Confusing Data Validation Rules Explained

Michael Beers, Pinnacle 21

April 30, 2020

PRESENTER – MICHAEL BEERS

- Principal Consultant at Pinnacle 21
- More than 10 years experience at both Pharma and CROs
- SME on FDA projects for CDER and CBER

INTRODUCTION

- Sources of Confusion
  - Standards Development Organization changes
    - Examples:
      - ADaM
        - ADaMIG Conformance Rules v3.0 (for ADaMIG v1.2) – Still in CDISC's review process
      - SDTM
        - SDTMIG Conformance Rules v1.1 (for SDTMIG v3.3) – Awaiting final publication
        - SDTMIG Conformance Rules V1.2 (for SDTMIG v3.4) – In Public Review
      - SEND
        - SENDIG Conformance Rules v1.0 (for SENDIG v3.0) – released March 2020
        - SENDIG Conformance Rules for SENDIG v3.1 – ??

INTRODUCTION

- Sources of Confusion, cont.
  - Regulatory Agency Changes
    - Examples:
      - Data Standards Catalog
        - Updated September 2019, April 2020
        - Requirement End dates for SDTM IG v3.1.2 Amendment 1, ADaM IG v1.0, and Define v1.0
      - Technical Conformance Guide
        - V4.5 – March 2020
      - PMDA – even more complicated
        - Watch recording of webinar “PMDA’s New Validation Rules” on Pinnacle 21 website

INTRODUCTION

- Sources of Confusion, cont.
  - Seemingly contradictory guidance between SDOs and regulatory agencies.
    - Examples:
      - Mapping planned arm for screen failure subjects
      - Core status for CDISC variables vs TCG expectations
  - Unclear or complicated validation messages
  - Misconception of certain concepts
    - e.g. Extensible codelists
  - Misunderstanding of the purpose of the validation rule
    - e.g. Duplicate records rule

DATA QUALITY RULES

These validation rules identify issues with the collection of the data, deficiencies in the data, or issues that may otherwise affect reviewability.

DATA QUALITY RULES – EXAMPLE 1

- Duplicate records (SD1117)
  - This validation rule looks to identify multiple observations collected for the same timepoint, based on a set of meaningful, common industry-wide keys
  - Assigned to Findings domains
  - Variables without standard definitions are not used as keys
    - Sponsor-defined variables, such as --SPID, --REFID
    - SUPPQUALs
  - This rule does not check for duplicate records against the sponsor defined keys in the define.xml

DATA QUALITY RULES – EXAMPLE 1

- Duplicate records (SD1117)
  - Common scenarios:

- Actual duplicate information, except for –SEQ

USUBJID	LBSEQ	LBTEST	LBORRES	LBORRESU	LBDTC
001-001	1	Hemoglobin	15	g/dL	2012-03-06T10:10
001-001	2	Hemoglobin	15	g/dL	2012-03-06T10:10

- Same timing information but one record is NOT DONE.

USUBJID	LBSEQ	LBTESTCD	LBORRES	LBORRESU	LBSTAT	LBNAM	VISIT	LBDTC
001-001	1	ALB	4.7	g/dL		ABC	UNSCHEDULED	2008-12-15
001-001	2	ALB			NOT DONE	ABC	UNSCHEDULED	2008-12-15
001-001	3	ALP	97	IU/L		ABC	UNSCHEDULED	2008-12-15
001-001	4	ALP			NOT DONE	ABC	UNSCHEDULED	2008-12-15

- Same timing information but results are different

USUBJID	LBSEQ	LBTEST	LBORRES	LBORRESU	LBDTC
001-001	1	Hemoglobin	15	g/dL	2012-03-06T10:10
001-001	2	Hemoglobin	18	g/dL	2012-03-06T10:10

- Only differentiated by a sponsor-defined variable.

USUBJID	LBSEQ	LBSPID	LBTEST	LBORRES	LBORRESU	LBDTC
001-001	1	123456	Hemoglobin	15	g/dL	2012-03-06T10:10
001-001	2	654321	Hemoglobin	15	g/dL	2012-03-06T10:10

DATA QUALITY RULES – EXAMPLE 1

- Duplicate records in -- Domain (SD1201)
  - This validation rule is the same as SD1117, however it is assigned to Events domains
  - This validation rule looks to identify multiple records for the same event at the same timepoint, based on a set of meaningful, common industry-wide keys

DATA QUALITY RULES – EXAMPLE 1

- Duplicate records in -- Domain (SD1201)
  - Common scenarios:
    - Actual duplicate information, except for –SEQ
    - Records that are only differentiated by a sponsor-defined variable (--SPID)
    - Records that are differentiated by the collection date (--DTC), instead of the start date (--STDTC) of the event
    - Records in the Disposition domain for multiple informed consent obtained
    - Records in the Medical History or Disposition domain for rescreened subjects
    - Multiple events, typically Medical History, that started in the same year/month, but exact date is unknown

DATA QUALITY RULES – EXAMPLE 1

- Duplicate records (SD1117)
- Duplicate records in -- Domain (SD1201)
  - Common explanation from sponsors:
    - "The keys defined by the check are not sufficient to identify a unique record for patient."
      - This is an actual explanation from a sponsor, and the actual keys listed in define.xml for this domain (Questionnaires) were: STUDYID, USUBJID, QSSEQ
    - Many times sponsors explain that --SPID, --GRPID, etc. need to be used, but the define.xml doesn't describe what these variables contain
    - If you can't correct your duplicate records, make sure to explain why the situation exists, and how to differentiate

