PharmaSUG 2019 - Paper SS-092

A Practical Guide to the Issues Summary in the Data Conformance Summary of Reviewer's Guides

Gary E. Moore, Moore Computing Services, Inc.

ABSTRACT

In the Reviewer's Guides for SDTM (SDRG) and ADaM (ADRG), the Data Conformance Summary is based on validation tools like the Pinnacle 21 Validator (formerly OpenCDISC) and/or proprietary in-house validators. This paper will discuss the purpose of a Reviewer's Guide. Using real life examples, it will illustrate when issues should be resolved by data modifications and when it is appropriate to provide explanations for non-compliance. The paper will contain ideas on the proper way to describe non-compliance issues.

INTRODUCTION

There are 3 sets of documents published by the FDA and Phuse that can be referenced for reporting conformance issues in the reviewer's guides: The **Study Data Technical Conformance Guide**[1], the **SDRG Template and Completion Guidelines**[2], and the **ADRG Template and Completion Guidelines**[3].

The **Study Data Technical Conformance Guide** describes the Study Data Reviewer's Guide (SDRG) and the Analysis Data Reviewer's Guide (ADRG) as follows:

> The SDRG should describe any special considerations or directions or conformance issues that may facilitate an FDA reviewer's use of the submitted data and may help the reviewer understand the relationships between the study report and the data.
>
> The ADRG provides FDA reviewers with context for analysis datasets and terminology, received as part of a regulatory product submission, additional to what is presented within the data definition folder (i.e., define.xml). The ADRG also provides a summary of ADaM conformance findings. The ADRG purposefully duplicates limited information found in other submission documentation (e.g., the protocol, statistical analysis plan (SAP), clinical study report, define.xml) in order to provide FDA reviewers with a single point of orientation to the analysis datasets. It should be noted that the submission of an ADRG does not eliminate the requirement to submit a complete and informative define.xml file corresponding to the analysis datasets.

The **SDRG Template** reports the conformance issues in section 4.2

4.2 Issues Summary

Dataset	Diagnostic Message	Severity	Count	Explanation

The **SDRG Completion Guidelines** instructions:

> 4.2. Issues Summary
>
> This required section summarizes findings from OpenCDISC* validation rules and/or corresponding sponsor-defined validation rules.
>
> • Insert findings from an SDTM conformance report (e.g., the OpenCDISC* report's Issues Summary tab or similar) into the table provided.
>
> • Annotate issues with a brief, non-technical explanation of the findings.
>
> • Do not include skipped validation checks or validation checks for which datasets do not exist.
>
> • If you are using an SDTM conformance checker other than OpenCDISC*, report only the diagnostic messages for validation rules that overlap with the OpenCDISC* rules.

• If your conformance diagnostics do not include severity, leave that column blank.

The **ADRG Template** reports the conformance issues in section 6.2

6.2 Issues Summary

(insert your text here and/or use following table)

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Count/ Issue Rate	Explanation

The **ADRG Completion Guidelines** instructions:

6.2 Issues Summary

This required section summarizes compliance findings.

• Summarize findings from an ADaM conformance report (e.g., the OpenCDISC* report's Issues Summary tab or similar) in table form. The table below may be used and sponsors may include additional columns if desired.

• List only those findings that appear in the submission.

• Annotate issues with a brief, non-technical explanation of the findings.

• Do not include skipped validation checks or validation checks for which datasets do not exist.

• If your conformance diagnostics do not include severity, leave that column blank.

• If non-automated issues were detected, these should be explained as well.

What does this mean to the author of the Reviewer's Guides? The Reviewer's Guides are to assist the FDA reviewer in understanding the data by providing a single document that includes a comprehensive description of the data including any conformance issues that the data may have. They provide FDA reviewers with a single point of orientation and annotate issues with a brief, non-technical explanation of the findings. Reporting the conformance issues and accomplishing these goals is the main focus of this paper.

	Goals For Conformance Issues Annotation
1	Help the reviewer understand the conformance issues the data may have.
2	Provide a single document, a single point of orientation for referencing the conformance issues.
3	Annotate issues with a brief, non-technical explanation of the findings.

** Note - The SDRG and ADRG Completion Guidelines still refer to Pinnacle 21 as OpenCDISC.*

PLANNING AND TIMING

When is the correct time to deal with conformance issues?

