**Paper SA05**

# Practical Operations of Electronic Study Data Submissions

*Priyanka Pujari, Parexel International, United Kingdom*

## ABSTRACT

Regulatory Authorities are continuously updating their electronic study data submission requirements, and this greatly affects timeline and cost for the New Drug Application (NDA). Therefore, in terms of the global drug development, it is more efficient to implement streamlined process for e-study data submission that would meet different requirements from multiple authorities (FDA and PMDA, for example) in order to achieve simultaneous submissions. The purpose of this presentation is to share experience on challenges, points to consider and solutions to implement when submitting e-study data to multiple authorities. In particular, understanding the key requirements, validation rules, documents for consultations and other aspects across Health authorities which should be considered at the development stage of each clinical study. By comprehending the requirements well and understanding how e-data are processed we would be able to implement a globally integrated way for e-study data submission in the future.

## Introduction

In the realm of pharmaceuticals and regulatory processes, the significance of high-quality study data standard deliverables cannot be overstated. Delivering Clinical Data Interchange Standards Consortium (CDISC) submissions for New Drug Application (NDA) require understanding the regulatory agencies requirements and CDISC standards. The path to achieving CDISC compliance is not without its challenges, as industry stakeholders have come to realize. It's become increasingly evident that the regulatory requirements can vary significantly between different FDA review divisions and even more so when extending across the global agencies like PMDA. This variance in expectations and guidelines can lead to issues and inconsistencies within submission packages, resulting in potentially costly delays and complications.

This paper offers a high-level overview of the regulatory requirements and the fundamental components that constitute e-study data submission packages. It highlights situations which can possibly occur if submission requirements are no followed proactively and share optimization ways for e-study data submission.

## Purpose Statement

With evolving industry standards and regulatory guidance's, it is vital to have a consistent approach in creating a submission ready package by implementing the requirements from beginning of the data analysis process.
This will ensure quality at every step and enable simultaneous submissions to multiple health authorities in a streamlined manner.

## Scope

We'll look from statistical programming perspective and focus on two regulatory authorities, the FDA and PMDA:

- To compare regulatory requirements and differences in their submissions review process
- Statistical programmer's responsibilities in the submission process for New Drug Application (NDA).
- Discuss best practices & use of various guidance's/validation tools used in submission process.

## Overview

CDISC compliant e-study data submission is mandatory in PMDA application from April 2020, while FDA has been requesting standard e-data packages earlier. The core of CDISC compliance lies in the validation rules prescribed by each regulatory agency. These rules can vary, especially in their approach to assessing the severity of non-compliance, potentially leading to the suspension of an application review. It is important to plan programming activities proactively and prepare fully compliant submission package from the beginning of the programming & analysis phase. This proposed solution not only equips the team for a quality submission but also serves as a motivating force, instilling confidence, and efficiency at every stage of the analysis process.

Putting forward three hypothetical situations to understand often overlooked submission related activities:
**Scenario 1**: Originally the study was meticulously prepared for FDA submission, adhering to all conformance and validation rules. However, a decision was made to submit the study to PMDA. The study uncovered 'REJECT' issues for source data where the sites were closed and could not be fixed.

**Scenario 2**: A study originally not intended for submission was added to the FDA submission package. This caused the statistical programming team to adapt their deliverables to meet the FDA's requirements with stringent timelines and a stressful environment.

**Scenario 3**: As the study deliverables were ready for submission, it was realized that the annotated CRF was not created. This resulted in unplanned time and effort being spent by study team to review and map raw data to SDTM with unanticipated findings which lead to SDTM dataset and spec updates.

**Key takeaway**
While both regulatory agencies have embraced CDISC standards, it's crucial to recognize that their specific requirements, validation rules, and even severity categories can substantially differ. What might pass muster with the FDA may not meet the standards set by the PMDA. Consequently, it becomes imperative to maintain a keen awareness of these distinctions and plan for Submission readiness before starting analysis process.