DATA QUALITY RULES – EXAMPLE 1

- Duplicate records (SD1117)
- Duplicate records in -- Domain (SD1201)
  - P21's plan is to:
    - Update SD1117 to look at domain-specific keys
    - Add a new rule to check for unique records using keys specified in define.xml
    - Be on the lookout for new rule SD1352, assigned to Interventions domains

DATA QUALITY RULES – EXAMPLE 2

- No records for 'SCRNFAIL' subject are found in IE domain (SD1032)
  - Purpose of rule: Identify screen failure subjects who have no failed inclusion/exclusion criteria
  - Sources of Confusion:
    - Unclear and insufficient mapping guidance
    - Contradictory guidance on identifying screen failure subjects
      - Between SDOs and regulatory agencies
      - Between versions of the IGs

DATA QUALITY RULES – EXAMPLE 2

- No records for 'SCRNFAIL' subject are found in IE domain (SD1032)
  - Common scenarios:
    - IE records are just not submitted for some reason
      - Typical explanation:

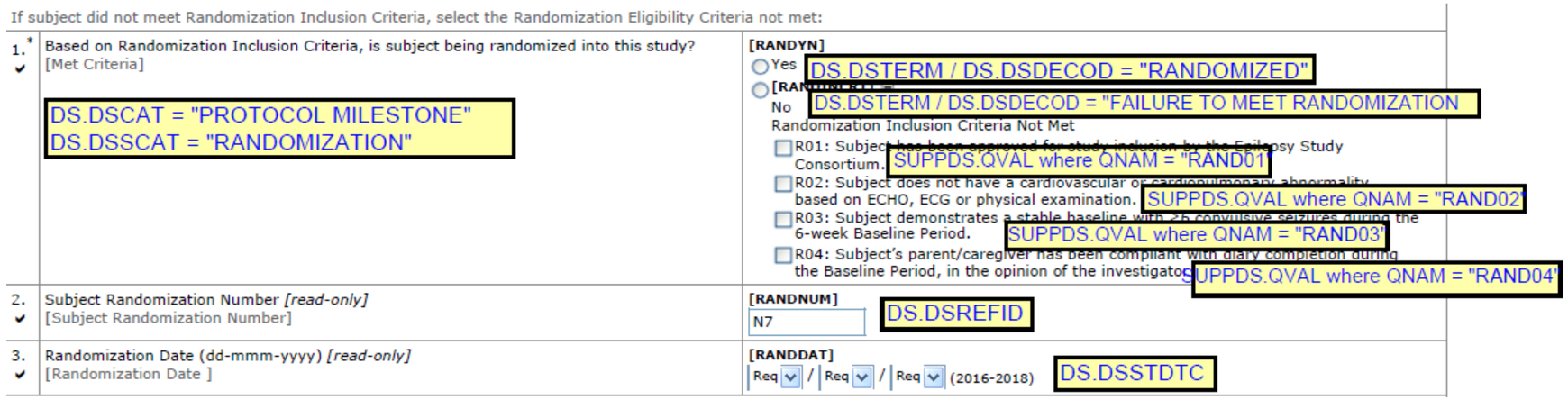
Dataset	Diagnostic Message	Severity	Count	Explanation
DM	No records for 'SCRFAIL' subject are found in IE domain	Warning	3418	Inclusion/Exclusion data for screen failure subjects who are not submitted.

    - Subjects identified as Screen Failure instead of Not Assigned
      - Typical explanation:

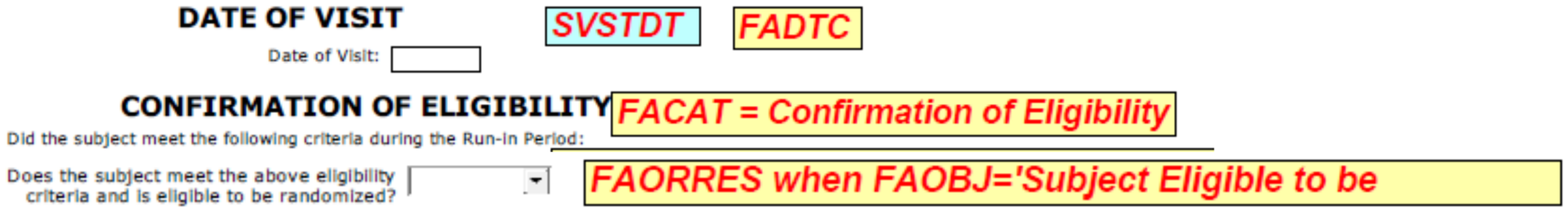
Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD1032	No records for 'SCRNFAIL' subject are found in IE domain	Warning	DM	62 (2.80%)	Sponsor made decision to close enrollment after the screening for these subjects.

DATA QUALITY RULES – EXAMPLE 2

- No records for 'SCRNFAIL' subject are found in IE domain (SD1032)
  - Common scenarios, cont.:
    - Subjects met IE criteria, but failed randomization criteria
      - May end up mapped anywhere, typically DS/SUPPDS:

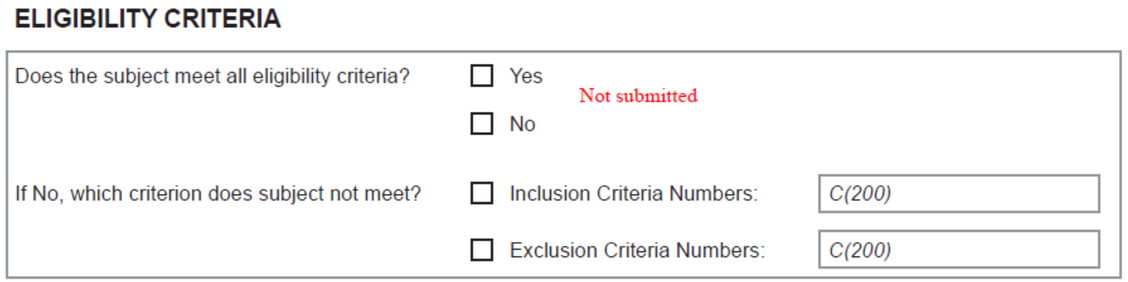


      - But not always:

