

Updates on validation of ADaM data

#DS11

PhUSE EU CSS 2018

Max Kanevsky, Sergiy Sirichenko
November 6, 2018

Data consumers and custodians

› Consumers

- Utilize data
- Statisticians, programmers, reviewers, data warehouses, automated analysis tools, sponsors, etc.

› Custodians

- Create data
- Data managers, programmers, EDC systems, vendors, etc.

› Validator

- Communication tool
- Provides executable business rules
- Helps custodians to prepare high quality data
- Helps consumers to ensure quality of data

Data Quality

- › Defined by *intended use*
 - absence of errors which matter
- › All users are different
 - Interested only in data issues directly related to him/her
 - May have supplementary task-specific requirements
- › Sometimes there are conflicting requirements
 - FDA reviewers prefer Conventional US units, PMDA asks for SI units

Source of data validation rules

- › Standards specifications
 - ADaM Model and ADaM IG
 - CDISC ADaM Validation Checks
- › FDA/PMDA business rules
 - high-level review and process related requirements
 - published by the Agencies for public use
- › Data management checks
 - may be promoted to industry-wide rules
- › Tool-specific requirements
 - to know in advance if there are any potential problems
- › Additional requests from users
 - Customs standards and business rules

Implementation

- There are no “exact” implementation of business rules
 - Multiple stakeholders
 - Programming limitations
 - Collapsed rules and split rules
 - Special attributes “Publisher ID” in P21 reports

Search Pinnacle 21 LLC (US) | https://www.pinnacle21.com/validation-rules/adam

PRODUCTS SERVICES COMMUNITY About Us

SDTM SEND ADAM DEFINE-XML MEDICAL DEVICES

Search ALL PMDA

P21/PMDA ID	CDISC ID	Message	Description	Domains	P21 Severity	PMDA Severity
AD0018	18	Variable label mismatch between dataset and ADaM standard	Variable Label in the dataset should match the variable label described in ADaM. When creating a new domain Variable Labels could be adjusted as appropriate. Currently this rule runs if a variable has a Label in the OpenCDISC ADaM metadata.	ADSL	Error	Warning
AD0019	19,20,21,22,23,24,25	Variable subject-population flag value is null	For subject-level character population flag variables: N = no (not included), Y = yes (included). Null values are not allowed.	ADSL	Error	Error

Stages in data standards life-cycle

- › Release of initial version
- › Slow implementation
- › Requirements by FDA/PMDA as a major driver for adoption
- › Development of standardized tools
- › Automation
 - Requires high level of standard compliance
 - Introduces additional business rules for study data
 - Increases users demand for high quality standardized data
 - Extends validation from std. compliance to data quality with focus on content
 - Allows evaluation and enforcement of data quality

3 groups of validation rules

1. Standard compliance
 2. Data Quality with focus on content
 3. Tool-specific to support automated process
- › Weights of each group vary across standards
 - Quite even for SDTM and SEND
 - Rigid structure and robust standard control terminology
 - ADaM validation focuses on standard compliance
 - Flexible structure and limited control terminology
 - › Major source of ADaM validation rules is CDISC

CDISC ADaM Conformance Rules

- The first version published in 2010
- Last updated in 2015
- New version is under public review
 - Will covers ADaM IG 1.0 and 1.1

