STUDY DATA VALIDATION – MEETING NEW REQUIREMENTS FROM INDUSTRY AND AUTHORITIES

Author: Michael Beers, Pinnacle 21

PHUSE SDE Boston 2023

DISCLAIMER

- The views and opinions presented here represent those of the speaker and should not be considered to represent advice or guidance on behalf of the regulatory agencies or standards development organizations.

AGENDA

- Overview of Study Data Validation
  - CDISC Conformance rules
  - Regulatory agency rules
  - Other
- Ensuring Compliance
  - How to be compliant
  - Dispositioning issues
- Areas for Improvement - Gaps in Validation
  - Therapeutic Area Specific data
  - Custom domains
  - Impact of gaps in validation

OVERVIEW OF STUDY DATA VALIDATION

DEFINITION OF STUDY DATA VALIDATION

8. Study Data Validation and Traceability

8.1 Definition of Study Data Validation

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

*FDA's Study Data Technical Conformance Guide, October 2022

WHY VALIDATION IS IMPORTANT

- Goal is to make drugs available ASAP
- Increased automation is needed for this goal
- Manual review process is inefficient
  - Data review time is limited
  - Data quality evaluation is difficult
- Automation relies on high quality data
  - Value of standard reporting and analysis for Reviewers increases demand for data quality
- Standardized data allows early evaluation and enforcement of data quality
- Validation is the key to ensure standardized data

TYPES OF VALIDATION RULES

8.2 Types of Study Data Validation Rules

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

2. FDA eCTD Technical Rejection Criteria for Study Data that assess conformance to the standards listed in the Catalog (See section 7.1, section 8.2.2, and Appendix F).

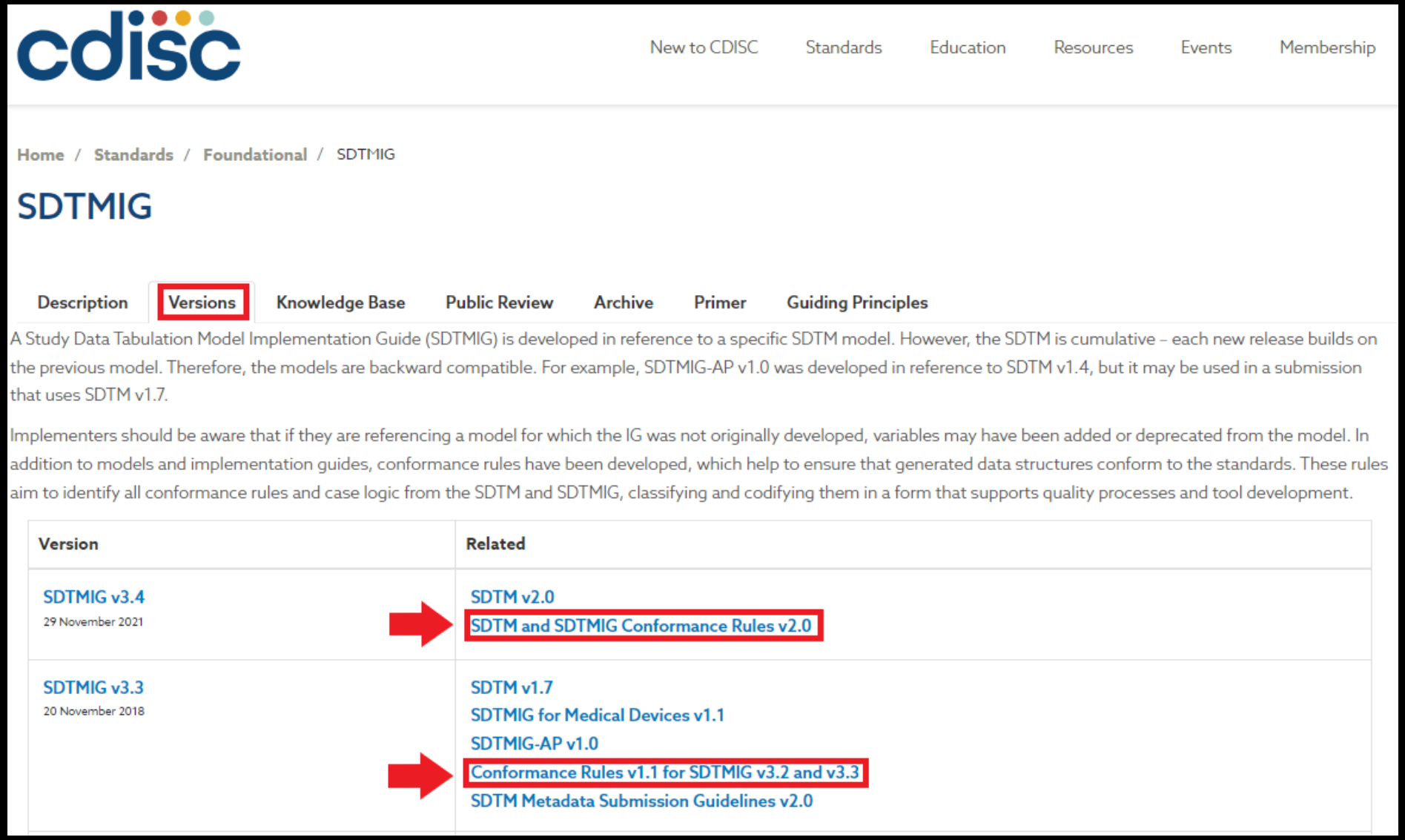
3. FDA Business and Validator rules to assess that the data support regulatory review and analysis.

8.2.1 FDA Business and Validator Rules

FDA Business Rules describe the business requirements for regulatory review to help ensure that study data are compliant and useful and support meaningful review and analysis. The list of business rules will grow and change with experience and cross-center collaborations. All business rules should be followed where applicable. The business rules are accompanied with validator rules which provide details regarding FDA's assessment of study data for purposes of review and analysis. The FDA Validator Rules also represent the latest understanding of what best supports regulatory review. The Study Data Standards Resources webpage page provides links to the currently available FDA Business and Validator rules.[67]

*FDA's Study Data Technical Conformance Guide, October 2022

LOCATION OF CDISC CONFORMANCE RULES



cdisc

New to CDISC | Standards | Education | Resources | Events | Membership

Home / Standards / Foundational / SDTMIG

SDTMIG

Description | Versions | Knowledge Base | Public Review | Archive | Primer | Guiding Principles

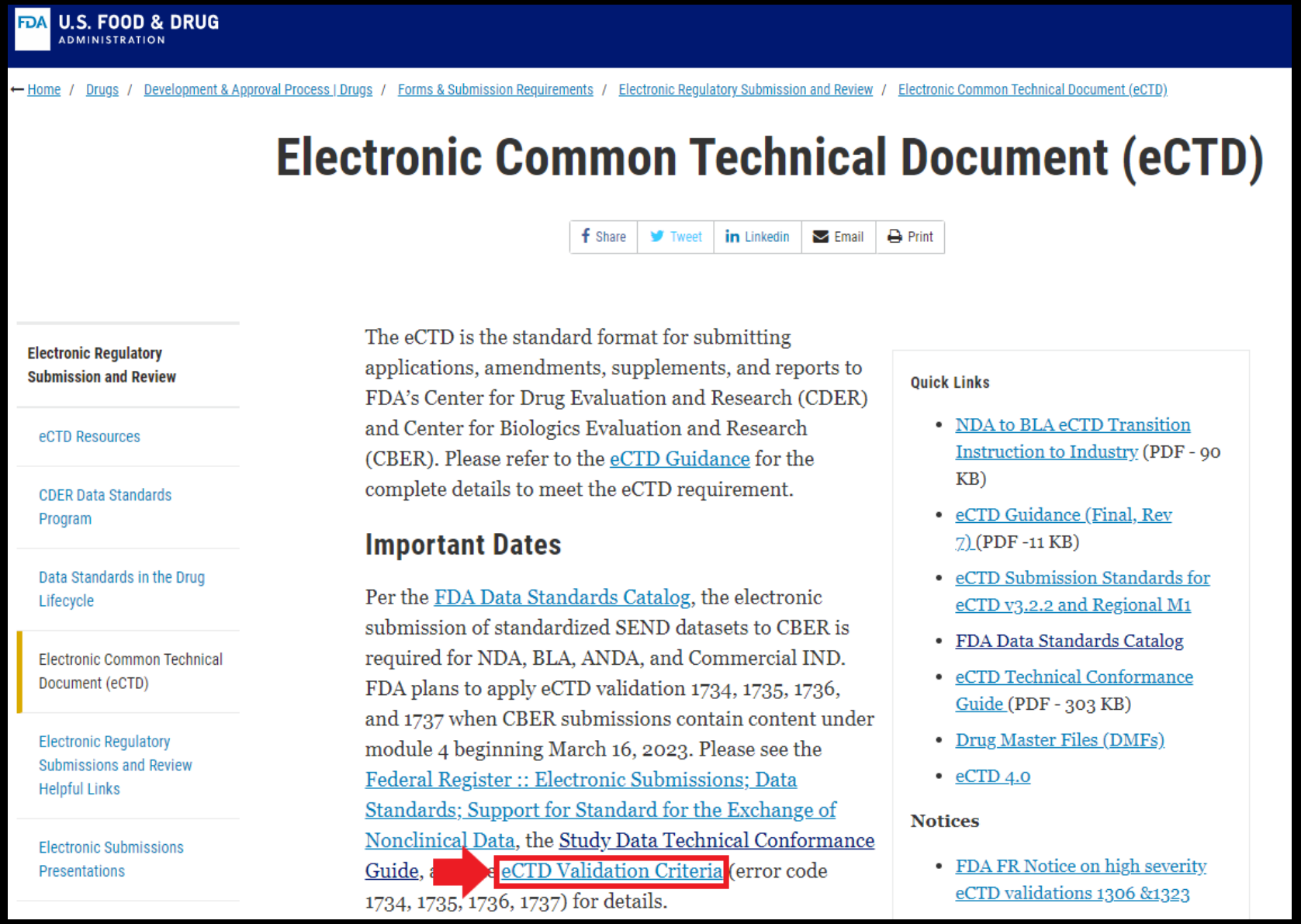
A Study Data Tabulation Model Implementation Guide (SDTMIG) is developed in reference to a specific SDTM model. However, the SDTM is cumulative – each new release builds on the previous model. Therefore, the models are backward compatible. For example, SDTMIG-AP v1.0 was developed in reference to SDTM v1.4, but it may be used in a submission that uses SDTM v1.7.