Some studies concentrate on SDTM and ADaM dataset creation and tables, listings, and figure creation and worry about conformance issues at the end of the project. This is the wrong time to address conformance issues. At this point, correcting any conformance issues will create a lot of downstream work. If they are not corrected, there will be an overabundance of issues and the need for annotations to explain them. What answer do you give the FDA reviewer when they ask why are there so many conformance issues? Typically, the true answers are poor planning, lack of time, or lack of budget. None

of these are good answers. When the reviewer asks to have these conformance issues fixed it will still take more time and budget. So, how do we fix this? We start with planning. Start the review of conformance issues from the beginning of SDTM and ADaM creation and continue it on an ongoing basis throughout the life of the project. If planned correctly, the conformance issues at the end of the project will be minimal.

SEVERITY LEVELS AND RATES

There are 3 severity levels in most compliance reports. The only severity levels that have to be reported in the Reviewer's Guide are the ERROR and WARNING levels. NOTE levels can be excluded from the Reviewer's Guide. The goal would be to remove all of the compliance issues, but that is often not obtainable. When it is not possible to remove all of the compliance issues, the goal is to remove as many of the ERROR and WARNING messages as possible and to minimize the Count/Issue Rate to the lowest possible amount.

EXAMPLES

The examples are being used as a guide to develop a strategy for handling and reporting compliance issues in the Reviewer's Guide. They are not based on a specific implementation or version of CDISC SDTM or ADaM. They are not based on a specific validator version. They are real world examples that have been sent by statistics and programming staff (CRO or internal) for review prior to submission.

TAKE CARE OF THE EASY STUFF

There are compliance issues that can easily be handled at any point in the life of the project (even the at the end). The work in annotating the compliance issue is almost as much work as correcting the issue in the dataset.

The following issues should always be corrected. They produce an Error level compliance issue that is easily corrected.

Example 1 – Use the Correct Formats

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	*TM does not have the ADaM required SAS time format	Error	We used TOD5 format rather than TIME5. It is fine to use the same as both are SAS time format. [WRONG]

In the format issue example, the time variable is stored as a SAS time value and the format does not affect the value or the variable, just the way it is displayed. There is no reason to not be compliant.

Example 2 – Use the Correct Labels

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0018: Variable label mismatch between dataset and ADaM standard	Error	[WRONG/Sometimes OK]

In the label issue example, the label does not affect the value of the variable. Usually there is no reason to not be compliant. There are some additional examples where labels do present compliance issues.

Example 3 – Use the Actual Lengths for Character Variables (<=200 Characters)

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
SDTM	SD1082: IETEST variable length is too long for actual data	Error	Not an Issue [WRONG]

Consider the nature of the data, and apply reasonable, appropriate lengths to variables. Use the actual required length for character variables.

Example 4.a – Include All Required Variables

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD1001: Required variable is not present	Error	ARM variable was not included [WRONG]

Include all required variables even if they are not used for production of output. ARM should have a similar definition to TRT01P. Look at the protocol, the SAP, the randomization and correctly populate ARM.

Example 4.b – Include All Required Variables

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0138: : CRITy is present and CRITyFL is not present	Error	CRITyFL was omitted. [WRONG]

Include all required variables even if they are not used for production of output. Variable CRITyFL must be present on the dataset if variable CRITy is present.

Example 5.a – Delete All Non-Required Variables That Were Not Used

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0177: Multiple baseline records exist for a unique USUBJID,PARAMCD,BASE TYPE	Error	BASETYPE was not used in ADLB, PARCAT1 was used. [WRONG]

When a non-required variable is creating a compliance issue and that variable was not used for output creation it should be dropped from the dataset.

Example 5.b – Delete All Non-Required Variables That Were Not Used

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
SDTM	SD1101: Missing MHENTPT variable, when MHENRTPT variable is present	Error	According to SDTM IG 3.1.3, MHENTPT is a permissible variable and it wasn't required on the subsequent analyses so was not kept in the dataset [WRONG]

MHENTPT was dropped because it was a permissible variable and was not needed. But, the explanation did not explain the problem. The problem is that MHENRTPT exists on the dataset when MHENTPT does not. MHENRTPT is the end relative to the reference time point and it cannot exist on the dataset without the MHENTPT. If you drop MHENTPT, then drop MHENRTPT instead of annotating the compliance error.

