

Best Practice for Explaining Validation Results in the Study Data Reviewer's Guide

Kristin Kelly, Pinnacle 21, LLC

Michael Beers, Pinnacle 21, LLC

EU PhUSE Connect 2018

Paper DS06

Agenda

Introduction

To Fix or Not to Fix

Rules that Require Investigation

Best Practices for Explanations

Pop Quiz!

Conclusion

Introduction (1)

- For SDTM, it is recommended to include a Clinical Study Data Reviewer's Guide (cSDRG) in the study package
 - This document should be named 'csdrg.pdf' (where 'c' designates clinical)
 - Provided as a PDF for each study in module 5 (m5) of the eCTD.
 - Recommended template developed by PhUSE that is widely used
- The cSDRG typically contains the following:
 - SDTM/SDTMIG version (e.g. SDTM v1.4/SDTMIG v3.2)
 - Dictionary/terminology (e.g. MedDRA, CDISC)
 - Trial design and other domain-specific information that cannot be described in the define.xml
 - Results from automated validation checks
 - Listed in Section 4 of the cSDRG
 - Issues that cannot be resolved due to data oddities with explanations

Introduction (2)

- It is important to be transparent about the submitted data
 - Not providing a complete response for a validation check results in having a reviewer spend time investigating why a particular issue remains
- Possible reasons for not fully explaining a check:
 - Insufficient knowledge of how to investigate why a particular check is firing
 - Difficulty determining which issues should be fixed vs those that should be explained
 - Severity of 'Warning' and not 'Error'
 - Sometimes it is not okay to leave it as is and simply provide an explanation
 - Results in some programming issues that fired a check remaining in the data as well as incorrect or incomplete responses in the cSDRG.

*Note: 'Reviewer' – any consumer of the data downstream (e.g. analyst that reviews the data before/after submission, a clinical programmer, a medical reviewer, etc)

To Fix or Not to Fix? (1)

SD0063 – SDTM/dataset variable label mismatch

- This rule checks that the variable label in the dataset matches the variable label defined in the SDTMIG
- Though the PE domain is mentioned, it is not stated which variable label does not match.

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD0063	SDTM/dataset variable label mismatch	Warning	PE	1 (6.67%)	Label is fine.

- Insufficient explanation in the cSDRG

To Fix or Not to Fix? (2)

SD0063 – SDTM/dataset variable label mismatch

- The complete explanation mentions the variable affected, provides revised label, and the reason the label was updated.

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD0063	SDTM/dataset variable label mismatch	Warning	PE	1 (6.67%)	Per the SDTMIG, the variable label in the dataset should match the label in the SDTMIG. In SDTMIG v3.2, there are several labels that were erroneously defined to be >40 characters. The variable label for PESTRESC, 'Character Result/Finding in Standard Format', exceeds 40 characters. To avoid truncation, PESTRESC was assigned the label 'Character Result/Finding in Std Format' in the dataset so that it is <= 40 characters due to SAS restrictions.

To Fix or Not to Fix? (3)

CT2003 – Variable code and decode values do not have the same code in CDISC Controlled Terminology

- There are some paired variables (--TESTCD/--TEST, TSPARMCD/TSPARM, etc.) in CDISC CT. The code variable value and decode variable value share the same NCI code
- Example:
 - Values in CDISC Controlled Terminology