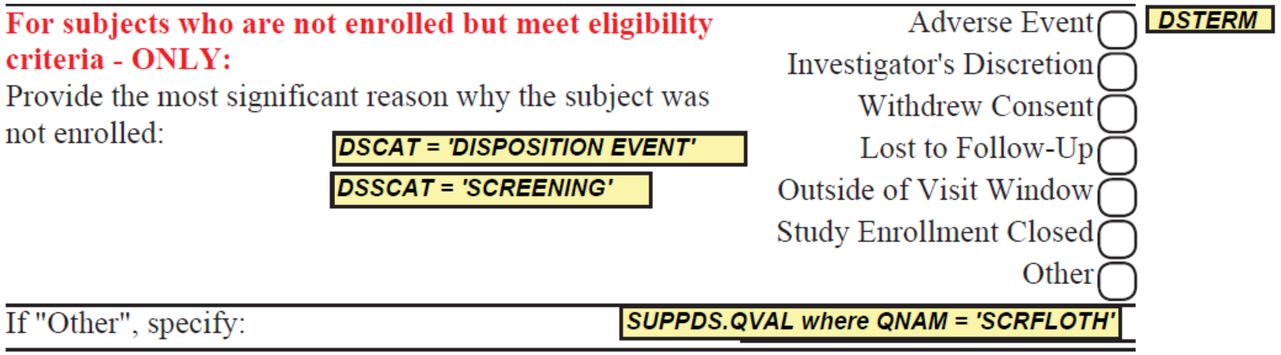


DATA QUALITY RULES – EXAMPLE 2

- No records for 'SCRNFAIL' subject are found in IE domain (SD1032)
  - Common scenarios, cont.:
    - CRF captures just the IE criterion failed, and value mapped as-is (typically to DS)



  - Subjects met IE criteria, but withdrew prior to randomization



DATA QUALITY RULES – EXAMPLE 2

- No records for 'SCRNFAIL' subject are found in IE domain (SD1032)
  - May be somewhat common for some screen failure subjects to not have records in the IE domain
  - Becomes problematic when many subjects are missing this information
    - Could affect the ability to determine possible bias in patient enrollment
  - Recommendation is to:
    - Verify you are using the correct ARM value
    - Verify that this failed criteria data is collected for screen failure subjects, and mapped appropriately to the IE domain
  - More official guidance would be beneficial to reduce industry-wide variation in mapping

CONTROLLED TERMINOLOGY RULES

These validation rules identify discrepancies between the values a sponsor used in their data compared to allowable values of controlled terminology lists.

CONTROLLED TERMINOLOGY RULES – EXAMPLE 1

- Value not found in extensible codelist (CT2002)
  - Purpose of rule: This validation rule will fire if a value in the dataset, for a variable with a CDISC-defined codelist, does not exactly match a value in the CDISC extensible codelist
  - Source of confusion:
    - Misconception of the concept of ‘extensible codelist’

- Variable value not found in extensible codelist when value-level condition occurs (CT2005)
  - Purpose of rule: Same as CT2002, but for Value Level instead of Variable Level

CONTROLLED TERMINOLOGY RULES – EXAMPLE 1

- Value not found in extensible codelist (CT2002)
- Variable value not found in extensible codelist when value-level condition occurs (CT2005)

  - Common scenarios:
    - A value is used that has no corresponding match in the CDISC codelist
      - This is the valid case for having CT2002/CT2005 fire for your study
    - A value is used that has a match in the CDISIC codelist but casing differs
    - A synonym is used for a value in the CDISC codelist
    - Values are combined that should be in split into separate SDTM variables
      - Example: Values of LEFT/RIGHT combined in the --LOC variable instead of the --LAT
    - OTHER is concatenated with the specified value (OTHER: specified value)

CONTROLLED TERMINOLOGY RULES – EXAMPLE 1

- Variable value not found in extensible codelist when value-level condition occurs (CT2005)
  - Example of confusion:
    - CT2005 fired for DSDECOD (when DSCAT = DISPOSITION EVENT)
      - Sponsor's explanation of issue:

Dataset	Diagnostic Message	Severity	Count	Explanation
DS	DSDECOD value not found in 'Completion/Reason for Non-Completion' extensible codelist when DSCAT == 'DISPOSITION EVENT'	Warning	4370	False positive result as no codelist exists and DSDECOD defined as per the CRF

CONTROLLED TERMINOLOGY RULES – EXAMPLE 1

- Variable value not found in extensible codelist when value-level condition occurs (CT2005)
  - Example of confusion, cont.:
    - CT2005 fired for DSDECOD (when DSCAT = DISPOSITION EVENT)
      - Here are some of the values being flagged:

Issue Details - CT2005 (DS)

Details | Records | Explanation

Search | Print | Copy | Download

DSDECOD	DSCAT	Failures	% Affected Records	% Total Records
SUBJECT DECISION	DISPOSITION EVENT	238	5.4%	0.7%
SEVERE NON-COMPLIANCE TO PROTOCOL	DISPOSITION EVENT	9	0.2%	< 0.1%

CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
WITHDRAWAL BY SUBJECT		An indication that a study participant has removed itself from the study. (NCI)	Withdrawal by Subject

Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition
Completion/Reason for Non-Completion	PROTOCOL VIOLATION		A significant departure from processes or procedures that were required by the protocol. Violations often result in data that are not deemed evaluable for a per-protocol analysis, and may require that the subject(s) who violate the protocol be discontinued from the study. Compare to protocol deviation. (CDISC Glossary)

CONTROLLED TERMINOLOGY RULES – EXAMPLE 1

- Value not found in extensible codelist (CT2002)
- Variable value not found in extensible codelist when value-level condition occurs (CT2005)

  - Common explanation from sponsors:
    - "Codelist is extensible."
  - Extending an extensible codelist when there is no corresponding value to use is an acceptable approach, however the other scenarios are not
  - A proper dispositioning of this validation issue would be to:
    - Correct the implementation to use the valid controlled terminology value where possible
    - List all actually valid extended values, if possible
  - Even better…submit your extended terms to CDISC (as early as possible) to have them added to CT

CONTROLLED TERMINOLOGY RULES – EXAMPLE 2

- Variable and Decode values do not have the same Code in CDISC CT (CT2003)
  - In CDISC controlled terminology, for paired variables (--TESTCD/--TEST and --PARMCD/--PARM), both values will use the same NCI Code value.
  - When a certain --TESTCD or --PARMCD value is used, you must use the --TEST or --PARM value that corresponds to the --TESTCD or --PARMCD with the same NCI code

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C74683	C65047		Laboratory Test Code	CYTYRO	Tyrosine Crystals
C74683	C67154		Laboratory Test Name	Tyrosine Crystals	Tyrosine Crystals

  - Purpose of rule: This validation rule will fire if that is not the case

CONTROLLED TERMINOLOGY RULES – EXAMPLE 2

- Variable and Decode values do not have the same Code in CDISC CT (CT2003)
  - Common scenarios:
    - A value from the synonym column is used instead of the CDISC Submission Value
    - Misspelling of one of the values
    - The wrong CT version was configured for the validation
    - --TESTCD and --TEST values for different tests were mixed up
      - This is the most concerning scenario

CONTROLLED TERMINOLOGY RULES – EXAMPLE 2

- Variable and Decode values do not have the same Code in CDISC CT (CT2003)
  - Example:
    - Values in CDISC Controlled Terminology

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C74683	C65047		Laboratory Test Code	CYTYRO	Tyrosine Crystals
C74683	C67154		Laboratory Test Name	Tyrosine Crystals	Tyrosine Crystals
C74756	C65047		Laboratory Test Code	CYTRPHOS	Triple Phosphate Crystals
C74756	C67154		Laboratory Test Name	Triple Phosphate Crystals	Triple Phosphate Crystals

- Values in the dataset:

USUBJID	LBTESTCD	LBTEST	LBORRES	LBORRESU	VISIT	LBDTC
001-001	CYTYRO	Triple Phosphate Crystals	1+		Visit 3	2016-05-13T10:27
001-001	CYTYRO	Triple Phosphate Crystals	2+		Visit 4	2016-06-13T11:30

CONTROLLED TERMINOLOGY RULES – EXAMPLE 2

- Variable and Decode values do not have the same Code in CDISC CT (CT2003)
  - Recommendation is to always fix this issue
    - Make sure you correctly map your paired variable values to controlled terminology
    - Make sure you configure your validation correctly

CONTROLLED TERMINOLOGY RULES – EXAMPLE 3

- Value for --DECOD not found in WHODrug dictionary (SD1344)
- Value for --CLAS not found in WHODrug dictionary (SD1345)
- Value for --CLASCD not found in WHODrug dictionary (SD1346)

  - Purpose of rule: New rules added to check concomitant medications coding against the WHODrug dictionary

  - Sources of confusion:

    - Not all WHODrug version formats are supported
    - How to handle mapping values greater than 200 characters from dictionary
    - Lack of official mapping guidance

CONTROLLED TERMINOLOGY RULES – EXAMPLE 3

- Value for --DECOD not found in WHODrug dictionary (SD1344)
- Value for --CLAS not found in WHODrug dictionary (SD1345)
- Value for --CLASCD not found in WHODrug dictionary (SD1346)

  - Not all WHODrug version formats are supported

FDA Data Standards Catalog v6.2 (04-10-2020)