Implementers should be aware that if they are referencing a model for which the IG was not originally developed, variables may have been added or deprecated from the model. In addition to models and implementation guides, conformance rules have been developed, which help to ensure that generated data structures conform to the standards. These rules aim to identify all conformance rules and case logic from the SDTM and SDTMIG, classifying and codifying them in a form that supports quality processes and tool development.

Version	Related
SDTMIG v3.4 29 November 2021	SDTM v2.0 SDTM and SDTMIG Conformance Rules v2.0
SDTMIG v3.3 20 November 2018	SDTM v1.7 SDTMIG for Medical Devices v1.1 SDTMIG-AP v1.0 Conformance Rules v1.1 for SDTMIG v3.2 and v3.3 SDTM Metadata Submission Guidelines v2.0

P21

LOCATION OF FDA ECTD REJECTION CRITERIA



U.S. FOOD & DRUG ADMINISTRATION

Home / Drugs / Development & Approval Process | Drugs / Forms & Submission Requirements / Electronic Regulatory Submission and Review / Electronic Common Technical Document (eCTD)

Electronic Common Technical Document (eCTD)

Share | Tweet | Linkedin | Email | Print

Electronic Regulatory Submission and Review

- eCTD Resources
- CDER Data Standards Program
- Data Standards in the Drug Lifecycle
- Electronic Common Technical Document (eCTD)
- Electronic Regulatory Submissions and Review Helpful Links
- Electronic Submissions Presentations

The eCTD is the standard format for submitting applications, amendments, supplements, and reports to FDA's Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER). Please refer to the eCTD Guidance for the complete details to meet the eCTD requirement.

Important Dates

Per the FDA Data Standards Catalog, the electronic submission of standardized SEND datasets to CBER is required for NDA, BLA, ANDA, and Commercial IND. FDA plans to apply eCTD validation 1734, 1735, 1736, and 1737 when CBER submissions contain content under module 4 beginning March 16, 2023. Please see the Federal Register :: Electronic Submissions; Data Standards; Support for Standard for the Exchange of Nonclinical Data, the Study Data Technical Conformance Guide, a[illegible] eCTD Validation Criteria (error code 1734, 1735, 1736, 1737) for details.

Quick Links

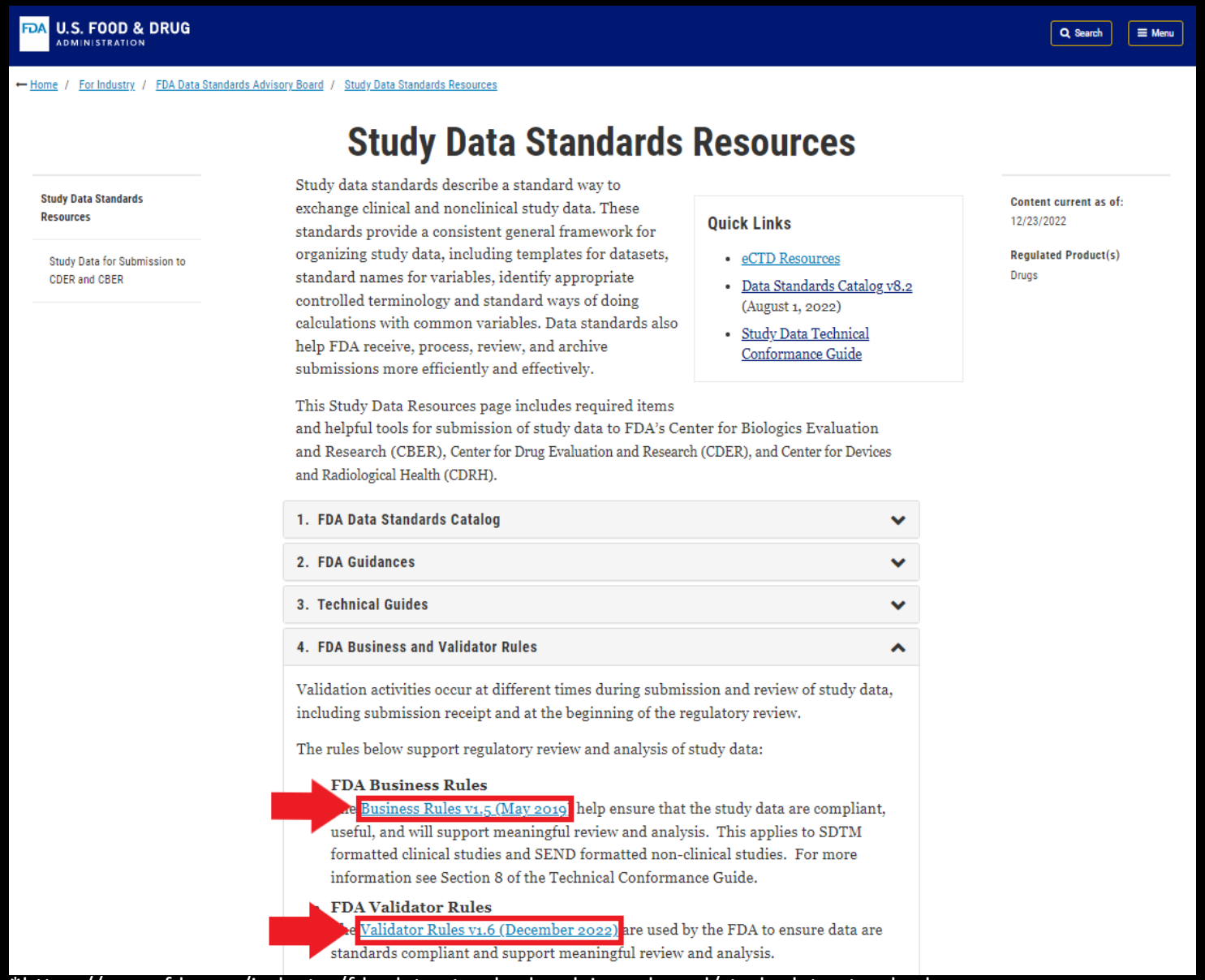
- NDA to BLA eCTD Transition Instruction to Industry (PDF - 90 KB)
- eCTD Guidance (Final, Rev 7) (PDF -11 KB)
- eCTD Submission Standards for eCTD v3.2.2 and Regional M1
- FDA Data Standards Catalog
- eCTD Technical Conformance Guide (PDF - 303 KB)
- Drug Master Files (DMFs)
- eCTD 4.0

Notices

- FDA FR Notice on high severity eCTD validations 1306 &1323

*https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/electronic-common-technical-document-ectd

LOCATION OF FDA BUSINESS/VALIDATOR RULES



U.S. FOOD & DRUG ADMINISTRATION

Search | Menu

Home / For Industry / FDA Data Standards Advisory Board / Study Data Standards Resources

Study Data Standards Resources

Study Data Standards Resources

Study Data for Submission to CDER and CBER

Study data standards describe a standard way to exchange clinical and nonclinical study data. These standards provide a consistent general framework for organizing study data, including templates for datasets, standard names for variables, identify appropriate controlled terminology and standard ways of doing calculations with common variables. Data standards also help FDA receive, process, review, and archive submissions more efficiently and effectively.

