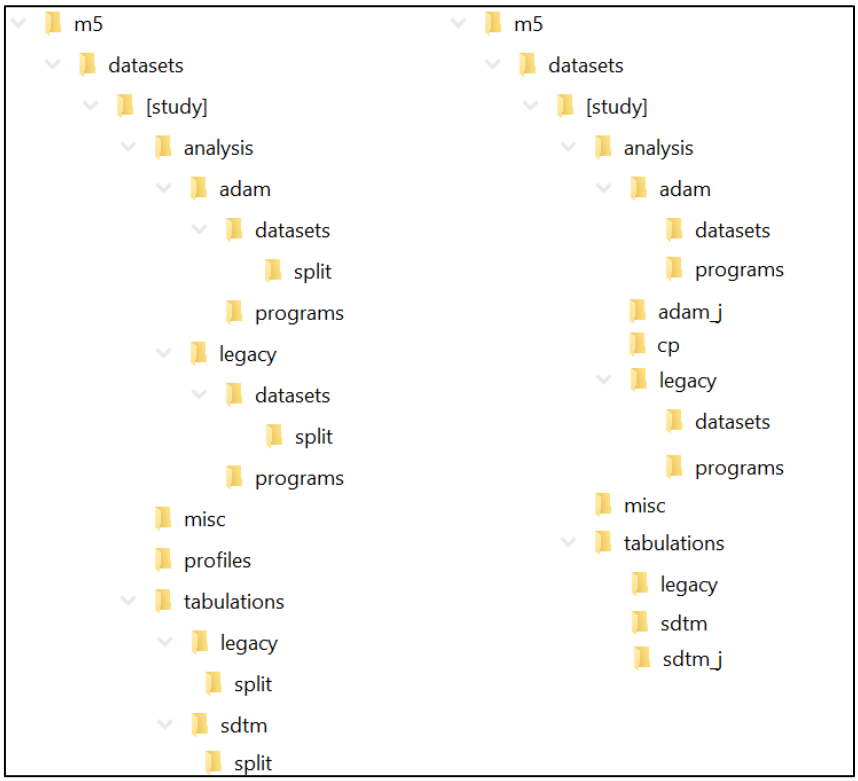


Figure 1 - Folder Structure for Study Datasets, for FDA (left) and PMDA (right) submissions.[1,9]

Additional or unauthorized subfolders should not be created, and empty folders should be removed. If organizing data according to the folder structure outlined in the SDTCG proves challenging, the applicant should consult the PMDA in

advance and submit data only after agreeing on the folder structure and file storage. This consultation process is not explicitly required by the FDA.

The top-level folder for storing clinical data within Module 5 should be named “m5”. The path length, including the file name, must not exceed 160 characters for PMDA submissions and 150 characters for FDA submissions. Folder names may only include alphabetic characters, numbers, underscores, and hyphens.

A folder should be created for each study or for each integrated analysis results in the “datasets” subfolder of the main directory. It should be named with the study number (e.g. study123) or the type of analysis (e.g. iss, ise) that allows the unique identification of each study. The study number used and the study identifier (STUDYID) in the trial summary (TS) should be identical wherever possible. This folder should not contain any files other than lower folders.

The folder of each study or integrated analysis is then organized as follows:

- “analysis”: to store analysis datasets, programs and documents. Files are stored only in the subfolders:
  - “adam”*:* in the “datasets” subfolder the applicant should store ADaM datasets, the ADRG and the define-XML file with the style sheet, while the subfolder “split” is planned in FDA submission for including any split ADaM datasets. If the latter subfolder is non-empty, an exhaustive explanation should be provided in ADRG for the split datasets. Software programs for ADaM datasets, tables and figures included in the submission, are placed in the “programs” subfolder.
  - “adam_j”: the folder is included in the PMDA submissions only to store the Japanese datasets that corresponds to the alphanumeric ADaM dataset contained in “datasets”, subdirectory of “adam”. No other files should be added in this folder.
  - “legacy”: its subfolders are structured similarly to the “adam” subdirectory with ADaM datasets being replaced by datasets in other than ADaM format.
  - “cp”: this folder is included in PMDA SDTCG for storing electronic study data on clinical pharmacology in formats other than the CDISC standard.
- “tabulations”: to store aCRF and datasets in which subject data and study-related information are displayed in a tabulation format; files are collected only in the subdirectories:
  - “legacy”: all legacy tabulation datasets are included in this folder; and the split datasets sourced from datasets larger than 5 GB in size are stored in the “split” subfolder for FDA submission.
  - “sdtm”: SDTM tabulation datasets are collected in this directory and, as for the other folders, any datasets extracted from those exceeding 5 GB are included in the “split” subfolder.
  - “sdtm_j”: the folder is included in the PMDA submissions only to store the Japanese SDTM datasets that corresponds to the alphanumeric SDTM dataset in “sdtm”. No other files should be included in this folder.
- “misc”: contains data that doesn’t fit into other specific subfolders but is essential for analysis and reporting. The folder could include miscellaneous datasets consisting of any data not captured in SDTM but used in ADaM or for outputs programming, but also look-up tables such as AEs of special interest, deviations recorded from source other than the CRF, or prohibited concomitant medications.
- “profiles”: to store patient profiles, but it is not applicable for PMDA.

While the overall hierarchy of the “m5” folder is broadly similar between the PMDA and FDA, reflecting a global standardization in eCTD submissions, subtle differences in path length limits, consultation requirements, and subfolder usage underscore the importance of tailoring the submission to the specific agency's guidelines.

DATA VALIDATION

Study data validation ensures that the data are compliant, reliable, and capable of supporting meaningful review and analysis. Validation activities take place at various stages, including when the data is first submitted and at the onset of the regulatory review process.

Both agencies utilize Pinnacle 21 Enterprise® to perform validation of data on compliance with CDISC standards, adherence to FDA and PMDA specific validation rules: the FDA utilizes the most recent version, while the PMDA uses version 5.1.2. If issues are found during validation, the FDA expects the sponsor to address them before submission or provide an explanation for any validation rule violations.[1] Similarly, the PMDA might require additional clarifications or corrections based on the validation results, but the specific nature of these requests may vary according to PMDA’s guidelines.[9] Before submitting electronic study data, the applicant should perform a validation and should take the necessary action for any violations that are identified, or document them in the RGs. These validations are performed on define-XML only, and on the same with the submitted datasets, for both SDTM and ADaM. Furthermore, three SDTM domains (AE, DM and EX) need to be included in the validation tool when validating ADaM packages to support traceability between SDTM and ADaM. Agencies also accept validation with Pinnacle 21 community version, with the latest validation engine.

Validation rules and their severity differ between the two authorities.

The PMDA rules and associated issues identified by Pinnacle 21 have been categorized by their level of importance, considering the characteristics of each rule, adherence to standards, ease of data use during the review process and for future applications, as well as the quality of clinical data expected by the agency. There are three distinct level of issues and importance:[9]

- REJECT: this rule must not be violated, as doing so will result in the suspension of the review until corrections are made. The issue must be resolved prior to submission.

- ERROR: this rule should not be violated, and any fixable errors must be corrected. If not, a clear explanation must be provided in both the Reviewer's Guide and Form A with a well-defined rationale. Additionally, these errors must be discussed during the 'Consultation on Data Format' meeting before the NDA submission. Unexplained errors would cause the review to be suspended until corrections or justifications are provided. If the PMDA requests additional clarification, Form A must be updated and resubmitted to the agency following the consultation meeting.
- WARNING: while this rule should not be violated, if the violation cannot be corrected, an explanation must be provided in both the Data Reviewer's Guide and Form A.

On the other hand, the FDA's validation rules do not assign specific severity to violations of their business rules, and CDISC does not provide severity level for non-conformance to their standards either. Nevertheless, based on the previous classification, two levels of issues and importance can be identified for FDA validation rules:

- REJECT: These are reported as "Error" in the Pinnacle 21 tool but are considered rejection criteria that must be resolved before submission.
- ERROR/WARNING: all other issues, while not assigned a severity in Pinnacle 21 Community, are grouped in this paper under this category for consistency with the PMDA classification. These issues should be documented and explained in the Reviewer's Guide with a clear rationale. If possible, they should be corrected.

