PharmaSUG China 2023 - Paper PO-136

Tips for Better Understanding and Resolving Pinnacle 21 Report Issues

Kaiying Li, Jiangsu Hengrui Pharmaceuticals Co.,Ltd., China

ABSTRACT

Pinnacle 21 validation tool is a well-known OpenCDISC Validator that provides compliance checks for the submission package against CDISC conformance rules and regulatory agency requirements. Pinnacle 21 validator provides a validation report that contains information about the submission package and the details of issues triggered in each dataset. Whether the reported issues should be fixed or documented in the Reviewer Guide remains a hot topic for sponsors, especially when preparing regulatory submissions. This paper will give an overview of the basics of Pinnacle 21 Validation rules, what these rules are, and how we programmers can quickly understand and check these rules with the relevant conformance rules. Additionally, this paper will discuss appropriate solutions for different types of Pinnacle 21 reported issues based on my experience and further illustrate with examples. This paper is intended to help programmers facilitate the process of data validation and improve the quality of submission packages by using Pinnacle 21 validation tool.

INTRODUCTION

Study data validation is one of the important steps during the preparation of submission data. It helps to ensure the study data is compliant, and to improve the quality of submission data to avoid the impact of regulatory review process. Regulatory agencies expect sponsors to evaluate the study data before submissions against the data conformance rules, which are the standards or the regulator’s requirements. Sponsors should either correct any discrepancies between their study data and the conformance rules or detailly explained the meaningful discrepancies in the relevant Reviewer Guide.

As a clinical statistic programmer, you will use Pinnacle 21 validation tools to identify the data conformance issues and check the accuracy of programming logic. Pinnacle 21 validation tool also helps to ensure your data packages meet regulatory requirements before data submission. Even if you put a lot of effort into your study data and the programming codes work perfectly, issues still trigger after running the validations. Sometimes these issues are triggered by mapping or programming errors that should be solved by our programmer. However, sometimes these triggering issues cannot be easily solved by programming because they are caused by the design of database or the true data issues. For the triggered issue, the first consideration is to solve it, not only redo and update the programming codes to fix the errors, but also trace back to the source of raw data and respond to the data management (DM) department if necessary. Explaining the reported issues in the Reviewer Guide should be the last choice for sponsors.

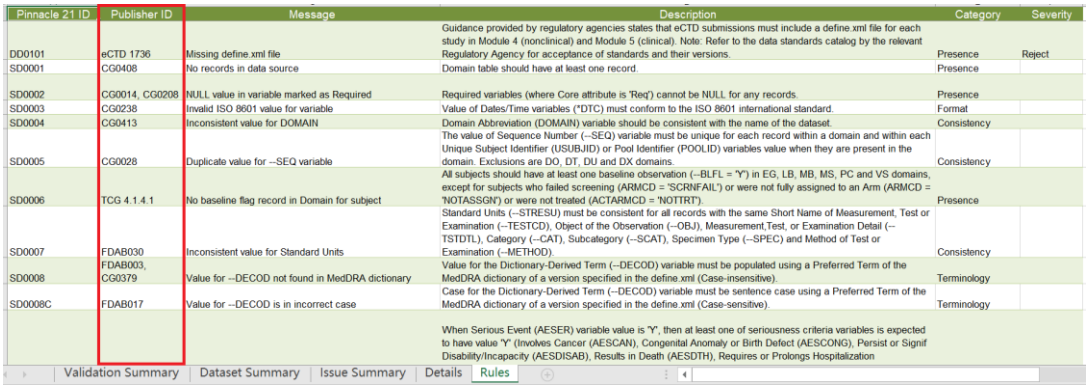
This paper will discuss based on Pinnacle 21 Community validation tool. Understanding how the validation rules constitute and what each rule checks for is critical to fixing or documenting the triggered issues. This paper will cover my experience in using Pinnacle 21 Community

and provide resolutions and new perspectives on different types of report issues that help improve the quality of study data and the submission package.