## STEP 1: Planning for Submission Readiness

Regardless of the specific regulatory authority, sponsors must possess a comprehensive understanding of both the relevant standards and regulatory requirements.
To facilitate this process, it is suggested to undertake the following key actions:

- Prepare a submission ready checklist by knowing all the

1. **Submission Components**

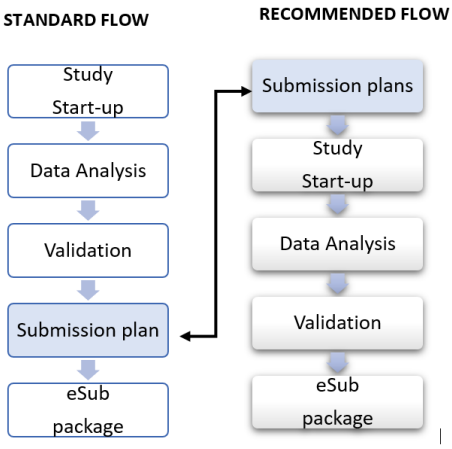
   Gain a thorough comprehension of all the components that comprise a submission package.

2. **Regulatory Requirements**

   Acquire a deep insight into the specific regulatory requirements.

3. **Study Data Traceability**

   Traceability permits an understanding of the relationships between the analysis results, analysis datasets, tabulation datasets, and source data.




Display 1: to demonstrate Standard flow versus Recommended flow of programming activities.

## 1. SUBMISSION COMPONENTS

When planning a study, it is important to be aware of all the Submission components as shown below:

- Study Data Tabulation Model (SDTM) dataset
- Annotated Case Report Form (aCRF)
- Analysis Data Model (ADaM) dataset
- Define-XML for SDTM and ADaM dataset
- Analysis Results Metadata (ARM) in PDF format*
- Study Data Reviewer's Guide (SDRG)
- Analysis Data Reviewer's Guide (ADRG)
- SAS® programs for creating ADaM dataset and analyses

* Submitted to PMDA only

Note: An often-overlooked aspect is the creation of annotated CRFs, which is typically deferred until the final stages of checklist development. However, this can be avoided by having a clear understanding of all components as part of the plan. Knowing the components for submission before starting the analysis process will help in planning and time management for the study and gathering the necessary guidance documents which will avoid rework.

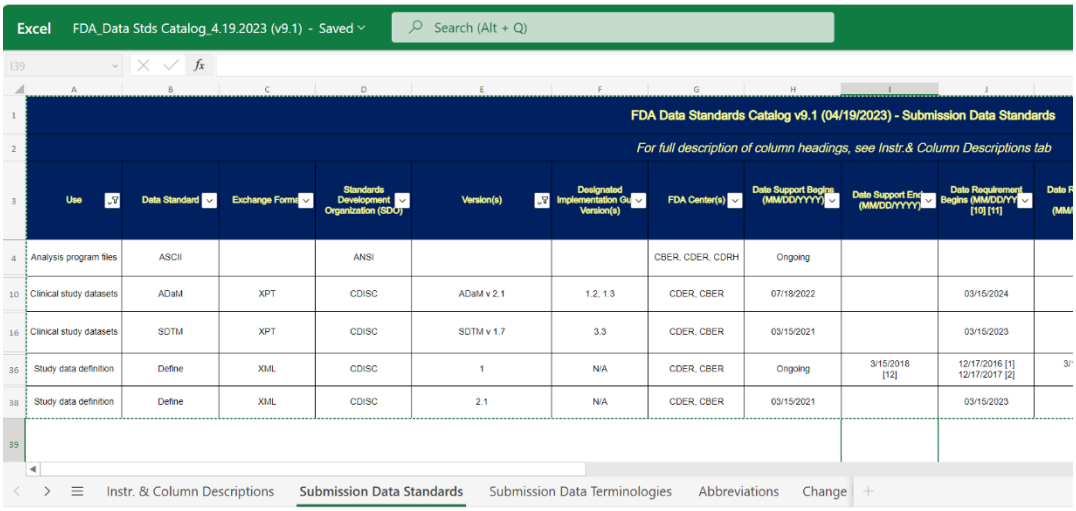
## 2. Regulatory Requirements

There are specific Guidance Documents provided by regulatory agencies like Data Standards Catalog, Study Data Technical Conformance Guide, Validation Rules etc. which are listed below with latest available versions and must be followed by Pharma Companies for successful e-submission of study data.