It is important to note that what is acceptable to the FDA may not necessarily be acceptable to the PMDA.

Let's compare the number of validation rules and their classification between the FDA and PMDA using SDTM-IG 3.3 and validation engines FDA 2304.3 and PMDA 2311.0, in Pinnacle 21 Community Version 4.0.2. The FDA has 503 validation rules, with only 6 identified as rejection criteria, whereas the PMDA recognize fewer violations (466), but with 8 rejection criteria, 196 errors and 262 warnings. The two agencies share only two rejection criteria. The remaining six rejection criteria for the PMDA are categorized as warnings or errors by the FDA. Similarly, three FDA rejection criteria are classified by the PMDA as errors (3) or warning (1).

Pinnacle 21 ID	Message	Description	FDA Severity	PMDA Severity
DD0101	Missing define.xml file	Guidance provided by regulatory agencies states that eCTD submissions must include a define.xml file for each study in Module 4 (nonclinical) and Module 5 (clinical). Note: Refer to the data standards catalog by the relevant Regulatory Agency for acceptance of standards and their versions.	Reject	Reject
SD0002	NULL value in variable marked as Required	Required variables (where Core attribute is 'Req') cannot be NULL for any records.	Error/ Warning	Reject
SD0056	SDTM Required variable not found	Variables described in SDTM IG as Required must be included in the dataset.	Error/ Warning	Reject
SD0062	Incompatible data source	Domain table must have a valid format (e.g., SAS transport (XPORT) v.5 or text-delimited).	Error/ Warning	Reject
SD0064	Subject is not present in DM domain	All Subjects (USUBJID) must be present in Demographics (DM) domain.	Error/ Warning	Reject
SD1020	Missing DM dataset	Demographics (DM) dataset must be included in every submission.	Reject	Reject
SD1073	Variable prohibited for use in SDTM	Variables described in IG as inappropriate for usage must be not included in the dataset.	Error/ Warning	Reject
SD1074	Variable which can be used only in SEND	Variables designed only for SEND pre-clinical studies must be not included in the SDTM dataset.	Error/ Warning	Reject
SD1115	Missing TS dataset	Trial Summary (TS) dataset must be included in every submission.	Reject	Error
SD2232	Missing SSTDTC Trial Summary Parameter	'Study Start Date' (SSTDTC) record must be populated in Trial Summary (TS) domain. It is expected for SDTM IG v3.1.2 data and required for data in all more recent SDTM versions.	Reject	Warning
SD2247	Invalid TSVAL value for SSTDTC	TSVAL variable value must be in ISO 8601 format when TSPARMCD='SSTDTC'.	Reject	Error
SD2247A	Incomplete date for study start date	TSVAL variable value must be in ISO 8601 format and include at least full date, when TSPARMCD='SSTDTC'.	Reject	Error

At the same time, 49 FDA errors or warnings are not considered as violations by the PMDA while 13 PMDA errors or warnings do not apply to the FDA. Among these are issues like errors for datasets larger than 5GB, which are handled

differently by the two agencies: although the message is the same and the severity is comparable, the description varies.

Pinnacle 21 ID	Message	Description	FDA Severity	PMDA Severity
SD1071	Dataset is greater than 5 GB in size	Datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets no larger than 5 GB.	Error/ Warning	
SD1142	Dataset is greater than 5 GB in size	Applicant should consult the PMDA beforehand if the dataset exceeds the file size specified by PMDA.		Warning

As previously discussed for data standards catalogs, there are no requirements for LOINC codes in PMDA submissions, while SNOMED CT is allowed but is not mandatory. As a result, the FDA has specific validation rules concerning these controlled terminologies, whereas the PMDA does not. Some of these differences are summarized in the table below.