BASIC KNOWLEDGE OF PINNACLE 21 VALIDATION RULES

Nowadays, Pinnacle 21 Community provides validation engines that enable sponsors to ensure their study data complies with three regulatory agencies, which are the Food and Drug Administration (FDA), the Japanese Pharmaceutical and Medical Devices Agency (PMDA), and the National Medical Products Administration (NMPA). Each validation engine is versioned based on the year and month released and a patch number. Pinnacle 21 recommends sponsors use the latest engine to validate study data in preparation for the imminent submission.

Pinnacle 21 validation rules are mainly composed of different types of conformance rules, which are CDISC conformance rules, business and validator rules of regulatory agencies, the Technical Conformance Guide, and eCTD Technical Rejection Criteria for Study Data. You can check the source of the validation rules by Publisher ID in the Rules tab in the validation reports. The Publisher ID is a key that helps us refer back to the related conformance rules and clarify what the rule is checking.

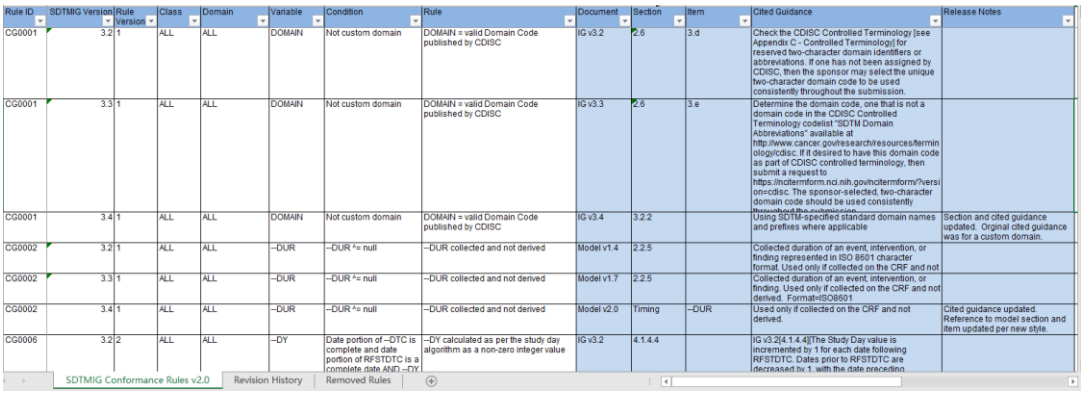


Pinnacle 21 ID	Publisher ID	Message	Description	Category	Severity
DD0101	eCTD 1736	Missing define.xml file	Guidance provided by regulatory agencies states that eCTD submissions must include a define.xml file for each study in Module 4 (nonclinical) and Module 5 (clinical). Note: Refer to the data standards catalog by the relevant Regulatory Agency for acceptance of standards and their versions.	Presence	Reject
SD0001	CG0408	No records in data source	Domain table should have at least one record.	Presence	
SD0002	CG0014, CG0208	NULL value in variable marked as Required	Required variables (where Core attribute is 'Req') cannot be NULL for any records.	Presence	
SD0003	CG0238	Invalid ISO 8601 value for variable	Value of Dates/Time variables (*DTC) must conform to the ISO 8601 international standard.	Format	
SD0004	CG0413	Inconsistent value for DOMAIN	Domain Abbreviation (DOMAIN) variable should be consistent with the name of the dataset.	Consistency	
SD0005	CG0028	Duplicate value for --SEQ variable	The value of Sequence Number (--SEQ) variable must be unique for each record within a domain and within each Unique Subject Identifier (USUBJID) or Pool Identifier (POOLID) variables value when they are present in the domain. Exclusions are DO, DT, DU and DX domains.	Consistency	
SD0006	TCG 4.1.4.1	No baseline flag record in Domain for subject	All subjects should have at least one baseline observation (--BLFL = 'Y') in EG, LB, MB, MS, PC and VS domains, except for subjects who failed screening (ARMCD = 'SCRNFAIL') or were not fully assigned to an Arm (ARMCD = 'NOTASSGN') or were not treated (ACTARMCD = 'NOTTRT').	Presence	
SD0007	FDAB030	Inconsistent value for Standard Units	Standard Units (--STRESU) must be consistent for all records with the same Short Name of Measurement, Test or Examination (--TESTCD), Object of the Observation (--OBJ), Measurement,Test, or Examination Detail (--TSTDTL), Category (--CAT), Subcategory (--SCAT), Specimen Type (--SPEC) and Method of Test or Examination (--METHOD).	Consistency	
SD0008	FDAB003, CG0379	Value for --DECOD not found in MedDRA dictionary	Value for the Dictionary-Derived Term (--DECOD) variable must be populated using a Preferred Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	Terminology	
SD0008C	FDAB017	Value for --DECOD is in incorrect case	Case for the Dictionary-Derived Term (--DECOD) variable must be sentence case using a Preferred Term of the MedDRA dictionary of a version specified in the define.xml (Case-sensitive).	Terminology	
			When Serious Event (AESER) variable value is 'Y', then at least one of seriousness criteria variables is expected to have value 'Y' (Involves Cancer (AESCAN), Congenital Anomaly or Birth Defect (AESCONG), Persist or Signif Disability/Incapacity (AESDISAB), Results in Death (AESDTH), Requires or Prolongs Hospitalization		