Example 6.a – Choose Proper Variable Names

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0185: illegal variable name: y is not in [1-9] for PARCATy(N)	Error	PARCAT was used instead of PARCAT1. [WRONG]

Rename the PARCAT to PARCAT1.

Example 6.b – Choose the Proper Variable

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0221: Inconsistent value for AVALCAT1	Error	AVALCAT1 was used instead of PCHGCAT1. [WRONG]

Rename the AVALCAT1 to PCHGCAT1.

TAKE CARE OF THE SLIGHTLY MORE COMPLICATED EASY STUFF

These are also easily handled, but sometimes require more documentation.

Example 7 – Explain or Fix Data Issues

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
SDTM	SD0035: Missing value for CMDOSU, when CMDOSE, CMDOSTXT or CMDOSTOT is provided	Error	This error is related to patient 376-040-0571. In this instance, dose units were correctly entered into the CMTRT, however, the unit “MG” was not entered into the CMDOSU variable. [RIGHT/WRONG]

The explanation is a good example of the correct balance between the level of detailed for a single source document and is still a brief and non-technical explanation. This particular issue would be very easy to correct by adding the unit ‘MG’ into the CMDOSU variable. But, depending on the work environment, this might require a Note to File that will be entered into the Trail Master File to explain why this was not collected in the database and had to be handled programmatically.

Example 8 – Use the Correct Variable Data Types

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0058: *DT is not a numeric variable	Error	Character (not numeric) fields were used for some of the Date/Time variables in the database; however, these impacted variables were only used for generation of listings, not for any of the summary tables .When 'analysis day' was required for any respective analysis, the numeric value was calculated separately. [WRONG]

Maintain the compliance data types. *DT is for the numeric SAS Date/Time value. There is no reason that the character versions of Date/Time were not populated in the *DTC variable and used as needed for the appropriate listings.

PLANNING PART 2

Example 9 – Ensure Version Compatibility

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	Variable label mismatch between dataset and ADaM standard	Error	ADaM IG V1.1 was followed in the study whereas pinnacle report is based on ADaM IG V1.0. Thus this discrepancy is present. [WRONG]

Using the latest version makes us look like we are ahead of the curve. But, the version of the SDTM, ADaM, and the validator/compliance checker should be determined prior to development. Typically, the data version should be in line with the start of the data creation. Why would you create a version of the data that is ahead of your validator? To avoid mismatches and potential errors in the compliance report, ensure that everything is compatible.

Example 10 – Choose Proper Variable Names

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0043: *DTM does not have the ADaM required SAS Datetime format	Error	PREPDTM is a time variable and date is not included. [UNFORTUNATE]

With planning and earlier review of compliance issues, this variable could have been renamed to something that did not have a reserved variable name suffix. In this case, Preparatory Dose Time was abbreviated to PREPDTM and creates a variable with the reserved variable name suffix. *DTM is reserved to indicate Date/Time variables. A simple renaming of the variable to PREPDSTM would avoid this compliance ERROR.

Example 11.a – Choose the Proper Values

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0124: Inconsistent value for PARCAT1 within a unique PARAMCD	Error	A given PARAMCD, for example BILI, can be from PARCAT1 = SERUM CHEMISTRY or URINALYSIS. The specific PARAM was used to distinguish these. [WRONG]

Providing unique PARAMCDs for PARCAT1 would correct this issue. For PARCAT1=SERUM CHEMISTRY keep PARAMCD=BILI and for PARACAT1=URINALYSIS use PARAMCD=UBILI. This is a common issue in lab data. This could actually be started at the SDTM level with the TESTCDs. This also should have created a compliance issue with the 1:1 mapping of PARAMCD and PARAM.

Example 11.b – Choose the Proper Values

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0141: Inconsistent value for PARAM within a unique PARAMCD	Error	Each PARAMCD has a PARAM value with and without units, except for URINALYSIS. [WRONG]

Another common lab data issue is that the units are not recorded on the NOT DONE or similar records and when PARAM is created from SDTM.LBTEST and LBSTRESU, you will get 2 different values of PARAM for the same PARAMCD. This should be fixed at the SDTM level by populating the LBSTRESU for all non-missing LBTEST or by utilizing the LBTESTCD=LBALL as described in the SDTMIG.

Example 12 – Include All Required Variables

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0198: Neither AVAL nor AVALC are present in dataset	Error	AVAL, AVALC were not needed. [WRONG]

Include all required variables even if they are not used for production of output. AVALC could easily be populated with the relevant data that is obviously being used in somewhere else in the dataset.