| FDA | PMDA |
|---|---|
| ***Standards for clinical analysis dataset*** | |
| Data Standards Catalog v.9.1 (19-Apr-2023) | Data Standards Catalog (28-Feb-2023) |
| Study Data Technical Conformance Guide V 4.6 (June-2023) | Technical Conformance Guide on e-Study Data Submissions (April 1, 2022) |
| Business Rules v1.5 (May 2019)<br>Validator Rules v1.6 (December 2022) | Study Data Validation Rules Version 4.0 (28-Feb-2023) |

**a. Data Standards Catalog**

The Data Standards Catalog provides a listing of the currently supported data standards with links to reference materials. The Catalog helps in understanding the requirement of certain standards and terminologies with the date the requirement begins and as needed, the date the requirement ends with information sources.



FDA Data Standards Catalog v9.1 (04/19/2023) - Submission Data Standards

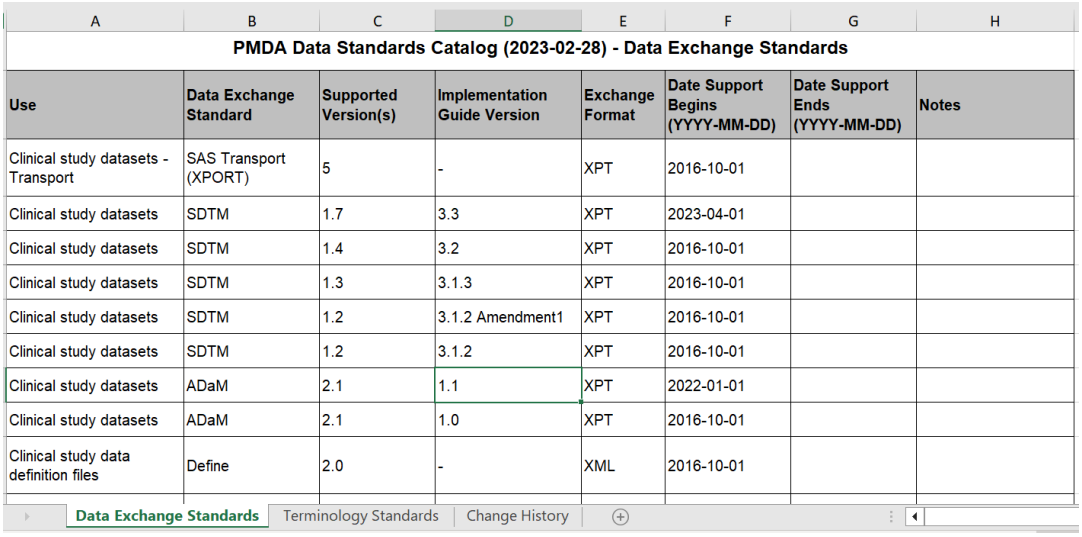
*For full description of column headings, see Instr.& Column Descriptions tab*

| Use | Data Standard | Exchange Format | Standards Development Organization (SDO) | Version(s) | Designated Implementation Guide Version(s) | FDA Center(s) | Date Support Begins (MM/DD/YYYY) | Date Support Ends (MM/DD/YYYY) | Date Requirement Begins (MM/DD/YYYY) [10] [11] |
|---|---|---|---|---|---|---|---|---|---|
| Analysis program files | ASCII | | ANSI | | | CBER, CDER, CDRH | Ongoing | | |
| Clinical study datasets | ADaM | XPT | CDISC | ADaM v 2.1 | 1.2, 1.3 | CDER, CBER | 07/18/2022 | | 03/15/2024 |
| Clinical study datasets | SDTM | XPT | CDISC | SDTM v 1.7 | 3.3 | CDER, CBER | 03/15/2021 | | 03/15/2023 |
| Study data definition | Define | XML | CDISC | 1 | N/A | CDER, CBER | Ongoing | 3/15/2018 [12] | 12/17/2016 [1] 12/17/2017 [2] |
| Study data definition | Define | XML | CDISC | 2.1 | N/A | CDER, CBER | 03/15/2021 | | 03/15/2023 |