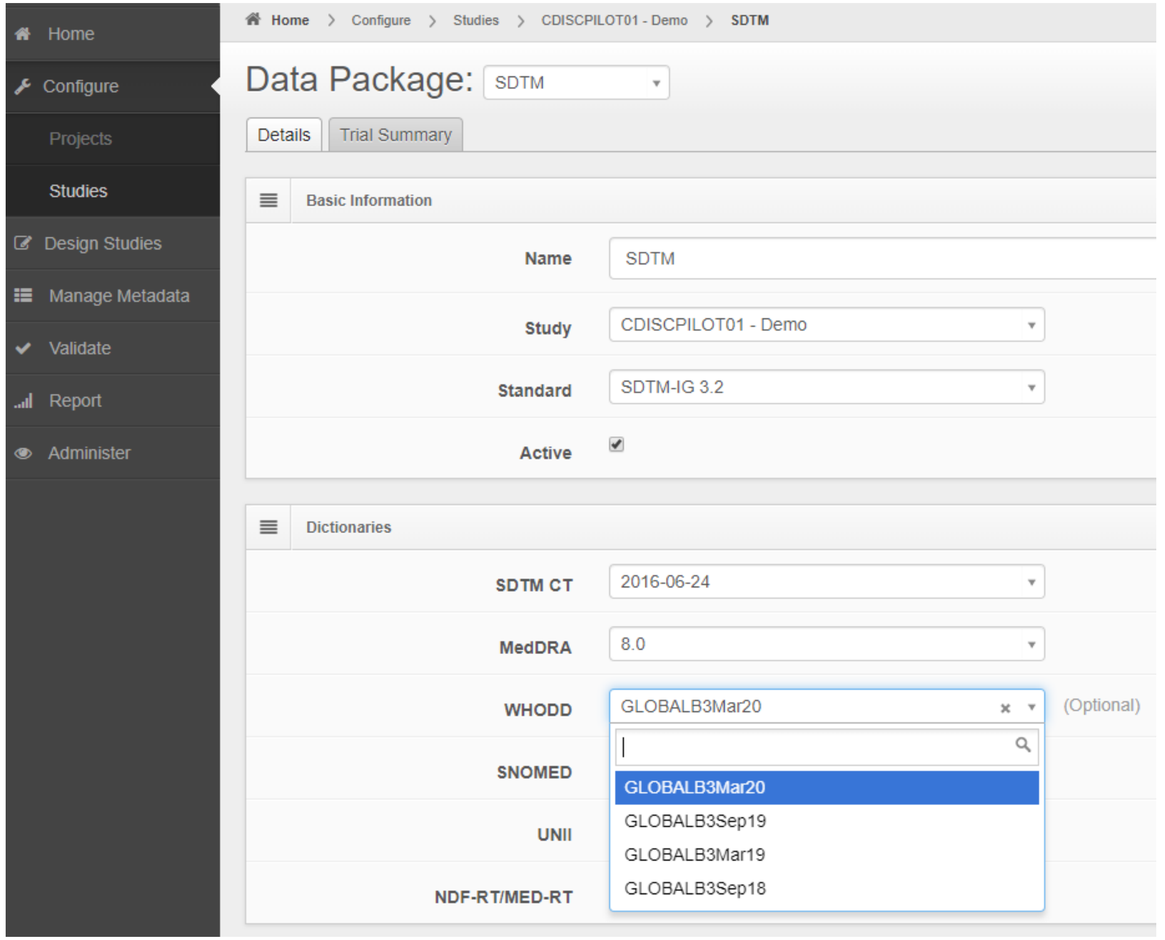
For full description of column headings, see Instr. & Column Descriptions tab

Use	Terminology Standard	Terminology Standards Development and/or Maintenance Organization	Version(s)	FDA Center(s)	Date Support Begins (MM/DD/YYYY)	Date Support Ends	Date Requirement Begins (MM/DD/YYYY)	Date Requirement Ends	Examples of Use	Statutory, Regulatory, or Guidance Authority
Medication	WHODrug Global	Uppsala Monitoring Centre	Current Version-B3 format	CBER, CDER	03/15/2018		03/15/2019		Use in SDTM CMDECOD and CMCLAS	Standardized Study Data

CONTROLLED TERMINOLOGY RULES – EXAMPLE 3

- Value for --DECOD not found in WHODrug dictionary (SD1344)
- Value for --CLAS not found in WHODrug dictionary (SD1345)
- Value for --CLASCD not found in WHODrug dictionary (SD1346)

- Not all WHODrug version formats are supported



CONTROLLED TERMINOLOGY RULES – EXAMPLE 3

- Value for --DECOD not found in WHODrug dictionary (SD1344)
- Value for --CLAS not found in WHODrug dictionary (SD1345)
- Value for --CLASCD not found in WHODrug dictionary (SD1346)

- Mapping values greater than 200 characters from dictionary
  - According to “How to use WHODrug for compliance with CM domain in the CDISC SDTM standard”*

CMDECOD is longer than 200 characters

For drugs with many ingredients, the generic name is longer than 200 characters. The SAS export format has a limitation to 200 characters per field, if this format is used for submission, the supplemental dataset needs to be utilised. Note that the guidelines state that the text should be truncated between words, in the case for long generic names the text should be truncated after the semicolon closest to 200 characters. Illustrations of the ordinary and supplemental datasets are shown in table 1 and 2.

Table 1. Illustration of SDTM dataset where CMDECOD is longer than 200 characters.

USUBJID	CMSEQ	CMTRT	CMMODIFY	CMDECOD	CMCLAS	CMCLASCD
AB-21-01	1	....	...	Ascorbic acid;Biotin;Calcium;Carbohydrates nos; Chloride;Choline;Chromium;Colecalciferol; Copper;Cyanocobalamin;Docosahexaenoic acid; Fats nos;Folic acid;Fructooligosaccharides; Iodine;Iron;Magnesium;	....	....

Table 2. Illustration of supplemental dataset for CM domain where CMDECOD is longer than 200 characters.

USUBJID	RDOMAIN	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
AB-21-01	CM	CMSEQ	1	CMDECOD1	Standardized Medication Name 1	Manganese;Nicotinic acid;Pantothenic acid;Phosphorus;Phytomenadione; Potassium;Proteins nos;Pyridoxine; Retinol;Riboflavin;Selenium;Sodium; Thiamine;Vitamin e nos;Zinc

*https://www.who-umc.org/media/2940/how-to-use-whodrug-for-compliance-with-cm-domain-in-the-cdisc-sdtm-standard-march-2017.pdf

CONTROLLED TERMINOLOGY RULES – EXAMPLE 3

- Value for --DECOD not found in WHODrug dictionary (SD1344)
- Value for --CLAS not found in WHODrug dictionary (SD1345)
- Value for --CLASCD not found in WHODrug dictionary (SD1346)

  - Enterprise currently supports WHODrug validation
  - Community will support it shortly
  - Recommend following Uppsala Monitoring Centre's guidance for mapping to the CM domain
  - If interested, see PhUSE US Connect 2020 paper "SUPPQUAL Datasets: Good Bad and Ugly", by Sergiy Sirichenko (Pinnacle 21)
    - 298 studies were analyzed…667 unique QNAMs used for ATC Classification coding
      - Official guidance for this mapping is critically needed

METADATA RULES

These validation rules identify issues with the define.xml, such as issues with the XML code, incorrect implementation of define.xml, or inconsistencies between the define.xml and the study datasets.