Quick Links

- eCTD Resources
- Data Standards Catalog v8.2 (August 1, 2022)
- Study Data Technical Conformance Guide

Content current as of: 12/23/2022

Regulated Product(s) Drugs

This Study Data Resources page includes required items and helpful tools for submission of study data to FDA's Center for Biologics Evaluation and Research (CBER), Center for Drug Evaluation and Research (CDER), and Center for Devices and Radiological Health (CDRH).

1. FDA Data Standards Catalog
2. FDA Guidances
3. Technical Guides
4. FDA Business and Validator Rules

Validation activities occur at different times during submission and review of study data, including submission receipt and at the beginning of the regulatory review.

The rules below support regulatory review and analysis of study data:

FDA Business Rules

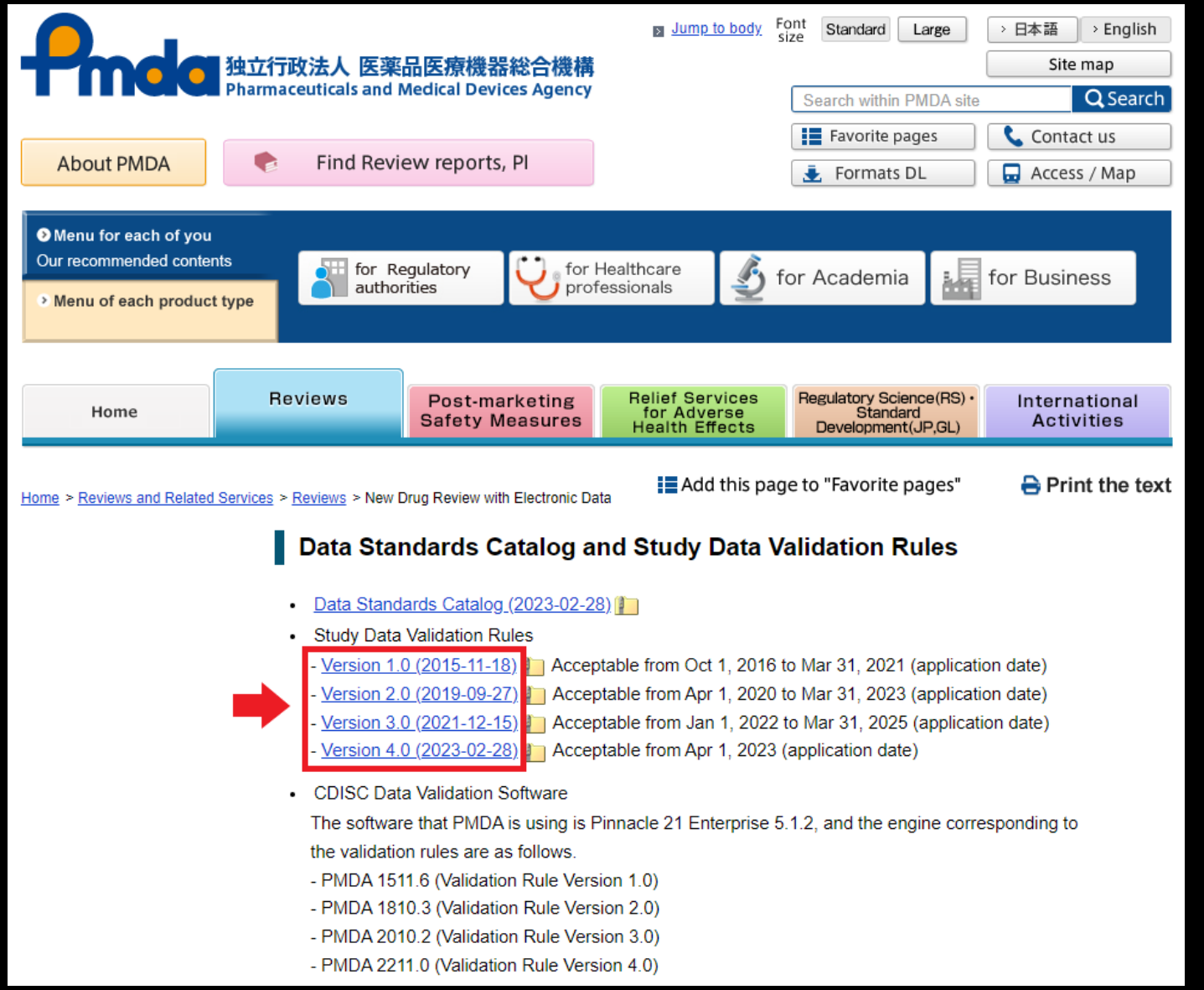
Business Rules v1.5 (May 2019) help ensure that the study data are compliant, useful, and will support meaningful review and analysis. This applies to SDTM formatted clinical studies and SEND formatted non-clinical studies. For more information see Section 8 of the Technical Conformance Guide.

FDA Validator Rules

Validator Rules v1.6 (December 2022) are used by the FDA to ensure data are standards compliant and support meaningful review and analysis.

*https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources

LOCATION OF PMDA VALIDATION RULES

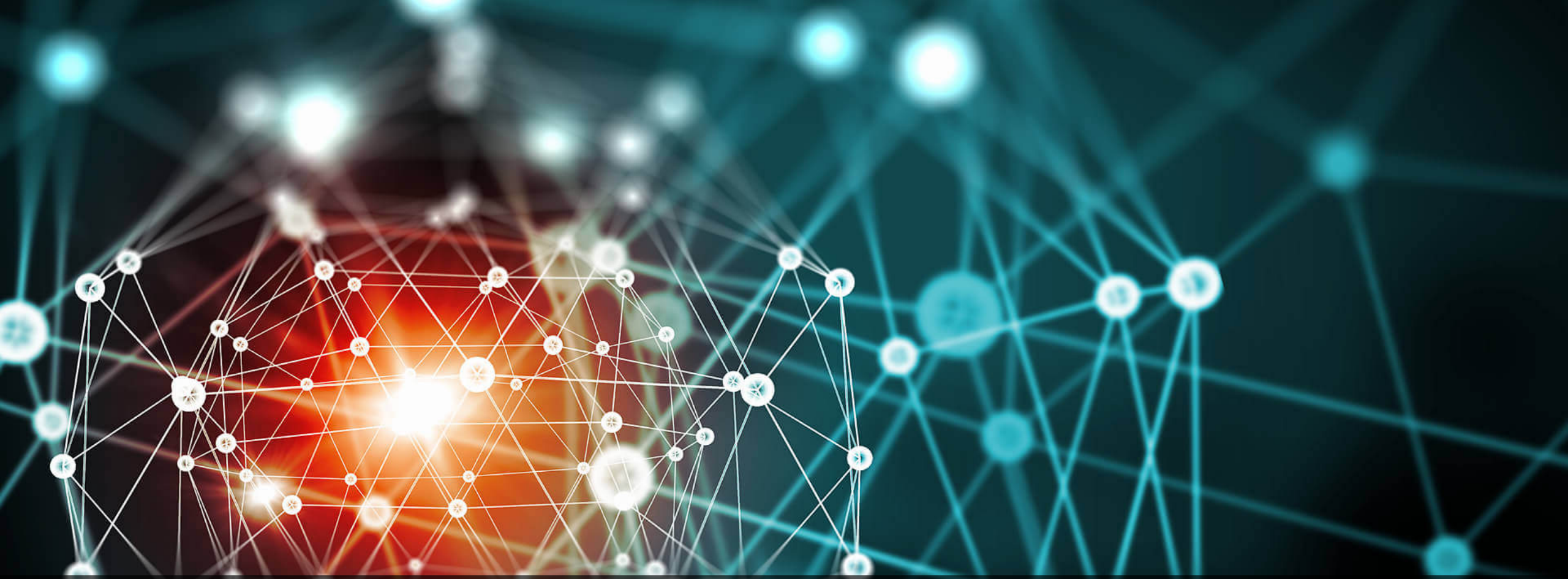


*https://www.pmda.go.jp/english/review-services/reviews/0002.html

OTHER SOURCES OF VALIDATION RULES

- Other FDA requirements
  - Bioresearch Monitoring (BIMO) Technical Conformance Guide
    - https://www.fda.gov/media/85061/download
- Domain-specific mapping guidance from other organizations
  - Uppsala Monitoring Centre's "How to use WHODrug for compliance with CM domain in the CDISC SDTM standard"
    - https://who-umc.org/media/2940/how-to-use-whodrug-for-compliance-with-cm-domain-in-the-cdisc-sdtm-standard-march-2017.pdf
    - This document describes how to map WHODrug data to the CM (and SUPPCM) SDTM domain.
    - This is important because CDISC does not provide much specific guidance on the mapping of this data.
    - While we don't have rules specifically checking CM/SUPPCM against this document, the guidance listed in the document factors into the rule algorithms

P21



ENSURING COMPLIANCE

ENSURING COMPLIANCE

FDA Validation Requirements

8.2.2 Support on Data Validation Rules

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

*FDA's Study Data Technical Conformance Guide, October 2022

ENSURING COMPLIANCE

PMDA Validation Requirements

3.6.1 Validation of data conforming to the CDISC standards

The PMDA will perform validation of data that conform to the CDISC standards using Pinnacle 21 Enterprise.