Instr. & Column Descriptions | Submission Data Standards | Submission Data Terminologies | Abbreviations | Change

Display 2: Below figure is a screenshot of the latest [FDA catalog version 4.19](#).

Similarly below Display 3, is the screenshot of the PMDA data standards catalog available at [New Drug Review with Electronic Data | Pharmaceuticals and Medical Devices Agency (pmda.go.jp)](#)



PMDA Data Standards Catalog (2023-02-28) - Data Exchange Standards

| Use | Data Exchange Standard | Supported Version(s) | Implementation Guide Version | Exchange Format | Date Support Begins (YYYY-MM-DD) | Date Support Ends (YYYY-MM-DD) | Notes |
|---|---|---|---|---|---|---|---|
| Clinical study datasets - Transport | SAS Transport (XPORT) | 5 | - | XPT | 2016-10-01 | | |
| Clinical study datasets | SDTM | 1.7 | 3.3 | XPT | 2023-04-01 | | |
| Clinical study datasets | SDTM | 1.4 | 3.2 | XPT | 2016-10-01 | | |
| Clinical study datasets | SDTM | 1.3 | 3.1.3 | XPT | 2016-10-01 | | |
| Clinical study datasets | SDTM | 1.2 | 3.1.2 Amendment1 | XPT | 2016-10-01 | | |
| Clinical study datasets | SDTM | 1.2 | 3.1.2 | XPT | 2016-10-01 | | |
| Clinical study datasets | ADaM | 2.1 | 1.1 | XPT | 2022-01-01 | | |
| Clinical study datasets | ADaM | 2.1 | 1.0 | XPT | 2016-10-01 | | |
| Clinical study data definition files | Define | 2.0 | - | XML | 2016-10-01 | | |

Data Exchange Standards | Terminology Standards | Change History

Display 3: screenshot of the PMDA data standards catalog available excel file.

As seen from the figures, the Standards Catalog can be used early in the planning process to identify the CDISC standard versions and initiate Specification writing based on the latest standards. It also gives guidance to the terminology standards, MedDRA versions and format for preparation of various submission components.

**b. The Study Data Technical Conformance Guide (Guide)**

This document provides specifications, recommendations, and general considerations on how to submit standardized study data using FDA-supported data standards located in the FDA Data Standards Catalog. Below is a snippet from the FDA technical conformance guide June 2023 version 5.4 showing the list of content which is useful in data analysis process.

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

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

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

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

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

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

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

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

The Technical Conformance Guide if referred to from the very beginning of the analysis process will aid in providing technical details at the data programming level for e.g., derivation of SUBJID & USUBJID in case of rescreen subjects as per FDA conformance guide is as below. This will allow uniformity in the analysis across studies and across industries.

*Contains Nonbinding Recommendations*

FDA U.S. FOOD & DRUG ADMINISTRATION

Subject Identifier (SUBJID)

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

Unique Subject Identifier (USUBJID)

The variable USUBJID is an identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.[25] Each individual subject should be assigned a single unique identifier across the entire application. This is in addition to the subject ID (SUBJID) used to identify subjects in each study and its corresponding study report. An individual subject should have the exact same unique identifier across all datasets, including between SDTM and ADaM datasets. Subjects that participate in more than one study should maintain the same USUBJID across all studies. It is important to follow this convention to enable pooling of a single subject's data across studies (e.g., a randomized control trial and an extension study).

Sponsors should not add leading or trailing spaces to the USUBJID variable in any dataset. For example, applications have been previously submitted in which the USUBJID variable for each individual subject appeared to be the same across datasets; however, in certain datasets, the actual entry had leading zeros added, or zeros added

The guide also provides specific file expected for submission and mitigation rules for e.g., PMDA conformance guide about program file format.