Check Number	IG Version	ADaM Structure Group	Machine-Testable Failure Criteria	Message Type	Guide	Section	Item	Cited Guidance
1	1.0	ADSL	ADSL dataset does not exist	Error	Model v2.1; ADaM IG v1.0	6; 2.3.1		Model v2.1, Section 6: ADSL and its related metadata are required in a CDISC-based submission of data from a clinical trial even if no other analysis datasets are submitted. ADaM IG v1.0, Section 2.3.1: ADSL and its related metadata are required in a CDISC-based submission of data from a clinical trial even if no other analysis datasets are submitted.
1	1.1	ADSL	ADSL dataset does not exist	Error	Model v2.1; ADaM IG v1.1	6; 2.3.1		Model v2.1, Section 6: ADSL and its related metadata are required in a CDISC-based submission of data from a clinical trial even if no other analysis datasets are submitted. ADaM IG v1.1, Section 2.3.1: ADSL and its related metadata are required in a CDISC-based submission of data from a clinical trial even if no other ADaM datasets are submitted.
2	1.0	ALL:SDTM	A variable is present in ADaM with the same name as a variable present in SDTM but the variables do not have identical labels	Error	Model v2.1; ADaM IG v1.0	4.1.2; 3	4 (General Variable Naming Conventions)	Model v2.1, Section 4.1.2: Any ADaM variable with the same name as an SDTM variable is required to be a copy of the SDTM variable, and its label, attributes, and values cannot be modified. ADaM adheres to the principle of harmonization known as "same name, same meaning, and same values." ADaM IG v1.0, Section 3, Item 4 (General Variable Naming Conventions): Any ADaM variable whose name is the same as an SDTM variable must be a copy of the SDTM variable, and its label, meaning, and values must not be modified. ADaM adheres to a principle of harmonization known as "same name, same meaning, same values."

New format

- › *Check Number* – for traceability across versions of rules
- › *IG Version* – reference to version of standard
- › *ADaM Structure Group* – reference to applied datasets
 - For example, ADSL or ALL:SDTM
- › *Machine – Testable Failure Criteria* – validation message
- › *Message Type* – Error, Warning or Note based on capability of business rule to avoid false-positive messages
- › *Guide* – reference to source document of rule
- › *Section* – reference to section within a source guide
- › *Item* – additional granularity of source
- › *Cited Guidance* – text from source document

Content of conformance rules

- › New ADaM IG 1.1
 - 173 total
 - 20 Warning
 - 11 Notes

- › Updated ADaM IG 1.0
 - Bugs fix
 - Rules for ADAE and TTE were removed

Validation of ADaM from Pinnacle 21

- Community version
 - Based in officially published rules from CDISC, FDA and PMDA
 - P21C 2.2.0 covers ADaM IG 1.0 utilizing the latest rules from CDISC published in 2015
 - Additional rules published by PMDA
 - Consistency checks between define.xml and actual study data, missing checks from ADaM team and the agency-specific rules
- Enterprise version includes a draft implementation of new unfinalized conformance rules from CDISC team
 - Still considered as a draft until a new official release from CDISC, as well as review and approval from PMDA and FDA teams

Updates

› ADaM IG 1.0

- Minor tuning according to a draft of updated rules from CDISC
- Additional validation for *ADAM OTHER* datasets is introduced

› ADaM IG 1.1

- Implemented as *Beta*
- Under review by PMDA (P21E 4.0.0)
- Expansion of ADaM rules from CDISC to other applicable datasets

Request for *ADAM OTHER*

- › P21C 2.2.0 and older versions of P21E identify dataset as BDS
 - If any of PARAM, AVAL, ADT, *ASTDT*, *AENDT* variables are present
- › OCCDS and ADAM OTHER was incorrectly validated as BDS
 - For example, *“Required variable PARAMCD is not present”*
- › Fixing this bug resulted in the validation being skipped for OCCDS and ADAM OTHER
- › Pinnacle 21 clients started requesting that this bug fix is reversed to support its *“off-label”* functionality to perform at least some validation for OCCDS and ADAM OTHER datasets

New validation for ADAM OTHER

- › ADAM OTHER is dataset named “AD**”, but not ADSL, ADAE, BDS or OCCDS
- › ADSL “Core” variables were assigned to ADAM OTHER as Permissible
- › Some global rules were applied
 - For example, AD0041: “*DT does not have the ADaM required SAS Date format”, AD0092: “Inconsistent value for TRTPN”, etc.