METADATA RULES – EXAMPLE 1

- Value for variable not found in user-defined codelist (SD0037)
  - Purpose of rule:
    - This rule looks to identify values in the dataset that are not listed in the associated codelist in the define.xml

- Variable value not found in user-defined codelist when value-level condition occurs (SD1228)
  - Purpose of rule:
    - Same as SD0037, but for Value Level Metadata

METADATA RULES – EXAMPLE 1

- Value for variable not found in user-defined codelist (SD0037)
- Variable value not found in user-defined codelist when value-level condition occurs (SD1228)

  - Common scenarios:
    - Casing difference or misspelling between the define.xml codelist and the data value
    - A value in the dataset is just missing from the codelist in the define.xml
      - May be because data was updated after define.xml creation
    - The wrong codelist was accidentally assigned for a variable in the define.xml

METADATA RULES – EXAMPLE 1

- Variable value not found in user-defined codelist when value-level condition occurs (SD1228)
  - Example:

Value Level Metadata - TS [TSVAL]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
TSVAL	TSPARMCD = "INTTYPE" (Intervention Type)	text	40	["DRUG"] <Intervention Type>	Protocol	

  - The codelist in the define.xml has these values:

Intervention Type [CL.INTTYPE]

Permitted Value (Code)
DRUG [*]

* Extended Value

  - The data has this value:

Issue Details - SD1228 (TS)

This value is not listed in the codelist above

Details | Records | Explanation

Search

TSPARMCD	TSVAL	Failures	% Affected Records	% Total Records
INTTYPE	BIOLOGIC	1	100%	1.5%

METADATA RULES – EXAMPLE 1

- Value for variable not found in user-defined codelist (SD0037)
- Variable value not found in user-defined codelist when value-level condition occurs (SD1228)

  - Typical explanations from sponsors:
    - "The codelist is extensible."
      - (This indicates that the sponsor doesn't understand what the rule is doing)
    - No explanation provided because sponsor didn't validate their define.xml with their data

  - Metadata issues should be corrected

METADATA RULES – EXAMPLE 2

- Variable value is longer than defined max length when value-level condition occurs (SD1231)
  - Purpose of rule: To check for situations where values in the dataset, for a variable, when a value-level condition is met, are greater than the length specified in the define.xml

METADATA RULES – EXAMPLE 2

- Variable value is longer than defined max length when value-level condition occurs (SD1231)
  - Common scenario:
    - Define.xml created for an ongoing study, updated data is received, but the define.xml is not refreshed with the latest metadata
    - The sponsor is completely unaware that this validation issue exists, because when validating the datasets, the define.xml was incorrectly excluded from the validation
    - The sponsor thinks the finding is a false-positive, due to hidden characters such as leading spaces

METADATA RULES – EXAMPLE 2

- Variable value is longer than defined max length when value-level condition occurs (SD1231)
  - Example:
    - SD1231 fired for the SUPPDS.QVAL where QNAM = 'ENTCRIT':

Rule	Message	FDA	PMDA	Rate	Status	Updated
SUPPDS - Supplemental Qualifiers for DS (1 Issues)						
SD1231	QVAL value is longer than defined max length 2 when QNAM == 'ENTCRIT'	Error	Warning	66.7%	Open	2020-02-20 12:41:04

    - SUPPDS.QVAL where QNAM = 'ENTCRIT' listed in define.xml with length of 2:

Value Level Metadata - SUPPDS [QVAL]

Variable	Where	Type	Length / Display Format
QVAL	QNAM EQ ENTCRIT (PROTOCOL ENTRY CRITERIA NOT MET)	text	2

    - The SUPPDS.QVAL variable (where QNAM = 'ENTCRIT') has these values:

QNAM	QVAL
ENTCRIT	C16
ENTCRIT	[1LS]C25

These values are longer than 2

METADATA RULES – EXAMPLE 2

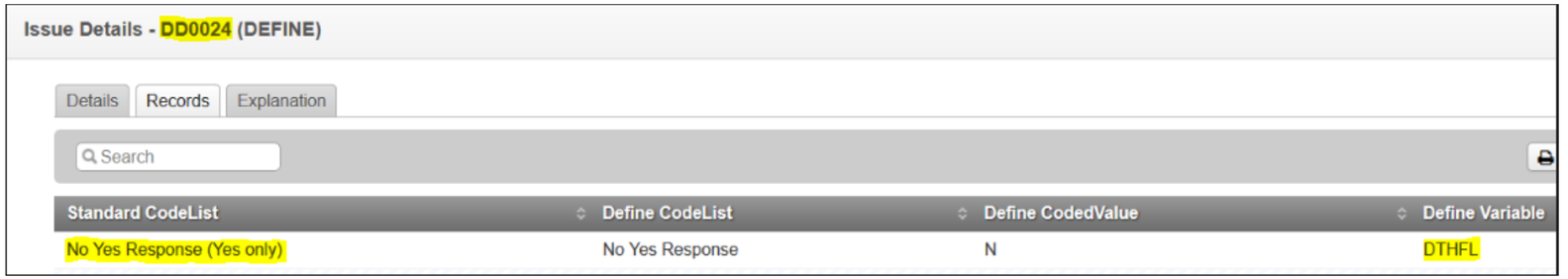
- Variable value is longer than defined max length when value-level condition occurs (SD1231)
  - When creating value-level metadata in a define.xml, care must be taken to ensure that the metadata accurately reflects the actual data
    - Always be sure to refresh an existing define.xml when data is updated
  - The recommendation is to always fix all define.xml related validation issues, so that a clean error-free define.xml is provided as part of the submission
  - Correct validation is essential
    - Validate the define.xml with the data to run data vs. define.xml cross-check rules
    - Validate the define.xml by itself to identify any XML related issues