Pinnacle 21 ID	Message	Description
SD2240	Invalid TSVCDREF value for INDIC	TSVCDREF variable value must be 'SNOMED' when TSPARAMCD='INDIC'.
SD2257	Invalid TSVAL value for INDIC	TSVAL for INDIC record must be a valid term from SNOMED CT.
SD2258	Invalid TSVALCD value for INDIC	TSVALCD for INDIC record must be a valid concept ID from SNOMED CT.
SD2259	TSVAL/TSVALCD value mismatch for INDIC	TSVAL and TSVALCD values must be populated from the same concept description record in SNOMED CT.
SD2266	Invalid TSVCDREF value for TDIGRP	TSVCDREF variable value must be 'SNOMED' when TSPARAMCD='TDIGRP'.
SD2267	Invalid TSVAL value for TDIGRP	TSVAL for TDIGRP record must be a valid term from SNOMED CT.
SD2268	Invalid TSVALCD value for TDIGRP	TSVALCD for TDIGRP record must be a valid concept ID from SNOMED CT.
SD2285	Value for --LOINC not found in LOINC coding system	Value for the LOINC Code (--LOINC) variable must be populated using a LOINC coding system version specified in the define.xml.

Similarly, Pinnacle 21 identifies 324 validation rules for PMDA and 263 for FDA using the same validation engines and ADaM-IG 1.1. The rejection criteria shared by both agencies are two: the absence of the ADSL dataset or the define.xml file in the submission. Additionally, submitting datasets in an incompatible format results in a rejection for the PMDA, while it generates an error/warning for the FDA. Lastly, the PMDA has 13 rejection criteria related to the validation of define.xml files, compared to 0 for the FDA.

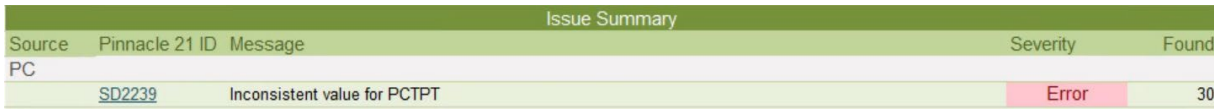
Difference in validation rules among regulatory agencies should be taken into account when creating SDTM and ADaM packages. The best practice is to ensure that checks are conducted early and continuously, in line with CDISC and regulatory agency compliance rules. This approach helps resolve errors as they arise and ensures that no validation rejections occur. The differences in validation rules and results can be addressed in the relevant section of the RGs by including the validation results of both authorities' rules, regardless of whether the submission is to the FDA or the PMDA.

Three examples from one of the above case studies are now presented, encountered during the preparation of submission packages for the PMDA and the finalization of those for the FDA, along with a description of the decisions made.

- When validating SDTM datasets built according to SDTM-IG 3.3, Pinnacle 21 identified a validation rule violation in the PC dataset for both regulatory bodies. This issue was flagged as "Error" in PMDA report (Figure 2), whereas severity did not appear at all in the FDA report (Figure 3). Once the packages were finalized in August 2023, the FDA 2201.4 and PMDA 2211.0 validation engines were utilized but the severity is the same in the current versions of the validation engines.

Issue Summary				
Source	Pinnacle 21 ID	Message	Severity	Found
PC				
	SD2239	Inconsistent value for PCTPT		30

Figure 2 - Issue identified by Pinnacle 21 with SDTM-IG 3.3 and validation engines FDA 2201.4 and 2304.3



Issue Summary				
Source	Pinnacle 21 ID	Message	Severity	Found
PC				
	SD2239	Inconsistent value for PCTPT	Error	30

Figure 3 - Issue identified by Pinnacle 21 with SDTM-IG 3.3 and validation engines PMDA 2211.0 and 2311.0

Upon concluding that the error could not be resolved, it was decided in agreement with the sponsor, to provide the clear below explanation in the cSDRG for both submissions, with more details than initially planned in the document for the FDA:

Out of the total number of records, 24 of them do not present any issues since each one has one plasma and one urine concentration measurement that corresponds to a single date and time. For example, the subject 001-2004 at Day 7 has two measurements collected on 2021-11-01T20:46: the first one from urine at planned time point "12-24 hours post dose", the second from plasma "12 hours post-dose".

For the remaining 6 issues, subject 001-1002 has the same plasma collection date/time on Day 7 for two consecutive planned time point: "0.75 hours post-dose" and "1 hour post-dose". For this subject, ACTTIME values at 0.75 and 1.0 hours post-dose were equal (1.16666666666667) and so nominal times (0.75 and 1.0) were used for these two data points in the analysis (this change was tracked under the Data TAB in the Phoenix WinNonlin® workflow).