The rules used for validation have been classified by the level of importance taking into consideration the characteristics of each rule and based on conformity to the standards, ease of use of the data in review, quality of the clinical study data which the PMDA should know beforehand, and future uses of the clinical study data by the PMDA. The levels of importance are shown below.

(a) Rules which, if violated, will cause the review to be suspended until corrections have been made

Very basic rules such as the presence/absence of necessary datasets for each clinical study

(b) Rules which, if violated, will require an explanation

In many cases, these rules are clearly stated in each standard and implementation guide, and if violated, the applicant should explain in the reviewer's guide about the reason for the violation and the reason why it is not possible to correct it.

(c) Rules which, even when violated, will not necessarily require any explanation

The reason for the violation possibly is requested separately for the above (c) from the perspective of the quality of the clinical study data.

Details of the environment in which the PMDA performs validations, individual rules and their importance are published on the PMDA's website. Before submitting electronic study data, the applicant should perform a validation, and should take the necessary action for any violations that are identified. Please keep in mind that these rules may be revised and that validation should always be performed after confirming the latest information. If any rules are revised, the content of the revision will be available for a certain period before the revised rules are applied.

*PMDA's Technical Conformance Guide on Electronic Study Data Submissions, April 2022

OVERVIEW OF PMDA SUBMISSION PROCESS

Approval
Application Date

Drug Development

Submission Preparation

Declare the submission date

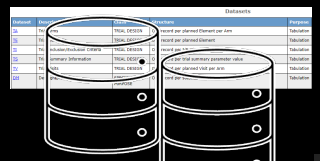
Submit data through Gateway

PMDA validation of submitted materials

PMDA review starts

Sponsors
should run validations beforehand and make sure no results of Reject found

Define



SDTM

ADaM

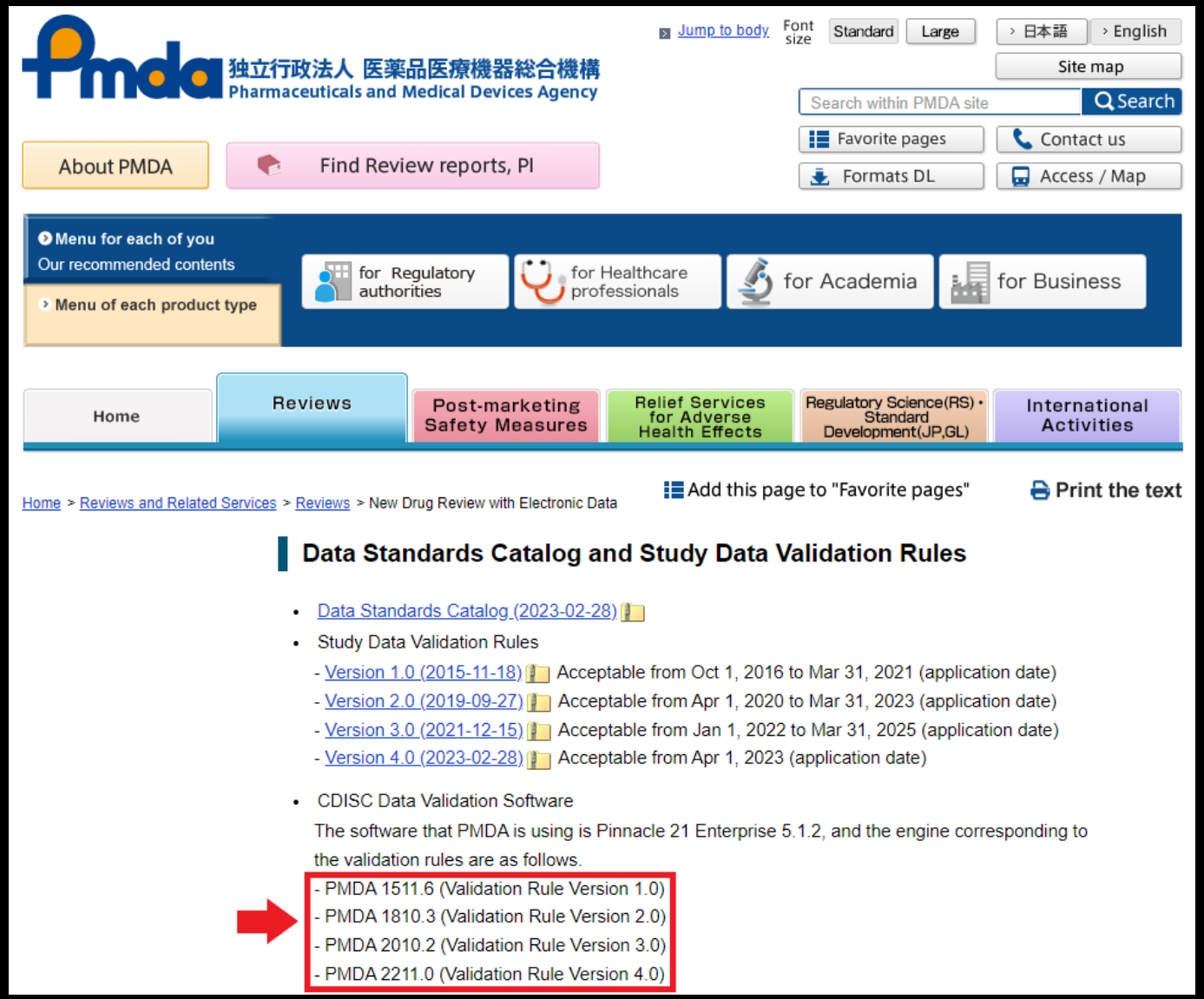
P21

Pre-NDA meeting
To confirm finalized submission package and submission date

PMDA review will not begin when:

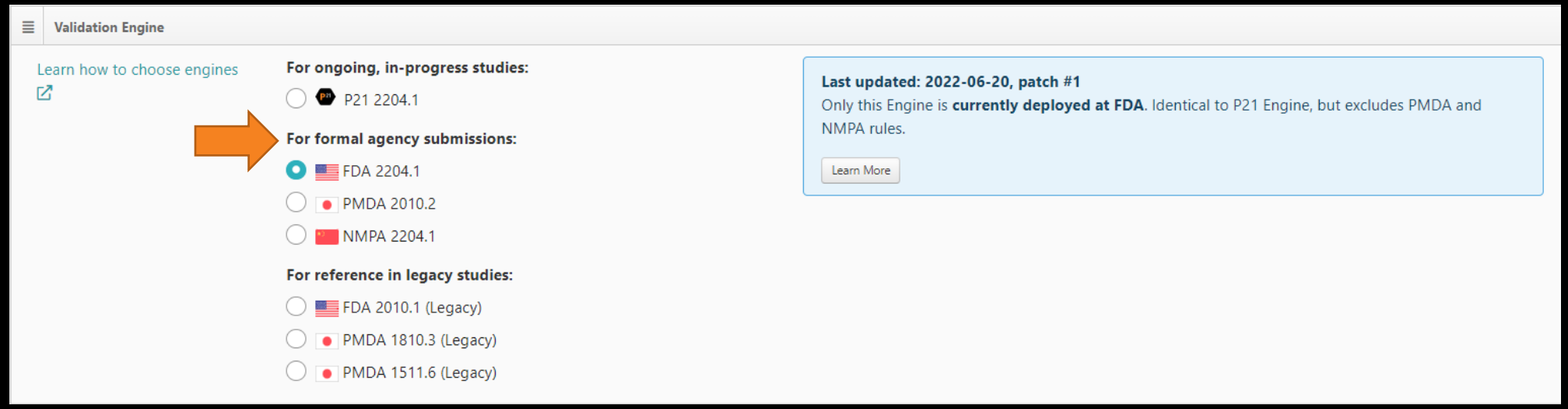
1. Abnormal termination of the validation occurs
2. Validation reports of a study/analysis have not been generated
3. Version of the validation rule not equal between the electronic study data system (Gateway) and Form A
4. Reject is detected