**4.1.6.2 File format of the programs**

As stated in 4.1.6.1, although the programs to be submitted are not limited to specific software or versions, the file name should include the extension attached by the analysis software. If the file name does not contain an extension, an explanation about the file format should be included in the reviewer's guide.

Moreover, awareness of the guidance will help in identifying differences in the technical requirements like the file format for FDA is different and specified as below:

*Contains Nonbinding Recommendations*

FDA U.S. FOOD & DRUG ADMINISTRATION

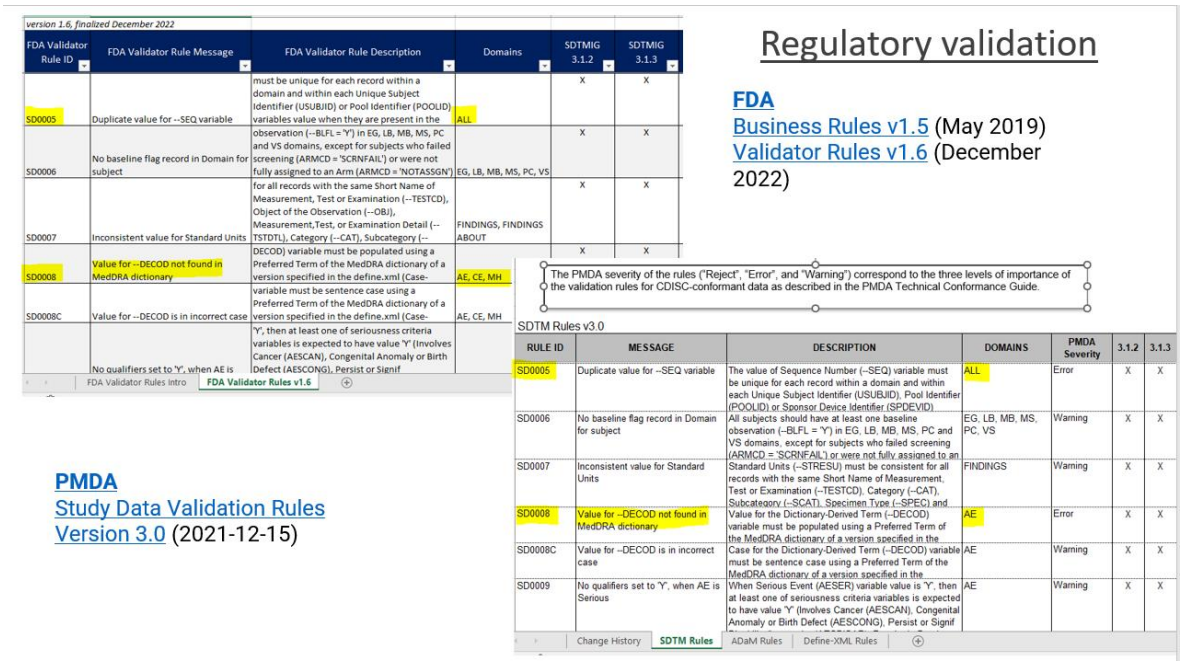
**4.1.2.10 Software Programs**

Sponsors should provide the source code used to create all ADaM datasets, tables, and figures associated with primary and secondary efficacy analyses. Sponsors should submit source code in single byte ASCII text format. Files with MS Windows executable extensions (.cmd, .com, and .exe) should NOT be submitted. For a list of acceptable file extensions, refer to the document entitled *Specifications for File Format Types Using eCTD Specifications*.[31]

**c. Study Data Validation Rules**

Study data validation rules help to ensure that the study data are compliant, useful, and will support meaningful review and analysis.
Please note that when submitting electronic study data to the PMDA via the gateway system, only one version of the validation rules must be selected for a single application, even if it involves multiple studies.