METADATA RULES – EXAMPLE 3

- Invalid Term in Codelist ‘No Yes Response (Yes Only)’ Codelist (DD0024)
  - Purpose of rule:
    - This validation rule fires when a variable should only have a value of ‘Y’ or null, per CDISC implementation guidance, but in the define.xml that variable references a codelist that contains other values
  - Common scenario:
    - A sponsor will create one No/Yes codelist, and have many variables reference it, regardless if all of the values of that variable apply

METADATA RULES – EXAMPLE 3

- Invalid Term in Codelist 'No Yes Response (Yes Only)' Codelist (DD0024)
  - Example:
    - DD0024 fired for the DM.DTHFL variable:



    - DTHFL listed in define.xml with the No Yes Response codelist:

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
DTHFL	Subject Death Flag		text	1	["N" = "No", "Y" = "Yes"] <No Yes Response>	Derived	Equals to Y when trial termination has been completed with DEATH in the DS dataset.

    - The codelist in the define.xml has these values:

No Yes Response [CL.NY, *C66742*]

Permitted Value (Code)	Display Value (Decode)
N [*C49487*]	No
Y [*C49488*]	Yes

METADATA RULES – EXAMPLE 3

- Invalid Term in Codelist ‘No Yes Response (Yes Only)’ Codelist (DD0024)
  - This variable in the example (DM.DTHFL), per CDISC guidance, should be ‘Y’ or null. By referencing a codelist with these other values, it becomes unclear if the sponsor is using these values that aren’t allowed
  - A proper dispositioning of this issue is, for variables where only values of ‘Y’ are allowed, to reference a separate codelist with only this value
  - Metadata issues should be corrected

REGULATORY CONFORMANCE RULES

These validation rules look for violations of regulatory agency (FDA and PMDA) guidance in their Technical Conformance Guides, such as missing requested data, and implementations inconsistent with the regulatory guidance.

REGULATORY CONFORMANCE RULES – EXAMPLE 1

- Regulatory Expected variable not found (SD1077)
  - Purpose of rule:
    - This validation rule looks to make sure that the EPOCH variable is provided in the appropriate domains

- Missing EPOCH value, when a start or observation date is provided (SD1339)
  - Purpose of rule:
    - Previously we were only checking for presence of EPOCH variable
    - This is a new rule to flag where an EPOCH value should have been provided but wasn't

REGULATORY CONFORMANCE RULES – EXAMPLE 1

- Regulatory Expected variable not found (SD1077)
- Missing EPOCH value, when a start or observation date is provided (SD1339)
  - Sources of confusion:
    - CDISC guidance listing this variable as permissible
    - Technical Conformance Guide stating that the EPOCH variable should be provided
    - Historical data

REGULATORY CONFORMANCE RULES – EXAMPLE 1

- Regulatory Expected variable not found (SD1077)
- Missing EPOCH value, when a start or observation date is provided (SD1339)
  - Common scenarios:
    - EPOCH is not provided for domains that the sponsor deems unnecessary, such as IE, etc.
      - Note: we've excluded some domains for this rule, such as MH, SU
    - EPOCH is missing due to no collected timing information (and therefore not possible to derive)
    - EPOCH is not provided for domains that the sponsor thinks conflicts with CDISC guidance, such as SV
    - EPOCH is just missing when it should have been provided
    - EPOCH is left null for historical data (CM, etc.)

REGULATORY CONFORMANCE RULES – EXAMPLE 1

- Regulatory Expected variable not found (SD1077)
- Missing EPOCH value, when a start or observation date is provided (SD1339)

  - Typical explanation from sponsors:
    - "EPOCH is a permissible variable as per SDTM IG 3.1.3, hence not included.“
    - It does not seem appropriate to disregard regulatory guidance based on CDISC variable core status
    - In most cases, EPOCH should be provided and populated

  - Now we have rules to check for both presence and population of the EPOCH variable. Possible next step…is the value of EPOCH correct?

REGULATORY CONFORMANCE RULES – EXAMPLE 2

- Missing new parameters listed in TCG
  - Missing EXTTIND Trial Summary Parameter (SD2273)
  - Missing NCOHORT Trial Summary Parameter (SD2274)
  - Missing OBJSEC Trial Summary Parameter (SD2275)
  - Missing PDPSTIND Trial Summary Parameter (SD2276)
  - Missing PDSTIND Trial Summary Parameter (SD2277)
  - Missing PIPIND Trial Summary Parameter (SD2278)
  - Missing RDIND Trial Summary Parameter (SD2279)
  - Missing SDTIGVER Trial Summary Parameter (SD2280)
  - Missing SDTMVER Trial Summary Parameter (SD2281)
  - Missing THERAREA Trial Summary Parameter (SD2282)

  - Purpose of rules:
    - To check for presence of trial summary parameters requested by FDA in the TCG