PMDA VALIDATION RULES & SUPPORTED ENGINES



*https://www.pmda.go.jp/english/review-services/reviews/0002.html

AGENCY-SPECIFIC VALIDATION ENGINES



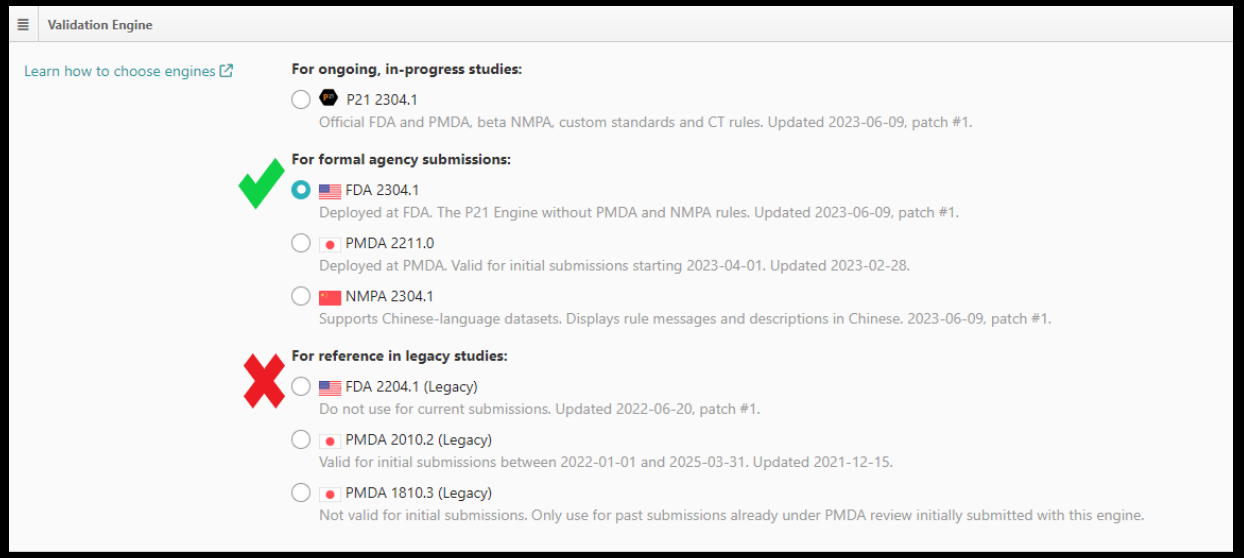
Regulatory agencies manage their own validation criteria

- Example 1 – Differences in supported versions of the standards
  - Agencies may differ in which versions of the standards they will accept
- Example 2 – Differences in validation rules
  - FDA rule SD1071 (Dataset is greater than 5 GB in size)
    - Description: Datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets no larger than 5 GB.
  - PMDA rule SD1142 (Dataset is greater than 5 GB in size)
    - Description: Applicant should consult the PMDA beforehand if the dataset exceeds the file size specified by PMDA.

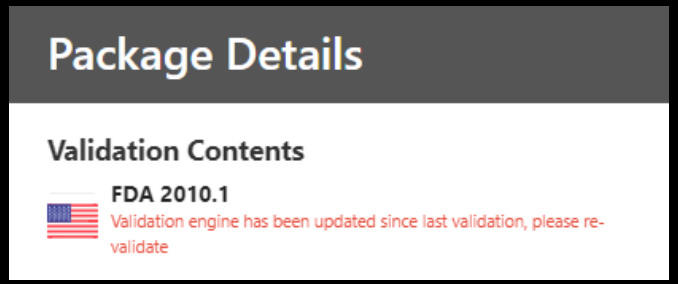
ENSURING COMPLIANCE IN ENTERPRISE

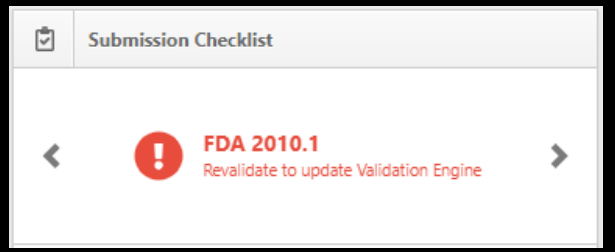
Always Use Latest Validation Engine Available

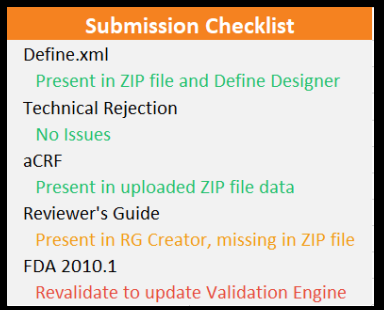
- Use P21 engine, but switch to agency-specific engine prior to submission



- Using an older (Legacy) engine will result in a warning message on multiple screens
  - Don’t ignore these messages







ENSURING COMPLIANCE IN COMMUNITY

Always Use Latest Validation Engine Available

Compatible Engines

Community provides different Validation Engines so that you can stay compliant with regulatory agencies. To do this, simply select which Engine you would like to use from the dropdown. By selecting the appropriate Engine, other dropdowns will be filtered to only show the standards and configurations that are compatible with the agency. For example, you won't find SEND in the standard dropdown for any PMDA Engine because the PMDA does not support SEND.

- **FDA** - Validation engine currently deployed at FDA. Use to prepare for imminent submission.
- **FDA Legacy** - Prior version of FDA engine. Provided for reference only. Do not use for imminent submission.
- **PMDA** - Validation engine for PMDA rules v3.0. Valid for use with applications from January 1, 2022.
- **PMDA Legacy** - Validation engines for older PMDA rules v1.0 and v2.0.
- **NMPA** - Beta validation engine for NMPA. Supports datasets with UTF-8 encoded Chinese characters and displays rule messages and descriptions in Chinese translation.

Incompatible Engines

The Validation Summary tab of the Validation Report will warn you if an incompatible engine has been used. The most common reasons for validation with incompatible engines are:

- Validation was ran using an unsupported configuration file
- A depreciated engine that is no longer compatible by PMDA or FDA was used

If this happens, ensure you are connected to the Internet and that you have the latest version of Pinnacle 21 Community. To see your version, go to "About Pinnacle 21 Community". After you have verified your Internet connection and release version, re-run the validation and choose a non-depreciated Engine from the dropdown and ensure the configuration file you select has not been modified or customized.

For any additional questions, please post on the P21 Community Forum. If you are an Enterprise client, please contact your CSM or email success@pinnacle21.com