Validation Summary | Dataset Summary | Issue Summary | Details | Rules

Figure 1. Pinnacle 21 Validation Rules and Rules' Details

Publisher ID can be identified by its prefix. A publisher ID starting with “FDAB” indicates that this validation rule comes from FDA business rules, “TCG” stands for FDA Technical Conformance Guide, and “eCTD” stands for eCTD (Electronic Common Technical Documents) Validation Criteria. The Publisher ID of the CDISC SDTM Conformance rules is prefixed with “CG”. Instead, the Publisher ID of the ADaM Validation rules is typically a number, which is associated with the Check Number in CDISC ADaM Conformance rules.



Rule ID	SDTMIG Version	Rule Version	Class	Domain	Variable	Condition	Rule	Document	Section	Item	Cited Guidance	Release Notes
CG0001	3.2	1	ALL	ALL	DOMAIN	Not custom domain	DOMAIN = valid Domain Code published by CDISC	IG v3.2	2.6	3.d	Check the CDISC Controlled Terminology [see Appendix C - Controlled Terminology] for reserved two-character domain identifiers or abbreviations. If one has not been assigned by CDISC, then the sponsor may select the unique two-character domain code to be used consistently throughout the submission.	
CG0001	3.3	1	ALL	ALL	DOMAIN	Not custom domain	DOMAIN = valid Domain Code published by CDISC	IG v3.3	2.6	3.e	Determine the domain code, one that is not a domain code in the CDISC Controlled Terminology codelist "SDTM Domain Abbreviations" available at http://www.cancer.gov/research/resources/terminology/cdisc. If it desired to have this domain code as part of CDISC controlled terminology, then submit a request to https://inciterm form.nci.nih.gov/inciterm form/?version=cdisc. The sponsor-selected, two-character domain code should be used consistently	
CG0001	3.4	1	ALL	ALL	DOMAIN	Not custom domain	DOMAIN = valid Domain Code published by CDISC	IG v3.4	3.2.2		Using SDTM-specified standard domain names and prefixes where applicable	Section and cited guidance updated. Orginal cited guidance was for a custom domain.
CG0002	3.2	1	ALL	ALL	--DUR	--DUR ^= null	--DUR collected and not derived	Model v1.4	2.2.5		Collected duration of an event, intervention, or finding represented in ISO 8601 character format. Used only if collected on the CRF and not	
CG0002	3.3	1	ALL	ALL	--DUR	--DUR ^= null	--DUR collected and not derived	Model v1.7	2.2.5		Collected duration of an event, intervention, or finding. Used only if collected on the CRF and not derived. Format=ISO8601	
CG0002	3.4	1	ALL	ALL	--DUR	--DUR ^= null	--DUR collected and not derived	Model v2.0	Timing	--DUR	Used only if collected on the CRF and not derived.	Cited guidance updated. Reference to model section and item updated per new style.
CG0006	3.2	2	ALL	ALL	--DY	Date portion of --DTC is complete and date portion of RFSTDTC is a complete date AND --DY	--DY calculated as per the study day algorithm as a non-zero integer value	IG v3.2	4.1.4.4		IG v3.2[4.1.4.4]The Study Day value is incremented by 1 for each date following RFSTDTC. Dates prior to RFSTDTC are decreased by 1, with the date preceding	