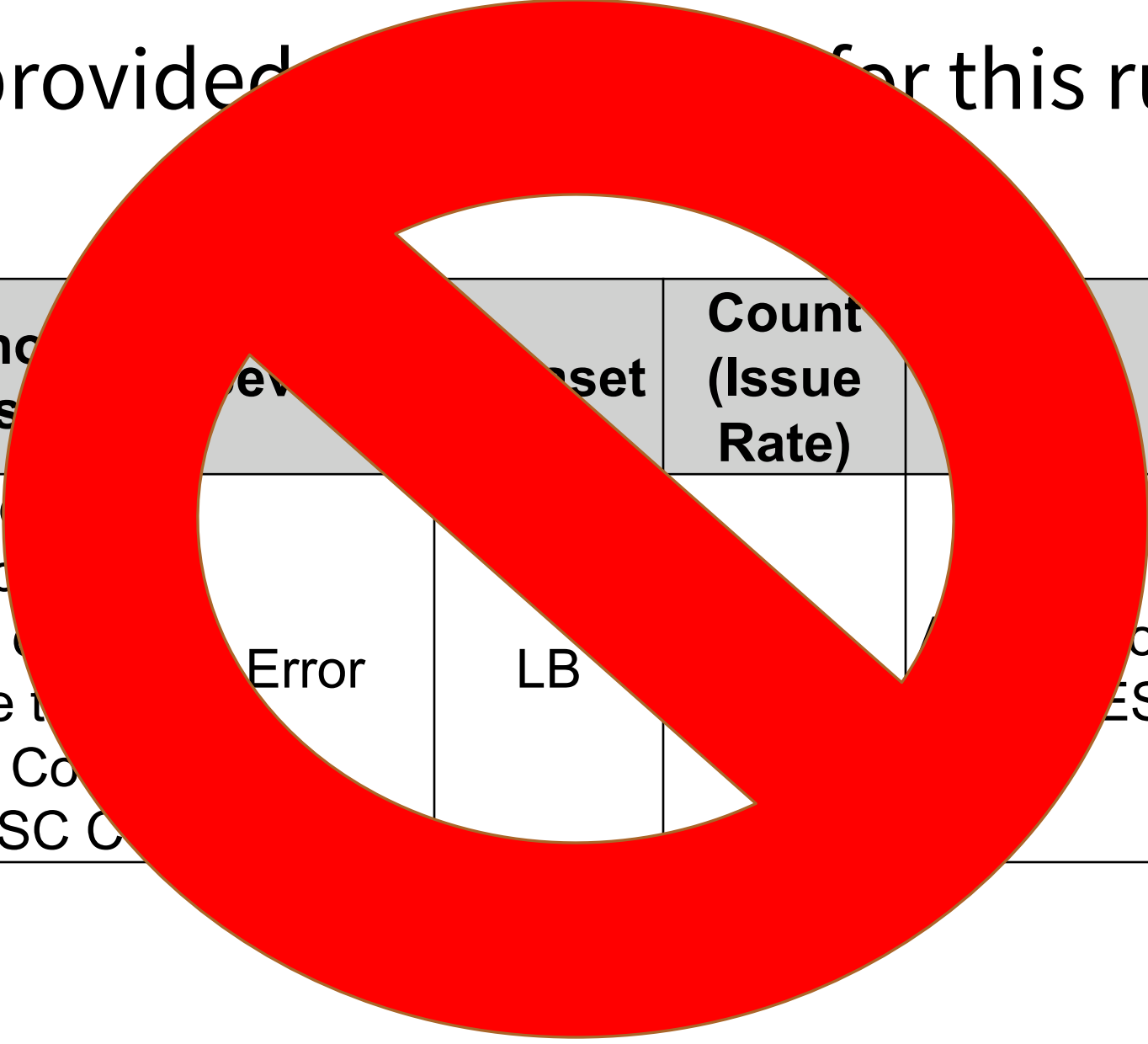
Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C74683	C65047		Laboratory Test Code	CYTYRO	Tyrosine Crystals
C74683	C67154		Laboratory Test Name	Tyrosine Crystals	Tyrosine Crystals
C74756	C65047		Laboratory Test Code	CYTRPHOS	Triple Phosphate Crystals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

To Fix or Not to Fix? (4)

- The sponsor provided [redacted] for this rule in the cSDRG:



Check ID	Diagnostic Message	Event	Dataset	Count (Issue Rate)	Explanation
CT2003	Code values have the same Code in CDISC C	Error	LB		Controlled Terminology, codelist EST is extensible.

- Incorrect explanation in the cSDRG
 - Not rectifying this issue in LB leads to uncertainty as to which test was performed

Rules that Require Investigation (1)

SD1117 – Duplicate Records

- This validation rule looks to identify potential duplicate records, based on a set of meaningful, common industry-wide keys.
- Variables without standard definitions are not used (sponsor-defined variables and SUPPQUAL variables)
- This rule does not check for duplicate records against the sponsor-defined keys in the define.xml.

Rules that Require Investigation (2)

SD1117 – Duplicate Records

- 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 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

Rules that Require Investigation (3)

SD1117 – Duplicate Records

- Common scenarios, cont.:
 - 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

- 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

Rules that Require Investigation (4)

- The explanation in the cSDRG was the following:

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD1117	Duplicate records	Warning	FT	324 (32.5%)	The keys defined by the check are not sufficient to identify a unique record for patient.

- Upon investigation:
 - The keys listed in the define.xml for FT (Functional Tests) were:
 - STUDYID, USUBJID, FTSEQ
 - The records flagged by the check could be differentiated using the FTSPID variable but a definition of what was contained in the variable in the define.xml was not provided

Rules that Require Investigation (5)

- A correct explanation might be the following:

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD1117	Duplicate records	Warning	FT	324 (32.5%)	The records flagged are for an assessment where the subject was asked to perform multiple trials of the same task. The records can be differentiated by the FTSPID variable which was populated with the trial number from the eCRF. This definition is provided in the define.xml.

Example: Search and Rescue (1)

SD1118 – Neither --STDTC, --DTC nor --STDY are populated

- Explanation in the cSDRG:

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD1118	Neither DSSTDTC, DSDTC nor DSSTDY are populated	Warning	DS	28(33.33%)	The date was not collected for these subjects.