SOMETHINGS ARE UNAVOIDABLE

Some compliance issues are unavoidable. Usually, these are issues with the study design, the database design, or the Statistical Analysis Plan (SAP). If something just does not exist in the data collected, you cannot include it in the SDTM. Sometimes the data collected has values that just are not compliant and to be true and consistent to the way the data was collected you have to include it and annotate it. Sometimes the SAP has instructions for data handling that are not compliant, but we have to follow the procedures of the SAP and annotate the non-compliance. There are exceptions. An example is acquiring a Note to File (NTF) for data exceptions and program around these.

Example 13 – SAP: Mapping of AVAL and AVALC and Inequalities

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0149: Inconsistent value for AVALC	Error	When AVALC has "<" "<=" ">" ">=" the symbol is stripped and then AVAL is the numeric part resulting in a 1 to many mapping. For example, for PARAMCD=BILI, AVALC may be 1.7 or <1.7 and AVAL in both cases is 1.7. [CORRECT]

Inequalities are a common compliance issue in the character and numeric results of lab data. They can be left blank in the numeric data, but this will require extra programming in the summary program. Many times, the SAP specifically says to remove the inequalities for the numeric results so that they can be summarized. The explanation above could have additionally referenced the SAP section and the Derivation in the Define.XML.

Example 14 – SAP and Data: They Are What They Are

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
SDTM	SD0006: No baseline result in MB for subject	Warning	These subjects had "No Growth" at the local lab for microbiology data, so these subjects had no baseline for microbiology data as defined in the SAP. The data was reported as received. [CORRECT]

Example 15.a –Data: It Is What It Is

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
SDTM	SD0026: Missing value for FAORRESU, when FAORRES is provided	Warning	There are no units recorded in the data when FAORRES is qualitative and not quantitative. [CORRECT]

You can only provide the data that was collected.

Example 15.b –Data: It Is What It Is

This is a non-validator example that could have been in Planning, but was included here as a data issue. Someone contacted me because the CRO they were working with was generating visits in SDTM.SV.

Question:

They are telling me that in the SDTM SV - that if a subject does not have a visit, they still have to put the visit in the database and use the last date the subject had a visit. For example:

Day 7 - Jan 1, 2018, Day 14 - Jan 7, 2018, EOT - Jan 14, 2018, TOC - Not done, FU - not done

They are coding TOC and FU to Jan 14, 2018. Which as you can imagine makes ALL of the ADaMs wrong. I know this is wrong because it is making up data, but I need you to tell me the CRO is full of ….

Response:

They are full of ….

From the SDTMIG v3.2

3.2.1 Table 3.2.1 SDTM Submission Dataset-Definition Metadata Example

Dataset	Description	Class	Structure	Purpose	*1*K*838H*	Location
SV	Subject Visits	Special Purpose Domains	One record per actual visit per subject	Tabulation	STUDYID, USUBJID, VISITNUM	sv.xpt

5 Models for Special-Purpose Domains…

Domain Code	Domain Description	Domain Document Name
SV	**Subject Visits** The subject visits table describes the actual start and end data/time for each visit of each individual subject.	Section 5 - SV Domain

Subject Visits (SV

SV – Description/Overview for the Subject Visits Domain Model…

The Subject Visits dataset provides reviewers with a summary of a subject's Visits. Comparison of an individual subject's SV dataset with the TV dataset ***[Section 7.3 - Schedule for Assessments: Trial Visits (TV)],*** which describes the planned Visits for the trial, quickly identifies missed Visits and "extra" Visits. Comparison of the values of STVSDY and SVENDY to VISIT and/or VISITDY can often highlight departures from the planned timing of Visits.

SV – Assumptions for the SUBJECT VISITS Domain Model

1. The Subject Visits domain allows the submission of data on the timing of the trial visits a subject actually passed through in their participation in the trial…

CHOOSE APPROPRIATE WORDING

It is important to be transparent with the Reviewer about compliance issues. But, the negative or positive response of the Reviewer may depend on the professional tone provided in the annotations. Choosing appropriate wording for the annotation is important for setting this tone.