SDTMIG Conformance Rules v2.0 | Revision History | Removed Rules

Figure 2. SDTM and SDTMIG Conformance Rules v2.0

HOW TO CHECK AND UNDERSTAND THE VALIDATION RULES IN A FAST WAY

The Pinnacle 21 Validation Report will be generated after each completion of the validation process and record all triggered issues about study data and relevant documents. You can find out what happens in each data set and the total number of observations with those issues in the Issue Summary tab. Additionally, the Detail tab documents the specific problems of each observation. But the message of issues and the description of the validation rules sometimes fail to clarify what exactly is happening in the data set. For example, SD1117 checks duplicated records for the Findings domain. You might be confused why this issue is triggered since the records are all unique under the key variables you provided in define.xml.

In addition to the message of issues and the descriptions of validation rules, the Publisher ID and the source code repository of Pinnacle 21 Community can help you better understand the validation rules. The Publisher ID can be used to check the conformance rule files which will give you a clearer picture of each check rule. Meanwhile, the source code provides the checking logic of Pinnacle 21 Validation Engine that can help to further understand the causes of the issue. You can browse the code for each rule by searching its Pinnacle 21 ID.

The check rule SD1117, as an example, states that the Finding Results should be unique within the same Test Short Name and the same Qualifier variables at the same timepoint for the same Subject. And referring to the related CDISC conformance rule CG0019, each record is unique per sponsor-defined key variables as described in the define.xml. If you confirmed that the key variables listed in the define.xml can make records unique but the issue SD1117 is still triggered, you can search the source code to see whether the key variables are also considered by P21C in its validation logic. After checking that the duplicates are not due to programming issues (e.g. key variables cannot make records unique) or true data issues, you can document this issue in the Reviewer guide and explain the discrepancy between the validation engine and your study data. Cross-checking can help you better understand the validation rules and how the validator checks the study data. When these issues come up again, you can react faster and fix them in an appropriate way.

```
<val:Unique ID="SD1116" PublisherID="FDAB018" Target="Metadata" oce:IgnoreContext="Yes" Message="Inconsistent variable length
within split datasets" Description="Split datasets should have matching variable lengths for future merges. Datasets should be
resized to the maximum length used prior to splitting." Category="Consistency" Variable="LENGTH" GroupBy="VARIABLE" Matching=
"Yes"/>
<val:Unique ID="SD1117" PublisherID="CG0019" Message="Duplicate records" Description="The structure of Findings class domains
should be one record per Finding Result per subject. No Finding Result with the same Test Short Name (--TESTCD) and the same
Qualifier variables at the same timepoint for the same Subject (USUBJID) are expected." Category="Consistency" Variable=
"%Domain%DTC" When="%Domain%TESTCD != '' @and %Domain%CAT != 'INTERPRETATION' @and %Domain%CAT != 'DIAGNOSIS'" GroupBy=
"USUBJID,POOLID,SPDEVID,FETUSID,%Domain%TESTCD,%Domain%OBJ,%Domain%CAT,%Domain%SCAT,%Domain%POS,ICIMPLBL,%Domain%RESLOC,%Domain%
METHOD,%Domain%SPEC,%Domain%ANTREG,%Domain%LOC,%Domain%LAT,%Domain%DIR,%Domain%PORTOT,%Domain%DRVFL,%Domain%EVAL,%Domain%EXCLFL,
VISITNUM,VISITDY,%Domain%ENDTC,%Domain%DY,%Domain%ENDY,%Domain%NOMDY,%Domain%TPTNUM,%Domain%RFTDTC,%Domain%STINT,%Domain%ENINT"
 Optional=
"POOLID,SPDEVID,FETUSID,%Domain%OBJ,%Domain%DTC,%Domain%CAT,%Domain%SCAT,%Domain%POS,ICIMPLBL,%Domain%RESLOC,%Domain%METHOD,%Dom
ain%SPEC,%Domain%ANTREG,%Domain%LOC,%Domain%LAT,%Domain%DIR,%Domain%PORTOT,%Domain%DRVFL,%Domain%EVAL,%Domain%EXCLFL,VISITNUM,VI
SITDY,%Domain%ENDTC,%Domain%DY,%Domain%ENDY,%Domain%NOMDY,%Domain%TPTNUM,%Domain%RFTDTC,%Domain%STINT,%Domain%ENINT"/>
<val:Condition ID="SD1118" PublisherID="FDAB038" Message="Neither %Domain%STDTC, %Domain%DTC nor  %Domain%STDY are populated"
Description="At least one value of three variables Start Date/Time (--STDTC), Collection Date/Time (--DTC) or Start Study Day
(--STDY) should be populated." Category="Presence" Test="%Domain%STDTC != '' @or %Domain%STDY !='' @or %Domain%DTC !='' "
Optional="%Domain%STDY,%Domain%DTC"/>
```