- While preparing the SDTM package for FDA, the sponsor agreed to use DM.RACE for collecting "Ethnic descent" from the CRF, taking values of "CAUCASIAN" and "JAPANESE". Since the RACE codelist is non-extensible, and these values do not apply to the exceptions listed in the SDTM-IG 3.3, this led to a validation rule violation that needed to be justified in the cSDRG. However, for the PMDA this was flagged as an error. Given that the error was fixable, it was decided together with the sponsor to resolve the issue in both FDA and PMDA packages, proceeding as follows:
  1. JAPANESE and CAUCASIAN have been mapped to ASIAN and WHITE, to comply with the RACE non-extensible codelist.
  2. The values captured on the CRF, have been represented using the supplemental qualifier RACEOR (Original Race).
  3. Both RACE and RACEOR have been included in ADSL and in all other ADaMs.
- The validation of ADaM datasets for the FDA submission did not identify any issues. However, an error related to the ARM variable in the ADSL dataset was detected during the PMDA's validation process (Figure 4).

Issue Summary

Source	Pinnacle 21 ID	Message	Severity	Found
ADSL				
	AD0196	Required ARM value is null	Error	11

Figure 4 - Issue identified by Pinnacle 21 with ADaM-IG 1.1 and validation engine 2211.0 and 2311.0

After a thorough investigation and research, it was concluded that this is an unfixable error due to an inconsistency between SDTM-IG 3.3 and ADaM-IG 1.1, and it became necessary to include the following detailed explanation in the ADRG:

False positive: The 11 subjects considered screen failures have ARM derived as per SDTM-IG 3.3 guidelines (section 5.2: "Name of the Arm to which the subject was assigned. If the subject was not assigned to an Arm, ARM is null and ARMNRS is populated."), meaning that they will have ARM=null in the DM domain. Per ADaM-IG, the ARM variable in ADSL is "Required", and should be a direct copy from its predecessor (DM.ARM). This is an inconsistency in CDISC between the SDTM rules for DM.ARM and the ADaM ADSL.ARM definition (Required).

HANDLING OF DATA WRITTEN IN JAPANESE TEXT

As anticipated in the previous section, the PMDA's SDTCG outlines how to handle data written in Japanese.[5]

Data collected in Japanese should be submitted if translating it into English risks losing certain information, or if the descriptions in Japanese are necessary and appropriate. To meet these requirements, two datasets must be generated and submitted:

- One dataset using ASCII character encoding
- A Japanese dataset, where only the Japanese items are written in Japanese, and other items are alphanumeric.

The two datasets must have the same structures, with the only differences being the data lengths for the Japanese items and the corresponding alphanumeric sequences. Both datasets must have the same number of records, record order, name, and label. The alphanumeric dataset should be included in the designated subfolder of the Module 5 directory, while Japanese SDTM and ADaM should be placed in the "sdtm_j" and "adam_j" subfolders, respectively. Only the alphanumeric dataset must be included in the submission module 5 directory if it doesn't contain any Japanese items.

In alphanumeric datasets, an English placeholder (e.g. "JAPANESE TEXT IN SOURCE DATA") should be used in place of the original Japanese text to clearly indicate that it is not the original data. If the Japanese data must be stored across multiple variables or records due to length restrictions, this English placeholder must be consistently placed in the corresponding records within the alphanumeric dataset. The chosen English character sequence must remain consistent throughout the study and be clearly documented in the RGs or define.xml files.

In some cases, however, it is useful to provide the English translation of the Japanese item, such as for questionnaires. If translation proves challenging, the translated items can be stored in external databases. In the alphanumeric dataset, a character sequence followed by a number can be included (e.g., "JAPANESE TEXT IN SOURCE DATABASE 01" and "JAPANESE TEXT IN SOURCE DATABASE 02") to distinguish these items. For same example, please refer to Attachment 3 in the PMDA SDTCG.[9]

CONCLUSION

Although agencies require CDISC standards and have similar guidelines, there are some differences that might not be fully harmonized in the future. These differences limit the ability for a sponsor to create one package that satisfies both the FDA and PMDA requirements. However, with good planning and thorough understanding of agencies guidelines, the differences between the submission packages for the two agencies can be limited.