Example 16.a – We Know, But Did It Anyway

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
SDTM	SD1082: IETEST variable length is too long for actual data	Error	We understand that according to the guidance, the variable length is longer than expected, but the size of the datasets are all still under the maximum limitation (1gb) and hence consistent length for this variable is maintained. [WRONG]

Understanding that it was wrong, but you did it anyway is never a good explanation. These should be corrected for length.

Example 16.b – We Know, But Did It Anyway

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
SDTM	SD1077: FDA Expected variable not found	Warning	[POOR] The SDTM IG 3.1.3 states that EPOCH variable is permissible and not required, so it was not implemented in SDTM domain. We understand that EPOCH is for reviewer's purpose; however, our final analysis does not rely on the missing variable (EPOCH). [BETTER] The SDTM IG 3.1.3 states that EPOCH variable is permissible and not required. Since this study was started prior to the FDA March 2016 Guidance and EPOCH does not affect our final analysis, it was not implemented in the SDTM domain.

The **Poor** explanation states that they understood this was expected because it made the review easier and decided not to include it anyway. The **Better** version states just the facts of the study without the statement of understanding.

Example 17 – Overly Simplistic Explanation

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	AD0193: illegal variable name: y is not in [1-9] for CRITy(FL/FN)	Error	[POOR] Not an issue: ADaM IG v1.1 was followed in the study whereas pinnacle report is based on ADaM IG v1.0. [BETTER] There are more than 9 criteria flags, CRITyFL, in the dataset. The validator is using ADaMID v1.0. ADaM IG v1.1; Sect. 3.1.1; 2.c - states that the lower-case letter "y" in a variable name is replaced with an integer [1-99, not zero-padded].

The **Poor** explanation assumes that the Reviewer has the ADaMIG memorized or is willing to leave this document to look it up. In order to provide the Reviewer with "a single document, a single point of orientation for referencing the conformance issues", the **Better** explanation gives the ADaMIG reference and relevant text.

Example 18 – Structure the Explanation: Most Relevant to Least Relevant

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Explanation
ADaM	SD0080: AE start date is after the latest Disposition date	Warning	[GOOD] Study was designed with phone follow up visit which is 30 days after last dose. The study disposition was collected at late follow up visit (Day 21 to Day 28). This study had 48 instances (AEs) that occurred between late follow up visit and phone follow up visit had start date later than disposition date. [BETTER] Last disposition date is the late follow up visit (day 21-28). AEs continued to be captured between the late follow up visit (day 21-28) and the 30 day phone follow up. There were 48 AEs captured during this interval.

The **Good** explanation describes the compliant issue well, but starts the explanation with non-relevant information. The Warning is about the latest Disposition date and the AE start date. The **Better** explanation begins with the specific details of the Warning and adds additional information as needed.

CONCLUSION

We looked at multiple examples to assist in reaching our conformance issues annotation goals, to help the review, provide a single source for reference, and to annotate issues with a brief, but thorough explanation of the findings. The examples have shown that planning and timing of compliance verification is important in order to provide maximum benefit with appropriate effort. Dataset creation should minimize the ERRORs and WARNINGs. The rate among the ERRORs and WARNINGs that are unavoidable should be minimized. We should always take care of the easy stuff. Document, explain, and fix the complicated easy stuff. Plan the dataset contents, the variable naming and values, and ensure version compatibility between datasets and the validator. Document and explain the unavoidable compliance issues. Choose the appropriate wording when explaining compliance issues.

If we can successfully follow these examples, we can create datasets and compliance explanations in the Reviewer's Guides that will maximize the benefit to the Reviewer.

REFERENCES

1: The Study Data Technical Conformance Guide. U.S. Department of Health and Human Services; Food and Drug Administration; Center for Drug Evaluation and Research (CDER); Center for Biologics Evaluation and Research (CBER); March 2018; https://www.fda.gov/media/88173/download

2: Study Data Reviewer's Guide Final Work Packages; Phuse Wiki; SDRG Package v1.2 2015-01-26; https://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide

3: Analysis Data Reviewer's Guide Final Work Package; Phuse Wiki; ADRG Package v1.1 2015-01-26; https://www.phusewiki.org/wiki/index.php?title=Analysis_Data_Reviewer%27s_Guide

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:	Gary E. Moore
Enterprise:	Moore Computing Services, Inc.
Address:	2006 Beckenham Cove
City, State ZIP:	Little Rock, AR 72212
Work Phone:	501-225-8689
E-mail:	gary@moorecsi.net
Web:	www.moorecsi.net