REGULATORY CONFORMANCE RULES – EXAMPLE 2

- Missing new parameters listed in TCG
  - Source of confusion:
    - Listed in TCG but not IG
      - Examples from

        TCG, Appendix B:

FDA Desired - Clinical	TSPARMCD	TSPARM	FDA Notes
Conditional	SDMDUR	Stable Disease Minimum Duration	If applicable.
Y	SENDTC	Study End Date	
Y	SEXPOP	Sex of Participants	
Y	SPONSOR	Clinical Study Sponsor	
Y	SDTMVER	SDTM Version	The value should be the exact term listed in the FDA Data Standards Catalog in Column E. If multiple SDTM Versions are used for a study the every version should be listed on each row.
Y	SDTIGVER	SDTM IGVersion	The value should be the exact term listed in the FDA Data Standards Catalog in Column F. If multiple SDTM IG Versions are used for a study the every version should be listed on each row.
Y	STOPRULE	Study Stop Rules	If no stopping rule, STOPRULE = 'NONE'.
Conditional	STRATFCT	Stratification Factor	If applicable. Use as many rows as needed.
Y	SSTDTC	Study Start Date	
Y	STYPE	Study Type	
Y	TBLIND	Trial Blinding Schema	
Y	TCNTRL	Control Type	
Conditional	TDIGRP	Diagnosis Group	Where HLTSUBJI = 'N'.
Y	THERAREA	Therapeutic Area	

REGULATORY CONFORMANCE RULES – EXAMPLE 2

- Missing new parameters listed in TCG
  - Recommendations
    - The TCG should be used to determine which parameters should be included in your TS domain
      - CDISC plans to eventually discontinue managing TS parameters in the Implementation Guides
    - If the parameter doesn't apply to your study, still include it but leave TSVAL null and use the TSVALNF (null flavor) variable

TRACEABILITY RULES

These validation rules identify issues with traceability between SDTM and ADaM datasets.

TRACEABILITY RULES

- Traceability rules not executed due to missing DM dataset (AD1024)
- Traceability rules not executed due to missing AE dataset (AD1025)
- Traceability rules not executed due to missing EX dataset (AD1026)

  - Purpose of rules: To ensure that a sponsor is including the correct SDTM datasets in their ADaM validation so that existing SDTM to ADaM traceability validation rules will be run

    - Added because the industry mostly does not do this properly

TRACEABILITY RULES

- Traceability rules not executed due to missing DM dataset (AD1024)
- Traceability rules not executed due to missing AE dataset (AD1025)
- Traceability rules not executed due to missing EX dataset (AD1026)

  - Sources of confusion:

    - New rules added recently
    - Adding SDTM data to an ADaM package
    - Only should be done for validation, but not in ADaM datasets eCTD folder in submission

TRACEABILITY RULES

- Traceability rules not executed due to missing DM dataset (AD1024)
- Traceability rules not executed due to missing AE dataset (AD1025)
- Traceability rules not executed due to missing EX dataset (AD1026)

  - Including the SDTM Demographics dataset is necessary to run these checks:
    - ADaM ADSL vs SDTM DM
      - The combination of STUDYID and USUBJID value does not exist in the SDTM DM domain (AD0053)
      - For the same USUBJID, the ADSL.AGE != DM.AGE (AD0204)
      - For the same USUBJID, the ADSL.AGEU != DM.AGEU (AD0205)
      - For the same USUBJID, the ADSL.SEX != DM.SEX (AD0206)
      - For the same USUBJID, the ADSL.RACE != DM.RACE (AD0207)
      - For the same USUBJID, the ADSL.SUBJID != DM.SUBJID (AD0208)
      - For the same USUBJID, the ADSL.SITEID != DM.SITEID (AD0209)
      - For the same USUBJID, the ADSL.ARM != DM.ARM (AD0210)
      - For the same USUBJID, the ADSL.ACTARM != DM.ACTARM (AD0367)

TRACEABILITY RULES

- Traceability rules not executed due to missing DM dataset (AD1024)
- Traceability rules not executed due to missing AE dataset (AD1025)
- Traceability rules not executed due to missing EX dataset (AD1026)

  - Including the SDTM Exposure dataset is necessary to run these checks:
    - ADaM ADSL vs SDTM EX
      - SDTM.EX is present but neither ADSL TRTSDT nor TRTSDTM are present (AD0061)
      - SDTM.EX is present but neither ADSL TRTEDT nor TRTEDTM are present (AD0061A)

  - Including the SDTM Adverse Events dataset is necessary to run these checks:
    - ADaM ADAE vs SDTM AE
      - Record key from SDTM AE is not traceable to ADaM ADAE (not enough ADAE recs) (AD0253)
      - Record key from ADaM ADAE is not traceable to SDTM.AE (extra ADAE recs) (AD0258)

TRACEABILITY RULES

- Traceability rules not executed due to missing DM dataset (AD1024)
- Traceability rules not executed due to missing AE dataset (AD1025)
- Traceability rules not executed due to missing EX dataset (AD1026)

  - Impact:
    - Traceability issues is clinical trial data are extremely important
    - Automated validation at the agencies always includes these SDTM datasets an ADaM validation
    - Industry's lack of knowledge/understanding of this process leads to unnoticed traceability issues (until seen at the regulatory agencies)
      - This results in traceability issues not being corrected or even explained in the reviewers guide

SUMMARY

SUMMARY

- A clear understanding of validation rules, and the issues that are identified by these rules, is critical to providing high quality standardized study data.

- Confusion regarding validation rules leads to:
  - Important issues not being corrected
  - Regulatory agencies not receiving the information they need
  - Data issues not being explained sufficiently


QUESTIONS?