ADaM “implicit” rules

- › Official CDISC ADaM rules are limited to IG text
- › Example of Treatment variables
 - ADSL – 41
 - BDS – 13
 - OCCDS – 3

“The treatment variable used for analysis must be included. Typically this would be TRTP, TRTA, TRTxxP, or TRTxxA. See the ADaM Implementation Guide version 1.1 [2] for more details on these variables. Additional dosing variables may also be included” (ADaM OCCDS v1.0)

- DOSEON - Treatment Dose at Record Start
 - DOSCUMA - Cumulative Actual Treatment Dose
 - DOSEU - Treatment Dose Units
- › ADSL and BDS *Treatment* variables may be and actually are used in OCCDS

ADaM rules removed by CDISC

- › Official CDISC ADaM rules are “programmable” and avoid false-positive messages
- › ADAE rules are removed
 - There is no formal rule to identify ADAE datasets
 - P21 assumes '*adae.xpt*' file name
- › TTE rules are removed
 - There is no formal rule to identify TTE datasets
 - P21 uses a Prototype based on presence of CNSR (Censor) variable
- › What we should do?
 - Discussing with FDA and PMDA

What is out of scope in CDISC ADaM rules?

- › Integrated data
- › Data traceability
- › Data quality
- › Tool-specific
- › Analysis-specific

Reducing noise

- › False-positive validation messages (FPVM) may be annoying and disruptive
- › 3 major sources
 1. Bugs
 - Reporting bugs to software developers will speed identification and resolution
 2. Warning messages are expected to have FPVM
 - Tuning of algorithms and better reporting will help
 3. User misinterpretation of real issues as FPVM
 - Education and knowledge sharing173 total
- › P21 team has started utilizing the industry metrics
 - to identify and resolve most common and notorious false-positive validation messages

Top 20 issues

Rule ID	Message	Affected studies	Average Issue Rate
CT2002	Variable value not found in extensible codelist	70.6%	13.8%
AD0047	Required variable is not present	56.5%	32.5%
AD0018	Variable label mismatch between dataset and ADaM standard	53.9%	12.3%
AD0198	Neither AVAL nor AVALC are present in dataset	48.9%	100.0%
AD9999	Dataset not validated	40.2%	100.0%
DD0084	Referenced File is missing	32.7%	100.0%
SD1231	Value level length exceed the length specified in define.xml	25.0%	59.1%
AD0124	Inconsistent value for PARCATy within a unique PARAMCD	23.9%	21.3%
DD0060	Define.xml/CDISC variable Label mismatch	23.9%	100.0%
AD0154	Multiple baseline records exist for a unique USUBJID, PARAMCD, BASETYPE	22.6%	31.9%
SD0037	Value for variable not found in user-defined codelist	22.6%	38.7%
AD0149	Inconsistent value for AVALC	21.7%	2.7%
DD0085	Missing Define XSL	21.4%	100.0%
AD0196	Required Variable value is null	17.2%	11.5%
AD1012	Secondary variable is present but its primary variable is not present	16.8%	41.9%
SD1229	Variable value is null when value level attribute is Mandatory='Yes'	15.8%	21.9%
SD1230	Value level datatype does not match the datatype specified in define.xml	15.4%	61.4%
AD0225	Calculation issue: PCHG != (AVAL - BASE)/BASE * 100	12.6%	20.4%
AD0141	Inconsistent value for PARAM within a unique PARAMCD	11.9%	5.4%
CT2001	Variable value not found in non-extensible codelist	11.9%	50.9%

Examples

- › AD0047 '*Required variable is not present*' is notorious for producing false-positive messages.
 - Due to a bug that incorrectly identified OCCDS datasets as BDS
 - Bug is now fixed
- › There are still some problems with correct implementation of ADaM standard
 - "*Inconsistent value of PARCATy within a unique PARAMCD*" (24% studies)
- › 7 out of 20 most common issues in analysis data are related to problems with correct implementation of define.xml file

Improving diagnostics

- › Another area for potential enhancement of data validation
- › Examples:
 - *Inconsistency* checks report only a new inconsistent value without a reference to original base value
 - “*Variable label mismatch between dataset and ADM standard*” report does not clarify expected standard label

Summary & Questions

- › Validation of analysis data for compliance with CDISC ADaM standard and additional business rules from regulatory agencies is an important step in preparation of study data for regulatory submissions
- › New version of Pinnacle 21 Validator includes a draft validation for ADaM IG 1.1 and extends CDISC conformance rules for needs of regulatory submissions

Max Kanevsky

Pinnacle 21

max@pinnacle21.com

Sergiy Sirichenko

Pinnacle 21

sergiy@pinnacle21.com