Figure 3. Source Code of Rule SD1117 in P21 Validator Configuration for SDTM-IG 3.3

The CHANGELOG.txt file in the configuration folder records changes in each validation engine. This file tells you which rules have been added or removed and which rules have been updated through bug fixes or maintenance. Knowing the changes in the validation engine facilitate reading the validation report and finding out the updates in the new report that helps you better fix or explain the issues.

The earlier study data validation begins, the easier the process of preparing for submission will be. In addition to being able to detect and solve data issues early, programming errors can also be detected and corrected in a timely manner. Starting data validation too late may result in some data issues not being resolved, or programmers spending more time re-running the programs.

REPORTED ISSUES AND RESOLUTIONS

Understanding the validation rules and how the validator functions are important steps in understanding the P21 Validator Report and solving the reported issues. The issue might be caused by dirty raw data, imperfect Electronic Data Capture (EDC) system, ongoing study status, validation engine bugs, incorrect mapping process, or programming errors. It is difficult for programmers to determine whether the issue can be resolved and whether the issue is related to the data that should be notified to the data management department. When the issue cannot be solved, programmers should provide detailed, study-specific explanations in the Reviewer Guide.

In this section, I would like to discuss the validation rules and how to determine the resolutions based on my experience. And I will choose some common issues as examples to explain the resolutions in detail. Regulatory agencies consider that all validation issues are important and that it is not appropriate to use the severity messages in the validation report to determine whether an issue should be resolved. Therefore, how to choose a proper way to treat the issues depends more on the programmers' experience and what is actually happening in the data. In general, the issues triggered in SDTM and ADaM data validations should be solved or explained in the Reviewer Guide. Any issues that arise in the validation of the define.xml file must all be solved.

In the validator report generated by P21C, validation rules in the Rules tab are divided into eight categories. Rules under each category may have similar properties and resolutions. Table 1 introduces the different categories. Based on these eight categories, you can distinguish rules in the same category, quickly locate related data issues, and jointly solve the related issues to improve efficiency.

Category	Contents
Consistency	Rules are checking the consistency within or between paired variables or cross-domain variables, and also for compliance with the conformance rules.
Cross-reference	Rules in this category are used to confirm the value of variables is valid across domains.
Format	Rules are used to confirm that the variable value meets the required length and that the data or variable is in the correct format and does not contain invalid characters.
Limit	Rules in this category are checking whether the variable is populated in a valid way. The programs of derivation or imputation should be checked carefully to avoid this type of issue.
Metadata	Rules are used to check whether the datasets or the variables are matched standard models and define.xml files.
Presence	Rules are used to check whether the dataset or variable exists as required. These issues have strong relationships with programming logic, and most of them should be fixed.
Structure	This category checks whether the structure of dataset meets the requirements of SDTMIG.
Terminology	Rules check the usage of terminology terms in the study data.

Table 1. Eight Different Categories and Content of Rules in Each Category.

EXAMPLES OF ISSUES DUE TO IMPERFECT PROGRAMMING LOGIC

Generally, there are multiple ways to solve the issues, not just one solution. In some cases, issues should be addressed by updating the programs rather than explained in the Reviewer Guide. But always remember that you should never change or modify the codes to just avoid validation issues when you are sure your programming does not violate CDISC standards.