ENSURING COMPLIANCE

Dispositioning Your Data Validation Issues

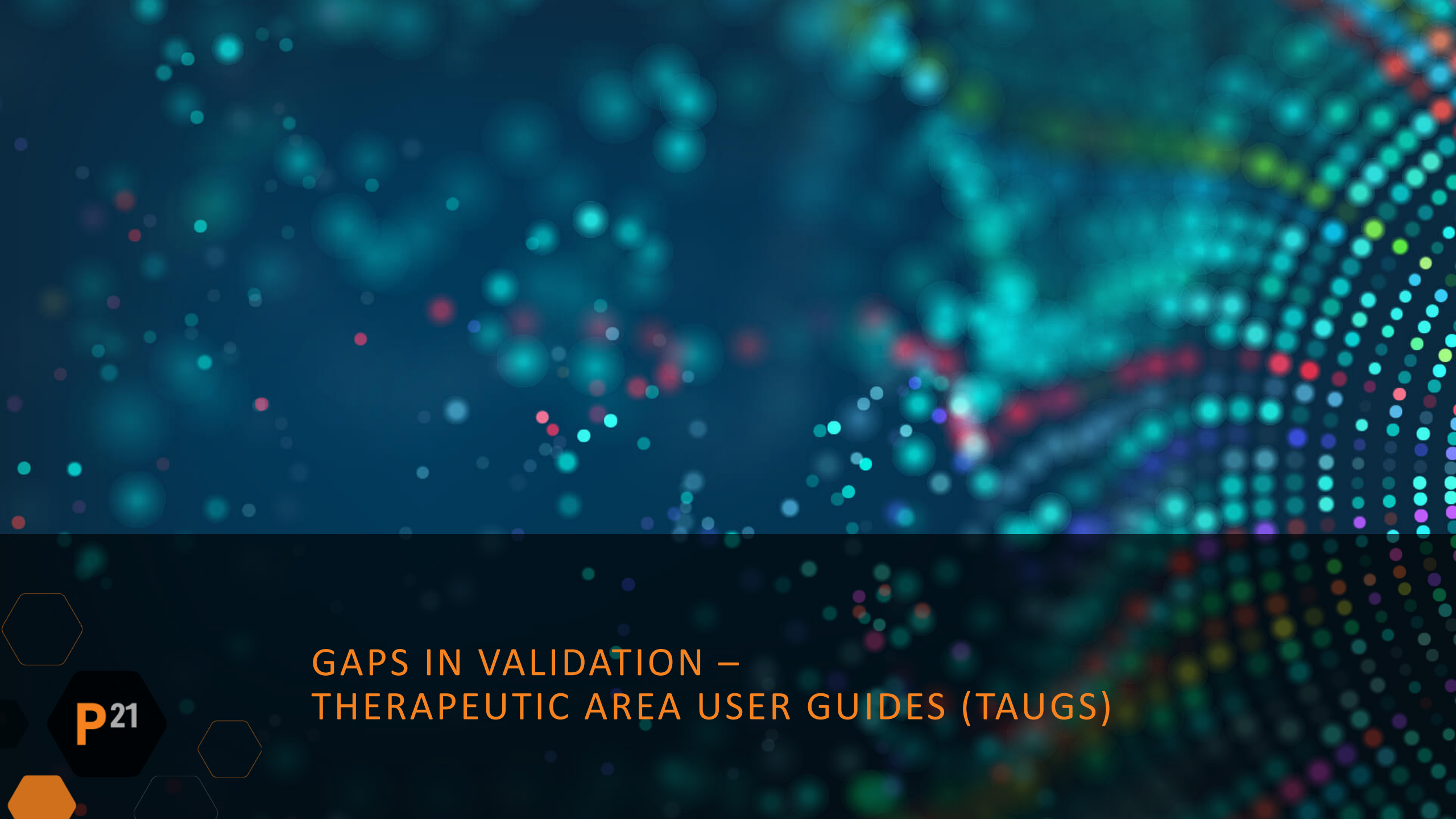
- Common problems when explaining issues in Reviewer's Guides
  - Poor quality issue explanations

Check ID	Diagnostic Message	FDA Severity	Dataset	Count	Explanation
SD0087	RFSTDTC is not provided for a randomized subject	Warning	DM	2	Populated as per source dataset
SD0088	RFENDTC is not provided for a randomized subject	Warning	DM	2	Populated as per source dataset

  - Internal Comments provided instead of actual explanations

Dataset	Diagnostic Message	Severity	Count	Explanation
LB	Inconsistent value for Standard Units		53	These have been discussed with sponsor and confirmed as fine to leave.

  - Not all issues listed/explained in Conformance Summary section of cSDRG
    - Sponsors sometimes only include issues they deem important, not all issues
    - Validation configurations weren't set up correctly/completely



GAPS IN VALIDATION – THERAPEUTIC AREA USER GUIDES (TAUGS)

GAPS IN VALIDATION

TAUG Validation Doesn't Currently Exist

- *Therapeutic Area User Guides (TAUGs) extend the Foundational Standards to represent data that pertains to specific disease areas.*

- There are currently no published validation rules for TAUG implementation

- Also, there is currently no (standard) way to know if:
  - A TAUG should have been used, or
  - A TAUG was used

- For validation purposes, we need answers to these two important questions, but there are issues with each…

GAPS IN VALIDATION

TAUG Validation – Should a TAUG Have Been Used?

How to know if a TAUG should have been used?

- Issue #1 – Use the Trial Summary INDIC (Trial Disease/Condition Indication) parameter?
  - Problem - TS Indication should use SNOMED, but not always implemented correctly

Indication problem examples – Prostate Cancer:

Trial Summary Indication
Carcinoma of prostate (disorder)
PROSTATE CANCER
Metastasis from malignant tumor of prostate
Adenocarcinoma of the prostate
CARCINOMA OF PROSTATE
Adenocarcinoma of prostate
Metastatic prostate cancer

Indication problem examples – Diabetes:

Trial Summary Indication
Type 2 Diabetes mellitus (disorder)
Diabetes mellitus type 2 (disorder)
Diabetes mellitus type 2
TYPE 2 DIABETES MELLITUS
Type II diabetes mellitus
TYPE 1 DIABETES
DIABETIES
Diabetes mellitus type 1 (disorder)
Type 2 diabetes mellitus with moderate to severe chronic kidney disease

GAPS IN VALIDATION

TAUG Validation – Should a TAUG Have Been Used?

How to know if a TAUG should have been used?

- Issue #2 – If a study indication doesn't have a TAUG, should one have been used anyway?
  - For example, should the PAIN TAUG only be used for pain studies, or should they be referenced for any study that collects data on pain?
  - Should a study potentially reference multiple TAUGs depending on the types of data collected?

GAPS IN VALIDATION

TAUG Validation – Was a TAUG Used?

How to know if a TAUG was used?

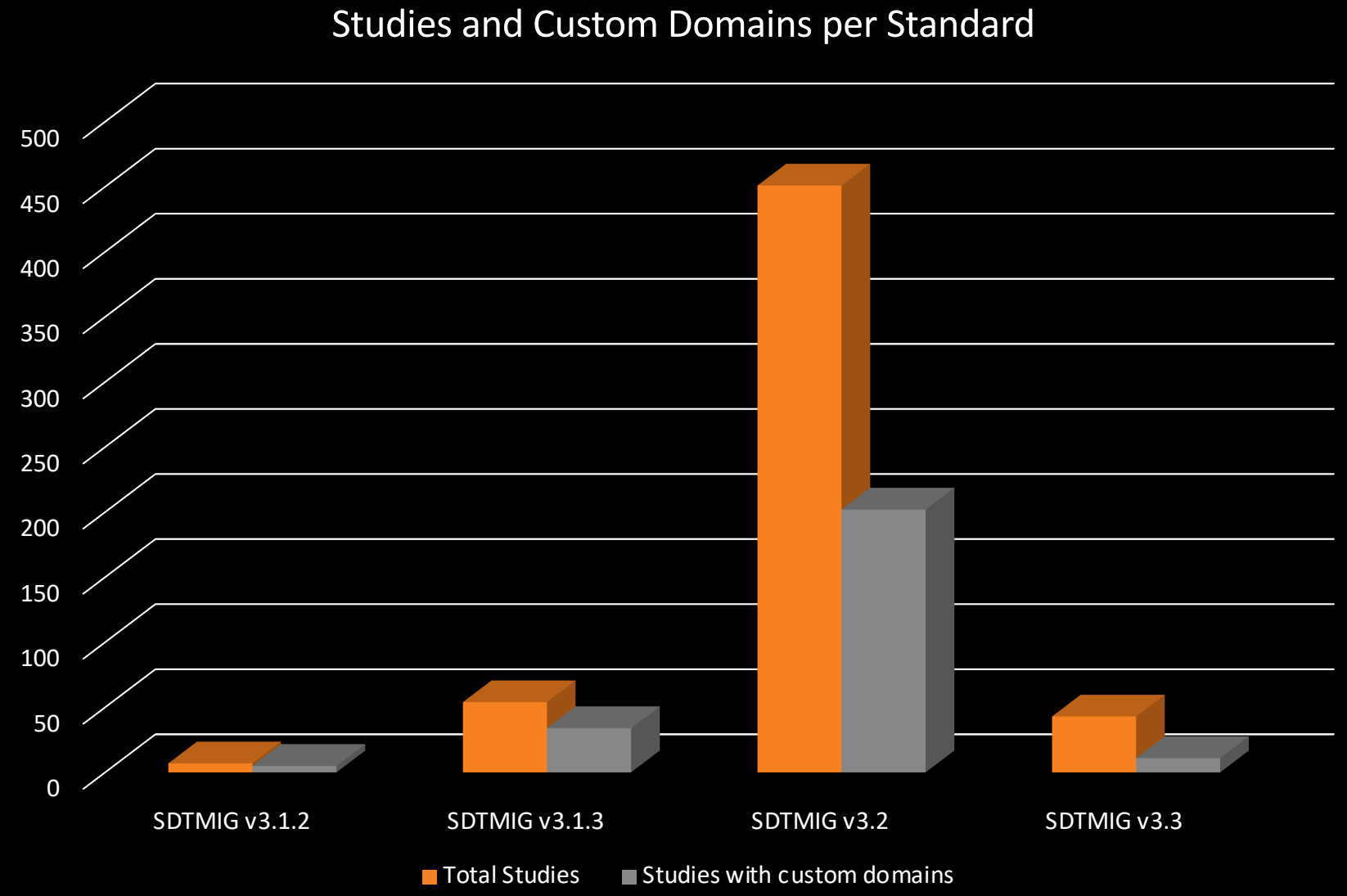
- Trial Summary parameter to indicate TAUG

FDA Desired - Clinical	TSPARMCD	TSPARM	FDA Notes
Conditional	CTAUG	CDISC Therapeutic Area User Guide	If applicable. the value should be the exact listing as in section 5.2 of the Technical Conformance Guide. Use as many rows as needed.