Below figure shows the difference between validation rules between FDA and PMDA versions. For e.g., the Rule SD0008 is applicable to AE, CM & MH domain as per the FDA validation rules version 1.6 and needs to be taken into consideration when writing specifications.




Display 4: Difference between FDA & PMDA regulatory validation rules.

We also need to keep in mind these rules are constantly evolving and the below latest PMDA validation rules version 4.0 shows the SD0008 rule rom SDTM is now same as FDA. Hence it is important to be aware of the change history before applying them to the data analysis process or copying over programs from sister studies where older rules were applicable and need update.

The PMDA severity of the rules ("Reject", "Error", and "Warning") correspond to the three levels of importance of the validation rules for CDISC-conformant data as described in the PMDA Technical Conformance Guide.

SDTM Rules v4.0

| RULE ID | MESSAGE | DESCRIPTION | DOMAINS | PMDA Severity | 3.1.2 | 3.1.3 | 3.2 | 3.3 |
|---|---|---|---|---|---|---|---|---|
| SD0005 | Duplicate value for --SEQ variable | The value of Sequence Number (--SEQ) variable must be unique for each record within a domain and within each Unique Subject Identifier (USUBJID) or Pool Identifier (POOLID) variables value when they are present in the domain. Exclusions are DE, DO, DT, DU, and DX domains. | ALL | Error | X | X | X | X |
| SD0006 | No baseline flag record in Domain for subject | All subjects should have at least one baseline observation (--BLFL = 'Y') in EG, LB, MB, MS, PC and VS domains, except for subjects who failed screening (ARMCD = 'SCRNFAIL') or were not fully assigned to an Arm (ARMCD = 'NOTASSGN') or were not treated (ACTARMCD = 'NOTTRT'). | EG, LB, MB, MS, PC, VS | Warning | X | X | X | X |
| SD0007 | Inconsistent value for Standard Units | Standard Units (--STRESU) must be consistent for all records with the same Short Name of Measurement, Test or Examination (--TESTCD), Object of the Observation (--OBJ), Measurement,Test, or Examination Detail (--TSTDTL), Category (--CAT), Subcategory (--SCAT), Specimen Type (--SPEC) and Method of Test or Examination (--METHOD). | FINDINGS, FINDINGS ABOUT | Warning | X | X | X | X |
| SD0008 | Value for --DECOD not found in MedDRA dictionary | Value for the Dictionary-Derived Term (--DECOD) variable must be populated using a Preferred Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive). | AE, CE, MH | Error | X | X | X | X |
| SD0008C | Value for --DECOD is in incorrect case | Case for the Dictionary-Derived Term (--DECOD) variable must be sentence case using a Preferred Term of the MedDRA dictionary of a version specified in | AE, CE, MH | Warning | X | X | X | X |

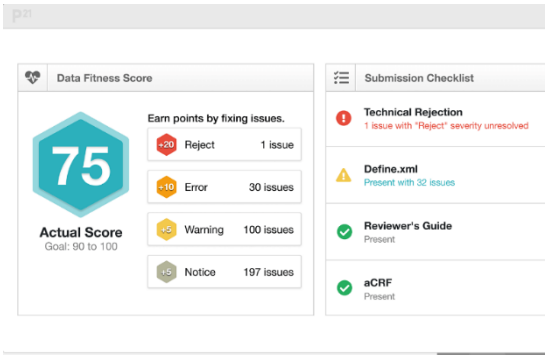
## 3. Study Data Traceability

A vital component of a regulatory review is an understanding of the provenance of the data (e.g., traceability of the sponsor's results back to the CRF data). Traceability enables the reviewer to understand the relationships between the analysis results (tables, listings, and figures in the study report), analysis datasets, tabulation datasets, and source data. It is necessary to make sure Study data traceability is not compromised at any step of data analysis. If the reviewer is unable to trace study data from the data collection of subjects participating in a study to the analysis of the overall study data, then the regulatory review of a submission may be compromised.