Example 1 – SD0036: Missing value for --STRESC, when --ORRES is provided

This issue will be triggered when variable –ORRES (Result or Finding in Original Units) is provided but the variable --STRESC (Character Result/Finding in Std Units) is not populated. When –ORRES is collected as inequalities and the units are not the standard units, you might consider that these results cannot be converted to standard unit results, then set the –STRESC as null and explain that data collected with characters attached cause the conversions to standard units fails. However, the SDTM Implementation Guide states that the value with the greater than (>) or less than (<) sign attached should be maintained and the numeric part should be converted when populating –STRESC, and –STRESN should be null. In these cases, the –STRESC should not be set to null when the original units can be

converted to standard units. Please update your programs instead of documenting the explanation in the Reviewer Guide.

lb.xpt

Row	LBTESTCD	LBCAT	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBSTAT	LBLOBXFL	VISITNUM	LBDTC
1	GLUC	CHEMISTRY	6.0	mg/dL	60.0	60.0	mg/L			1	2016-02-01
2	ALT	CHEMISTRY	12.1	mg/L	12.1	12.1	mg/L			1	2016-02-01
3	BACT	URINALYSIS	MODERATE		MODERATE					1	2016-02-01
4	RBC	URINALYSIS	TRACE		TRACE					1	2016-02-01
5	WBC	URINALYSIS	++		2+					1	2016-02-01
6	KETONES	CHEMISTRY	BLQ	mg/L	BLQ		mg/L			1	2016-02-01
7	MCHC	HEMATOLOGY			33.8	33.8	g/dL		Y	3	2016-02-15
8	HCT	HEMATOLOGY						NOT DONE		2	2016-02-08
9	LBALL	HEMATOLOGY						NOT DONE		3	2016-02-29
10	LBALL							NOT DONE		4	2016-02-22
11	WBC	HEMATOLOGY	<4, 000	10^6/L	<4,000		10^6/L			6	2016-02-07
12	BILI	CHEMISTRY	<0.1	mg/dL	<1.71		umol/L			6	2016-02-07

Figure 4. Examples of --STRESC Conversion in SDTM IG v3.4

Example 2 – SD1344: Value for --DECOD not found in WHODrug dictionary

This issue occurs when the value of --DECOD cannot be recognized as a drug name in WHODrug. Firstly, you should rule out the following possibilities. You need to confirm that you have populated –DECOD with the exact term from the WHODrug dictionary. What's more, the version of WHODrug chosen in the validation engine should be consistent with the one used during the mapping process and should be correctly listed in the define.xml. If the problem is not one of the above, you may need to consider whether and how the value of –DECOD is split. For a drug with many ingredients, the generic name may exceed 200 characters, which is the variable length limit in the SAS® V5 export format. SDTMIG states that sponsors can split the long string between words. However, if a long generic name is truncated between words, the validation engine will not be able to recognize the drug name successfully. In this case, the long generic name splitting should be performed after the semicolon to the nearest 200 characters. Each additional 200 characters of the string should be stored in a record in the SUPPQUAL dataset.

Drug Name	Preferred Name	Length
FENGHAN GANMAO GRANULE	ANGELICA DAHURICA VAR. FORMOSANA ROOT;CINNAMOMUM CASSIA TWIG;CITRUS AURANTIUM FRUIT PEEL;EPHEDRA SINICA HERB;GLYCYRRHIZA URALENSIS ROOT WITH RHIZOME;PERILLA FRUTESCENS VAR. CRISPA LEAF;PLATYCODON GRANDIFLORUS ROOT;PRUNUS ARMENIACA SEED;PUERARIA LOBATA ROOT;SAPOSHNIKOVIA DIVARICATA ROOT;ZINGIBER OFFICINALE RHIZOME	314

Table 2. Example of Preferred Name which is longer than 200 characters

USUBJID	CMSEQ	CMTRT	CMDECOD
ABC-01-123	1	FENGHAN GANMAO GRANULE	ANGELICA DAHURICA VAR. FORMOSANA ROOT;CINNAMOMUM CASSIA TWIG;CITRUS AURANTIUM FRUIT PEEL;EPHEDRA SINICA HERB;GLYCYRRHIZA URALENSIS ROOT WITH RHIZOME;PERILLA FRUTESCENS VAR. CRISPA LEAF;

Table 3. Example of CM dataset where the Preferred Name using to populate CMDECOD is longer than 200 characters.

USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
ABC-01-123	CMSEQ	1	CMDECOD1	Standardized Medication Name 1	PLATYCODON GRANDIFLORUS ROOT;PRUNUS ARMENIACA SEED;PUERARIA LOBATA ROOT;SAPOSHNIKOVIA DIVARICATA ROOT;ZINGIBER OFFICINALE RHIZOME

Table 4. Example of SUPPCM dataset saving the additional text of CMDECOD.

EXAMPLES OF ISSUES THAT ARE DATA ISSUES AND SHOULD NOTIFY DM

As I mentioned above, some issues are due to true data issues and you should check with DM to confirm the source data. In most of the cases I have come across, the source data was not entered into the EDC system or was entered with errors. Additionally, the design of EDC system also affects the quality of source data. Therefore, it is necessary to run data validation prior to database locks or data exactions for important milestones and inform DM of the true data issues.

Example 1 – SD0026: Missing value for --ORRESU, when --ORRES is provided

This is the most common case that should notify DM when it occurs. There is a large amount of results data and the original result units are often missed by mistakes, especially for the laboratory test results. This issue can be avoided and should not be left in the dataset. Besides that, it might take too much time to check this issue just by running the validation for each data exaction. You can include a checking algorithm in the programs as well to help check for this data issue so that DM can be notified and help resolve this data issue as early as possible.

Example 2 – SD0021: Missing End Time-Point value

This is a recurring issue. You may consider informing DM immediately when it occurs. But in some cases, missing the end time-point is acceptable, and you should rule out these acceptable cases before notifying DM. What's more, the relative time variables --ENRTPT and --ENTPT can also be used precisely to indicate the explicit timing descriptions or date/time value.

For the adverse event, which is the most common case with missing end time-points, only when the Outcome of Adverse Event (AEOUT) has the value of "NOT RECOVERED/NOT RESOLVED" or "RECOVERING/RESOLVING" or "UNKNOWN" that the AEENDTC can be missing. Other outcomes that do not have an end time-point should inform DM and check that

the date is indeed not collected. For medical history, concomitant medication or procedures, if the “Still Ongoing” question is collected as Yes, missing end time-point is acceptable, and the relative time variables –ENRTPT and –ENTPT can be used. Beyond that, the data without an end date should be notified to DM.

EXAMPLES OF ISSUES CAUSED BY P21 CHECKING LOGIC

Besides programming errors and true data issues, reported validation issues may also be caused by P21’s checking logic. Some of these issues are set by P21 for special purposes to warn users. Some are due to imperfect checking logic. These issues often cannot be fixed, and you can document them in the Reviewer guide.

Example 1 – SD2237: ACTARM does not equal ARM

This issue is always triggered when the subject is randomized but not treated, or when the subject is in a cross-over study with two treatments and multiple periods, but no treatment is given in one or more periods. In the first case, the value of actual treatment arm variables will be set to blank, which is required by FDA TCG. And in the second case, sponsors will often use “NOTTRT” or other values to represent the “Not Treated” period. The ACTARM value is not the same as the ARM value.

There is a discrepancy between FDA TCG and this check rule. FDA TCG states that the ACTARM should be null and ARM should be populated when the subject was randomized in the treatment group but not treated. Lately, P21 has clarified that this rule is a warning and is not a case of false positives. When this rule is triggered and a warning is also triggered, sponsors need to explain the specific situation in the cSDRG (clinical Study Data Reviewer’s Guide) that why the subject was neither treated nor received their planned treatment. For a submission to FDA, you should better follow the FDA TCG and explain this issue in the Reviewer guide.

EXAMPLES OF INAPPROPRIATE EXPLANATIONS FOR ISSUES

When the validation issues persist after the database lock and cannot be solved, sponsors should document these discrepancies between conformance rules in the Reviewer Guide. And it is sponsor’s responsibility to provide appropriate and comprehensive explanations. The explanations of discrepancies should be detailed and avoid simply repeating the wording of the conformance rules.

Example 1 – SD0021: Missing End Time-Point value

After the missing end date has been confirmed by DM, and the relative time variables are not suitable for this case, this issue will still appear in the validation report and should be explained in the Reviewer Guide. Sponsors might simply provide an explanation stating that the end date is unknown.