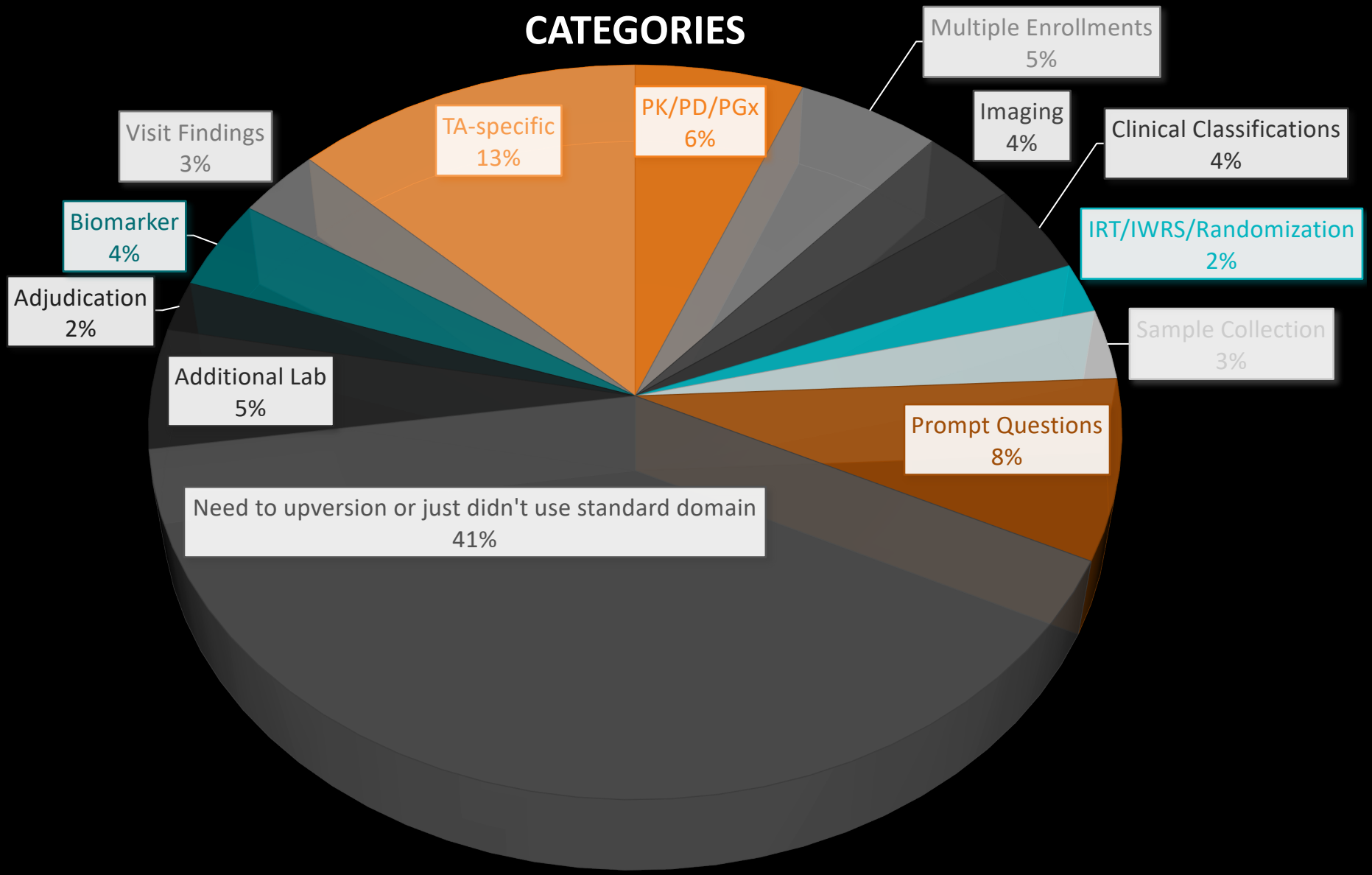
- This is great, however the parameter is listed as *Conditional*

GAPS IN VALIDATION – CUSTOM DOMAINS

CUSTOM DOMAINS USED PER VERSION OF SDTMIG



CUSTOM DOMAINS USED PER VERSION OF SDTMIG



CUSTOM DOMAINS – MULTIPLE ENROLLMENTS

Actual custom domains used in industry:

Custom Domains - Demographics for Multiple Enrollments	
Dataset	**SDTMIG**
XD - Demographics Support	3.3
DC - Demographics as Collected	3.2
XM - Multiple Participation	3.2
XM - Demographics (Multiple Screenings)	3.2
XD - Demographics for Re-Screened Subjects	3.2
XD - Demographics	3.2
XM - Additional Demographics	3.2
XD - Demographics for Re-Screened Subjects	3.2
XM - Demographics Re Screened Subjects	3.2
DC - Demographics as Collected	3.2
XI - Previous Subject Identifiers	3.2
XI - Previous Subject Identifiers	3.2

Regulatory Guidance:

DM Domain (Demographics)

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

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

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

*FDA's Study Data Technical Conformance Guide, October 2022

CUSTOM DOMAINS – ADDITIONAL LAB DATA

Actual custom domains used in industry:

Custom Domains - Additional Lab Data	
Dataset	**SDTMIG**
XB - Laboratory Test Results (Conventional)	3.3
ZT - Other Test	3.3
XB - Laboratory CRF Data/Local Laboratory Data	3.2
XB - Laboratory Test Results (Conventional)	3.2
XL - Non-Safety Laboratory Results	3.2
ZL - Laboratory Test Results - US Units	3.2
XL - Laboratory Test Results XL	3.2
XL - Laboratory Test Results XL	3.2
XD - Other Tests Results	3.2
XF - Hormones	3.2
ZL - Laboratory Test Results - US Units	3.1.3
ZM - Lab Miscellaneous	3.1.3

Regulatory Guidance:

4.1.1.3 SDTM Domain Specifications

LB Domain (Laboratory)

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

*FDA's Study Data Technical Conformance Guide, October 2022

FAQs on Electronic Study Data Submission

4. Questions on CDISC-conformant electronic study data

A: The use of SI units is required for all variables and parameters of test results to be stored in the Findings class domain of the SDTM dataset, as long as SI units are applicable. However, it is acceptable to store only the data in mmHg in the SDTM dataset without storing the converted data in SI unit for test results (e.g. blood pressure) collected in mmHg as conventional unit.

When the original data and the converted data in SI units are stored together in the SDTM dataset, store the converted data into "--STRESC" (or "--STRESN" if necessary) and the original data into "--ORRES". Also, in the reviewer's guide or dataset definition document, describe how each datum (original and converted) is stored, as well as the conversion equation.

In the case of multicenter or multiregional studies, etc., where different measurement units are used for capturing data by center or region, data in multiple units could be obtained for a single parameter (laboratory test). In such cases, it is possible to store data in uniform units other than SI units into "SUPP--" if necessary.

*PMDA's FAQs on Electronic Study Data Submission, June 2022

CDISC Guidance:

Standardized Lab Units

Regulatory Strategy

CDISC does not provide a specific solution when a sponsor is asked to provide both. If a Sponsor anticipates an FDA request for conventional units, a common solution is to produce two versions of the affected datasets (e.g., LB): One with standardized results in SI units (i.e., value in LBSTRESU is an SI unit) and one with standardized results in US conventional units (the value in LBSTRESU is a conventional unit).