### Pinnacle 21- a final checkpoint

Pinnacle 21, also known as OpenCDISC Validator, provides great compliance checks against CDISC outputs like SDTM, ADaM and Define.xml. The validation report provides detailed list of issue summary across datasets along with a summary report shown below.

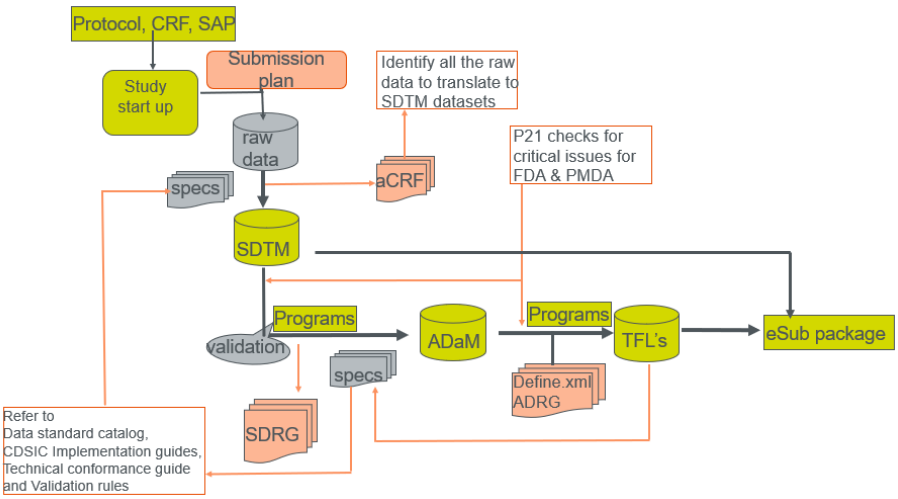
- This report acts as a last stop before datasets head off for their health authority review.
- It checks the data for compliance with CDISC standards, controlled terminology, and dictionaries like MedDRA and WHODrug.
- Additionally, it evaluates data against each agency's business rules to prevents issues that might delay the submission.
- Excellent validation tool for data quality and is suggested to be used in parallel to data analysis.




Display 5: An online example of P21 summary (Validation | Pinnacle 21)

## Summary and Recommendation

Submitting standardized study data to regulatory authorities is a critical and complex process and it requires strategy, planning, and communication. Sponsors need to proactively identify and prepare for challenges throughout the submission life cycle.
The recommendation is regardless of the plan for submission or knowledge of Health authority to be submitted to, the plan and process of submission needs to be followed from the beginning of data analysis and throughout every programming milestone. This does involve additional effort but will help to prepare for compliant submission package in a lean, resourceful, and cost-effective way. The below suggested easy-to-navigate flowchart highlights the basic understanding of this process and how the optimization method at each step helps in preparing for the submission and allows to have all the guidance document and regulatory requirements for the analysis.




Display 6: flowchart with proactive implementation of Submission requirements at every programming milestone.

## Disclaimer

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

References:

New Drug Application (NDA) | FDA

New Drug Review with Electronic Data | Pharmaceuticals and Medical Devices Agency (pmda.go.jp)

FDA_Data Stds Catalog_4.19.2023 (v9.1).xlsx (live.com)

Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications Guidance for Industry | FDA

Documentation | Pinnacle 21

Lessons Learned for Successful e-Study Data Submission to PMDA toward the End of Transitional Period (lexjansen.com)

Industry+Experiences+Submitting+Standardised+Study+Data+to+Regulatory+Authorities.pdf (phuse.s3.eu-central-1.amazonaws.com)

Awareness from Electronic Data Submission to PMDA and FDA - Lesson & Learnt from Hands-on Experiences (pharmasug.org)

Challenges and solutions for e-data submission to PMDA even after submission to FDA | Request PDF (researchgate.net)

REQUIRED (lexjansen.com)

Effective Approach for ADaM Submission to FDA and PMDA (pharmasug.org)

## CONTACT INFORMATION

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

Priyanka Pujari
Parexel International, United Kingdom
Navigation House South Quay Drive,
Sheffield, S2 5SU
priyanka.pujari@parexel.com