A successful approach, for example, is to ensure that checks are performed early and on an ongoing basis against CDISC and agency compliance rules. This enables errors to be revolved whenever possible and ensures that there are no reject validation outcomes.

Additionally, it is also recommended to engage in dialog with regulatory authorities as early as possible.

It would be helpful also to adopt the latest CDISC models and IG of SDTM, ADaM and controlled terminologies to ensure compliance with both agencies. Using ARM is advised to facilitate the review of key analyses with programs submitted as .sas files, and conversion to SI units in Finding domains is recommended to meet PMDA requirement.

Alignment and collaboration among the teams involved in preparation and submission activities is also fundamental for successful results.

From our experience, working on PMDA packages for the first study provided valuable insight into the PMDA guidelines, enabling us to better understand their specific requirements. This knowledge allowed us to proactively address and anticipate changes likely to be requested by the PMDA SDTCG when preparing the datasets and the RGs for the FDA submission package in the second study. By incorporating these adjustments early on, we optimized the process, saving time and resources, reducing costs, and ensuring smoother regulatory compliance with both agencies.

WHAT COMES NEXT?

At Alira Health, we are actively discussing ways to update our procedures and optimize our processes to enable the preparation of a single submission package that meets the common requirements of both regulatory authorities, allowing for submission to both agencies with minimal modifications.

Looking ahead, one crucial next step could be a closer alignment of the validation rules between the FDA and PMDA, with a stronger collaboration between the two agencies than what is currently in place. For instance, they could work together to develop a shared validation tool or fully harmonize their validation criteria and the severity levels of their rules. This would significantly simplify the process for pharmaceutical companies and CROs.

REFERENCES

1. Study Data Technical Conformance Guide. U.S. Food & Drug Administration. Version 5.8, September 2024. www.fda.gov/media/153632/download. Accessed September 2024.
2. FDA Data Standards Catalog. U.S. Food & Drug Administration. September 2024. www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-catalog. Accessed September 2024.
3. Providing Regulatory Submissions in Electronic Format – Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications. Guidance for Industry. U.S. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). September 2024. www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-regulatory-submissions-electronic-format-certain-human-pharmaceutical-product-applications. Accessed September 2024.
4. PMDA Profile of Services. Japan Pharmaceuticals and Medical Devices Agency. September 2023. www.pmda.go.jp/files/000241469.pdf. Accessed September 2024.
5. Revision of Technical Conformance Guide on Electronic Study Data Submission. Japan Pharmaceuticals and Medical Devices Agency. April 2024. www.pmda.go.jp/files/000267935.pdf. Accessed September 2024.
6. Data Standards Catalogs. Japan Pharmaceuticals and Medical Devices Agency. March 2024. www.pmda.go.jp/files/000267624.zip. Accessed September 2024.
7. Procedures for Preparation of Explanation of Electronic Study Data (Form A). Japan Pharmaceuticals and Medical Devices Agency. June 2022. www.pmda.go.jp/files/000247169.pdf. Accessed September 2024.
8. Procedures for Preparation of Explanation of Electronic Study Data (Form B). Japan Pharmaceuticals and Medical Devices Agency. June 2022. www.pmda.go.jp/files/000247164.pdf. Accessed September 2024.
9. Revision of Notification on Handling of Submission of Electronic Study Data for New Drug Applications". Japan Pharmaceuticals and Medical Devices Agency. April 2024. www.pmda.go.jp/files/000267942.pdf. Accessed September 2024.
10. SESUG Paper 259-2019, Once Click to Analysis Results Metadata. S. Ravikiran, P. Gathoni. www.lexjansen.com/sesug/2019/SESUG2019_Paper-259_Final_PDF.pdf. Accessed September 2024.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Luca Semeraro
Alira Health
Galleria Ugo Bassi 1 – 40121, Bologna, Italy
Email: luca.semeraro@alirahealth.com