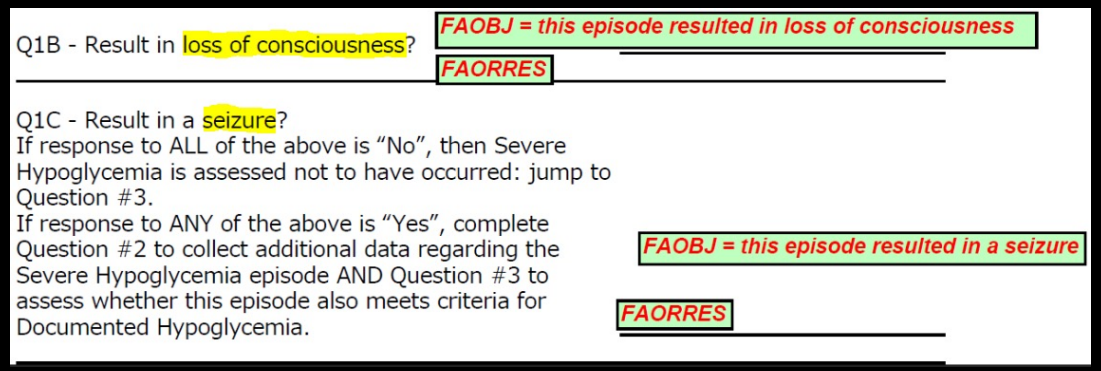
*https://www.cdisc.org/kb/articles/standardized-lab-units

IMPACT OF GAPS IN VALIDATION

IMPACT OF GAPS IN VALIDATION

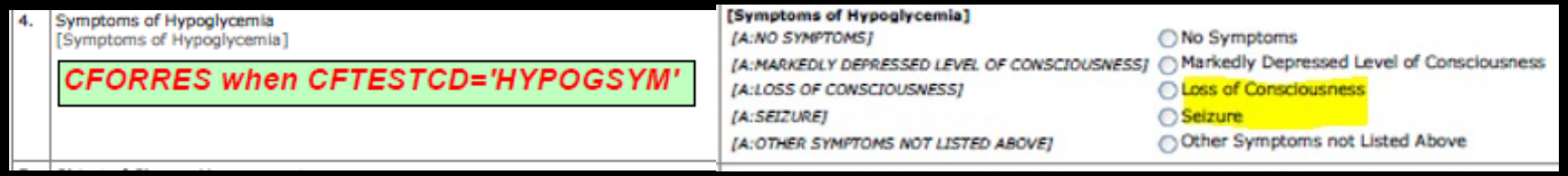
Lack of Validation for TAUG Implementation Results in Inconsistencies

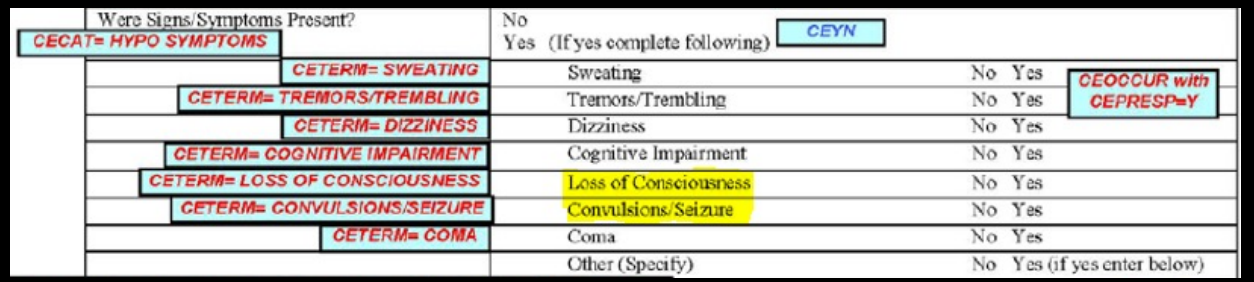
- Example: Diabetes indication
  - Hypoglycemia signs/symptoms data – same data mapped and collected inconsistently



Study A – Mapped to the FA domain

Study B – Mapped to the CF domain



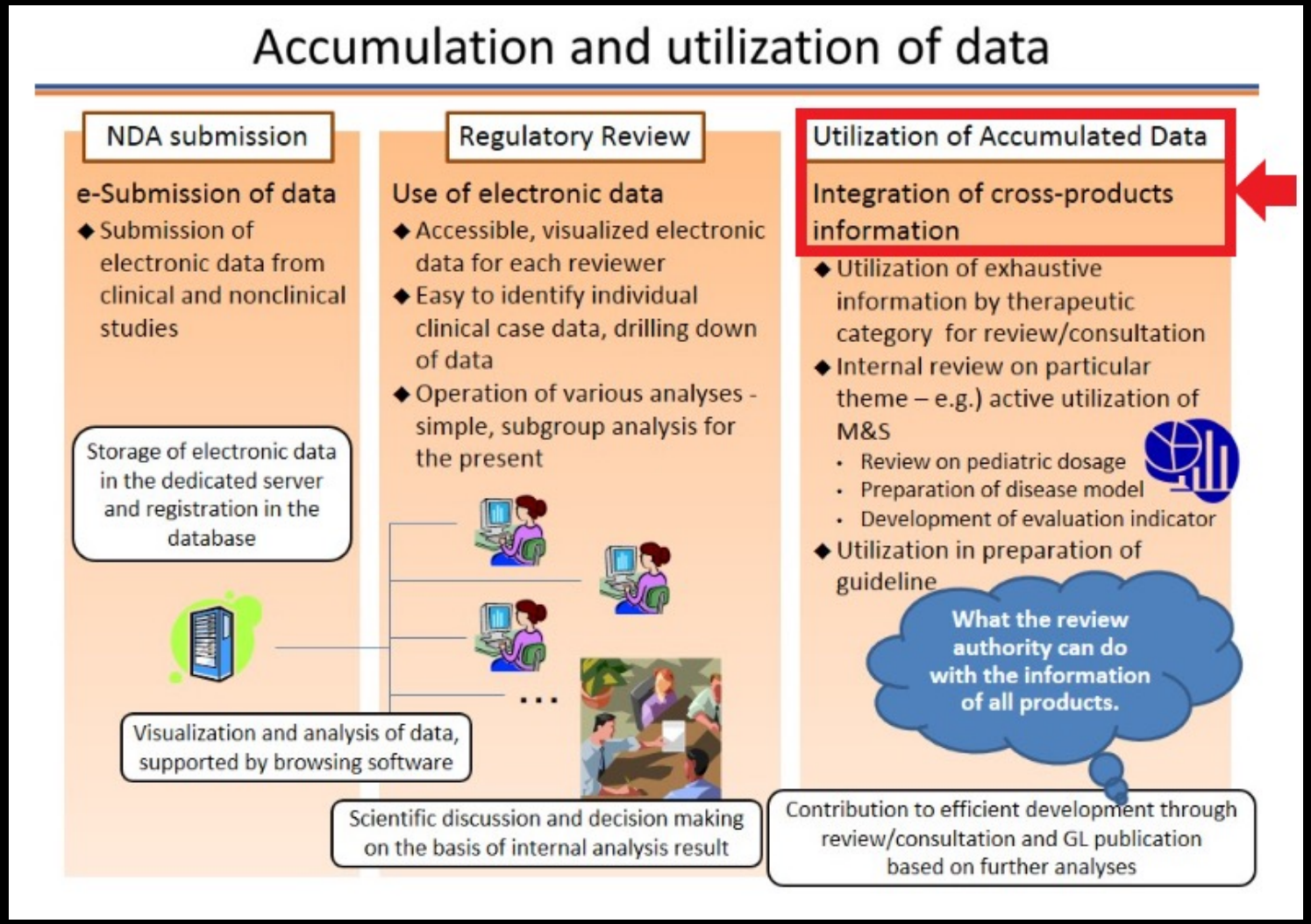


For reference, this is how it is collected and mapped in the TAUG

IMPACT OF GAPS IN VALIDATION

- Validation is too general on custom domains
  - Domain-specific checks not run and domain-specific codelists not checked properly
  - Example: Additional Lab Data
    - In SDTMIG v3.3, there are 155 validation rules run on the LB domain, but there are only 144 validation rules run on a custom findings domain.
    - Variables such as –TESTCD and –TEST would not be checked against the LBTESTCD and LBTEST codelists, even though the controlled terminology should have been used.
- Validation might not even occur on custom domains
  - Custom special-purpose domains, such as for Multiple Enrollments
  - This results in an unvalidated dataset being submitted to the regulatory agency
- Similar data across the industry mapped inconsistently

IMPACT OF GAPS IN STANDARDIZED DATA



*https://www.pmda.go.jp/english/review-services/reviews/0002.html

CONCLUSION

- Ensuring compliance to regulatory requirements is a critical part of preparing for a submission
  - Pinnacle 21, both Enterprise and Community, provide agency-specific engines for this purpose

- To reach actual fully-standardized data across the industry, we need to fill the gaps that currently exist in validation


QUESTIONS?