

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

Kristin Kelly, Pinnacle 21, LLC

US PhUSE Connect 2018

Paper DS13

# Agenda

Introduction

To Fix or Not to Fix

Investigation Example

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)

### SD1043 – Inconsistent value for --TESTCD within –TEST

- This rule is based on FDA Business Rule, FDAB009:
  - ‘All paired variables must have a one-to-one relationship. Examples include Short Name and Name of Test; Parameter Name and Parameter Code or Number; Variable Name and Variable Label, etc’
- There can be only one –TESTCD value for a given –TEST
- Sometimes mistaken for a data oddity that cannot be fixed

| Check ID | Diagnostic Message                            | Severity | Dataset | Count (Issue Rate) | Explanation  |
|----------|-----------------------------------------------|----------|---------|--------------------|--------------|
| SD1043   | Inconsistent value for --TESTCD within --TEST | Error    | LB      | 2(4%)              | As per data. |

- Should not be listed in the cSDRG at all since it's a programming issue that should be fixed.

## To Fix or Not to Fix? (4)

### SD0013 – --STDTC is after –ENDTC

- Rule appears in validation report due to a data oddity and not a programming error

| Check ID | Diagnostic Message           | Severity | Dataset | Count (Issue Rate) | Explanation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----------|------------------------------|----------|---------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SD0019   | --STDTC is after -<br>-ENDTC | Error    | AE      | 3 (6.67%)          | <p>This check fired for 3 records in AE where AESTDTC is after AEENDTC:</p> <p>USUBJID = XYZ-008, AESTDTC = 2011-09-14, AEENDTC = 2011-09-12</p> <p>USUBJID = XYZ-300, AESTDTC = 2011-05-23, AEENDTC = 2011-05-19</p> <p>USUBJID = XYZ-415, AESTDTC = 2011-09-05, AEENDTC = 2011-10</p> <p>For the first two records flagged, the start and end dates were missed during data cleaning before database lock.</p> <p>For the third record, though AEENDTC of '2011-10' is after AESTDTC of '2011-09-05', the dates cannot be compared because AEENDTC is a partial date. A complete date was not obtained prior to database lock.</p> |

# 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 | <a href="#">SD1118</a> | 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<br>CONSENT<br>OBTAINED | INFORMED<br>CONSENT<br>OBTAINED | 1        | SCREENING                          | 2015-04-16 |
| ABC-123-008 | 2     | COMPLETED                       | COMPLETED                       | 10       | FOLLOW-<br>UP/EARLY<br>TERMINATION |            |
| ABC-123-008 | 3     | COMPLETED                       | COMPLETED                       | 11       | STUDY<br>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 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. SD1011 – Invalid ISO 8601 value for PCELT<sub>M</sub> variable
  - ‘Details’ tab
    - ‘P24H’ value flagged
  - Recommended Action:
    - A. Fix
    - B. Fix with modification and explain
    - C. Explain
  - Update PCELT<sub>M</sub> = ‘PT24H’

# Pop Quiz!

1. CT2002 – LBTESTCD value not found in 'Laboratory Test Code' extensible codelist
  - 'Details' tab
    - 'LBALL' value flagged
  - Recommended Action:
    - A. Fix
    - B. Fix with modification and explain
    - C. Explain
  - Provide explanation in the cSDRG

| Dataset | Diagnostic Message                                                     | Severity | Count | Explanation                                                  |
|---------|------------------------------------------------------------------------|----------|-------|--------------------------------------------------------------|
| LB      | LBTESTCD value not found in 'Laboratory Test Code' extensible codelist | Warning  | 276   | ① LBALL used to indicate Test Not Done cases as per SDTM IG. |

## 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](mailto:kkelly@pinnacle21.com)