Dataset	Pinnacle 21 Rule ID	Diagnostic Message	Count	Explanation
CM	SD0021	Missing End Time-Point value	78	The end date is not collected.

Table 5. Example of Inappropriate Explanation for SD0021

This example of explanation fails to provide the reviewers with a clear reason to understand the discrepancies between conformance rules. A comprehensive explanation should include the details of the record that caused this issue and why this validation rule occurred and retained. Sometimes, you may also need to consider the data collection and the design of EDC system. In this case, the End Date is not collected as it is unknown and the question "Still Ongoing" in CRF is also not collected. A full explanation should be provided. Table 6 provides a complete explanation example of SD0021 triggered in CM domain.

Dataset	Pinnacle 21 Rule ID	Diagnostic Message	Count	Explanation
CM	SD0021	Missing End Time-Point value	78	The variables CMENDTC and CMENRTPT are missing due to an unknown end date, and the CM is still ongoing but the ongoing status was not collected.

Table 6. Example of Complete Explanation for SD0021

Example 2 – SD1117: Duplicate records

The issues of duplicate records are also common issues that appear in the validation reports. Rule SD1117 reports duplicate records for Findings domain. What's more, rule SD1201 is checking for Events domain and SD1352 is for Interventions domains. These rules, indicated by P21, aim to identify multiple observations, events and interventions collected on the same subject at the exact same time. Duplicate records can lead to unexpected results during the analysis and even confuse data review.

For this issue, sponsors might provide a simple explanation that the records are not exactly duplicated and that this issue is caused by P21 bugs. However, this explanation does nothing to clarify what triggers this issue. There are two common situations that can cause this issue. The first case is that the records do duplicate due to special reasons, such as the same test being taken twice on the same day. The second is that besides the keys used in the P21 check, other variables also contribute as the key variables of the dataset, which are specified in the define.xml file. Therefore, sponsors should provide different interpretations depending on the actual situation of the dataset.

Dataset	Duplicate Reason	Explanation
PE	Records do duplicate and the key variables listed in define.xml cannot differentiate the records	For ABC-01-001, the PE tests were taken twice on Day 29, while only the “If abnormal, please describe” answer for “Skin” were different.
MI	The records are not duplicated because MITSTDTL has different values.	Records are unique among the key variables listed in the define.xml, which takes MITSTDTL into account.

Table 7. Examples of Complete Explanations for SD1117

CONCLUSION

Pinnacle 21 Validation tool provides excellent and efficient conformance checks for the submission data. For best practice, it is necessary to understand the check rules that are triggered in your study and the resolutions to these reported issues in different situations. This paper may give different resolutions or perspectives on validation issues and help you improve the quality of your submission packages to support regulatory reviews.

REFERENCES

[1] U.S. Food & Drug Administration, *Study Data Technical Conformance Guide* (current version at time of access – v5.3/ May 2023), Accessed May 2023 - https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document

[2] U.S. Food & Drug Administration, *FDA Validator Rules* (current version at time of access – v1.6/December 2022), Accessed May 2023 - https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources

[3] Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 3.4. November 2021

[4] Pinnacle 21 Community Forum. Available at https://www.pinnacle21.com/forum

[5] Kelly, Kristin. 2018. 'Best Practice for Explaining Validation Results in the Study Data Reviewer's Guide'. PharmaSUG 2018, Phuse US Connect 2018, and Phuse EU 2018. https://www.pharmasug.org/proceedings/2018/SS/PharmaSUG-2018-SS13.pdf

[6] Ajay, Gupta. 2019. 'Common Pinnacle 21 Report Issues: Shall we Document or Fix?'. PharmaSUG 2019. https://www.lexjansen.com/pharmasug/2019/DS/PharmaSUG-2019-DS-119.pdf

[7] WHODrug B3- and C3- Formats – Implementation Guide, How to use WHODrug for compliance with CM domain in the CDISC SDTM standard - https://who-umc.org/media/164798/whodrug-b3-and-c3-formats-implementation-guide-april-2018.pdf

CONTACT INFORMATION

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

Kaiying Li
Jiangsu Hengrui Pharmaceuticals Co.,Ltd.
+86 18819454885
Kaiying.Li@hengrui.com