- ‘Details’ tab: 28 records where DSTERM/DSDECOD = ‘COMPLETED’, VISITNUM = 10 (VISIT = ‘FOLLOW-UP/EARLY TERMINATION’)

Issue Details									
Dataset	Record	Count	Variables	Values	Rule ID	Publisher ID	Message	Category	FDA
DS	2		STUDYID, DOMAIN, USUBJID, DSTERM, DSDECOD, DSCAT, VISITNUM, DSDTC, DSSTDTC, DSSTDY	ABC-123, DS, ABC-123-001, COMPLETED, COMPLETED, DISPOSITION EVENT, 10, null, null, null	SD1118	FDAC146	Neither DSSTDTC, DSDTC nor DSSTDY are populated	Presence	Warning

Example: Search and Rescue (2)

- Further investigation revealed the following:
 - Protocol – 28 planned subjects
 - TS domain - TSVAL = 28 where TSPARMCD/TSPARM = 'ACTSUB'/ 'Actual Number of Subjects'
 - DM domain – 28 randomized and treated subjects
 - RFSTDTC = First dose, populated for all subjects
 - ARMCD/ARM, ACTARMCD/ACTARM = Values based on planned treatment arms in the study
 - DS domain – Subjects had 2 records where DSDECOD = 'COMPLETED' where DSSTDTC was populated for one record and not the other

USUBJID	DSSEQ	DSTERM	DSDECOD	VISITNUM	VISIT	DSSTDTC
ABC-123-008	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	1	SCREENING	2015-04-16
ABC-123-008	2	COMPLETED	COMPLETED	10	FOLLOW- UP/EARLY TERMINATION	
ABC-123-008	3	COMPLETED	COMPLETED	11	STUDY COMPLETION	2015-09-23

Example: Search and Rescue (3)

- 2 CRFs for study completion
 - VISIT = 'STUDY COMPLETION' – collected a date
 - VISIT = 'FOLLOW-UP/EARLY TERMINATION' – did not collect a date
- If issue had been investigated, the explanation in the cSDRG may have been the following:

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD1118	Neither DSSTDTC, DSDTC nor DSSTDY are populated	Warning	DS	28(33.33%)	This check fired for all 28 subjects enrolled where DSDECOD = 'COMPLETED' and DSSTDTC was null. There were two records for each subject where DSDECOD = 'COMPLETED' for the VISITs = 'FOLLOW-UP/EARLY TERMINATION' and 'STUDY COMPLETION'. DSSTDTC was collected on the CRF for the 'STUDY COMPLETION' visit only but the term 'COMPLETED' was collected twice for both visits. In the protocol, these two visits are defined as one visit. The records in DS where DSSTDTC is populated indicates that all completion information was collected for each subject, accordingly.

Best Practices (1)

- Keep a log of specific checks that typically appear with the example explanations provided. Use best judgement before just copy and pasting.
- Maintain a knowledge base of validation checks
 - Authored by different stakeholders (e.g. data management, programmers, and publishers) that would best be able to explain which issues might be a programming/mapping issue that should be fixed or needs to be explained
- Review both the 'Rule Message' and the 'Rule Description' in conjunction with the 'Details' tab in the report for a particular rule.
- Refer to the records in the physical dataset that were flagged.
 - Reviewing the entire record may be helpful
- Refer to the protocol, define.xml, and aCRF if necessary. Sometimes reviewing the dataset is not sufficient.

Best Practices (2)

- Be transparent!
 - Provide as much detail as necessary, e.g. USUBJID, Variable values flagged, VISIT information, etc.
 - Assure that the information provided is sufficient as to not bring questions from a reviewer.
 - Construct the explanation in a way that it can be understood without looking in the 'Details' tab.
- Remember that the base SDTM standard is contained in two documents, the SDTM and the SDTMIG!
 - Except for a few FDA-specific rules (e.g. adding EPOCH to domains), the validation rules were developed based on the information in BOTH documents.
 - Keep these documents handy when reviewing reports and authoring explanations for validation issues.
 - **HINT:** Section 4 contains many of the conventions that are being checked by validation rules.

Pop Quiz!

1. CT2002 – CMDOSU value not found in 'Unit' extensible codelist
 - 'Details' tab
 - 'OTHER', 'DROPS' and 'UNK' values flagged
 - Recommended Action:
 - A. Fix
 - B. Fix with modification and explain
 - C. Explain
 - Provide explanation in the cSDRG

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
CT2002	CMDOSU value not found in UNIT codelist	Warning	CM	3 (<0.1%)	This rule fired for 'OTHER', 'DROPS', and 'UNK' values in CMDOSU. These values could not be mapped to CDISC CT but the UNIT codelist is extensible.

Pop Quiz!

2. SD0037 - Value for variable not found in user-defined EGTESTcodelist
 - 'Details' tab
 - 'ECG Maximum Atrial Rate' value flagged
 - Recommended Action:
 - A. Fix
 - B. Fix with modification and explain
 - C. Explain
 - Update EGTEST codelist in the define.xml by adding 'ECG Maximum Atrial Rate'

Conclusion

- Per the TCG, it is important to include a Reviewer's guide with the SDTM tabulation data to help further explain study conduct as well as results from automated validation tools.
- Because each issue that remains is to be explained, it is important that comprehensive explanations are provided to ensure transparency and to increase reviewability.
 - The more detailed an explanation, the less questions will be asked, which in turn, could shorten the duration of a review.
- Though the cSDRG is one small piece of a submission, it is an integral part of facilitating review in an effort to get drugs to patients faster.

Questions?

Name: Kristin Kelly

Organization: Pinnacle 21, LLC

E-mail: kkelly@pinnacle21.com

Name: Michael Beers

Organization: Pinnacle 21, LLC

E-mail: mbeers@pinnacle21.net