PharmaSUG China 2018 - Paper CD-24

Common Pinnacle 21 Report Issues: Shall we Document or Fix?

Ajay Gupta, PPD, Morrisville, NC

ABSTRACT

Pinnacle 21, also known as OpenCDISC Validator, provides great compliance checks against CDISC outputs like SDTM, ADaM, SEND and Define.xml. This validation tool provides a report in Excel or CSV format which contains information as errors, warnings, and notices. At the initial stage of clinical programming when the data is not very clean, this report can sometimes be very large and tedious to review. If the programmer is fairly new to this report s/he might not be aware of some common issues and will have to fully depend on an experienced programmer to pave the road for them. Indirectly, this will add more review time in the budget and might distract the programmer from real issues which affect the data quality. In this paper, I will discuss some common issues with the Pinnacle 21 report messages created from running against SDTM datasets and propose some solutions based on my experience. Also, I will discuss some scenarios when it is better to document the issue in reviewer's guide than doing workaround programming. While the author totally agrees that there is no one fit for all solution, my intention is to provide programmers a direction which might help them to find the right solutions for their situation.

INTRODUCTION

In 2004, the Clinical Data Interchange Standards Consortium (CDISC) recommended the use of the Study Data Tabulation Model (SDTM) standard for submitting clinical data to a regulatory agency. Since that time the pharmaceutical and biotechnology industries have worked tirelessly to implement this standard for the submission of clinical data and its related metadata in order to facilitate the review process of determining the safety and efficacy of a drug. The Pinnacle 21 Community was established to build a framework for the implementation of the CDISC Standard. In fact, this community of professionals created the Pinnacle21 Validator tool that performs numerous checks on clinical data to ensure compliance with the standard. This tool validates both collected data as well as its respective metadata.

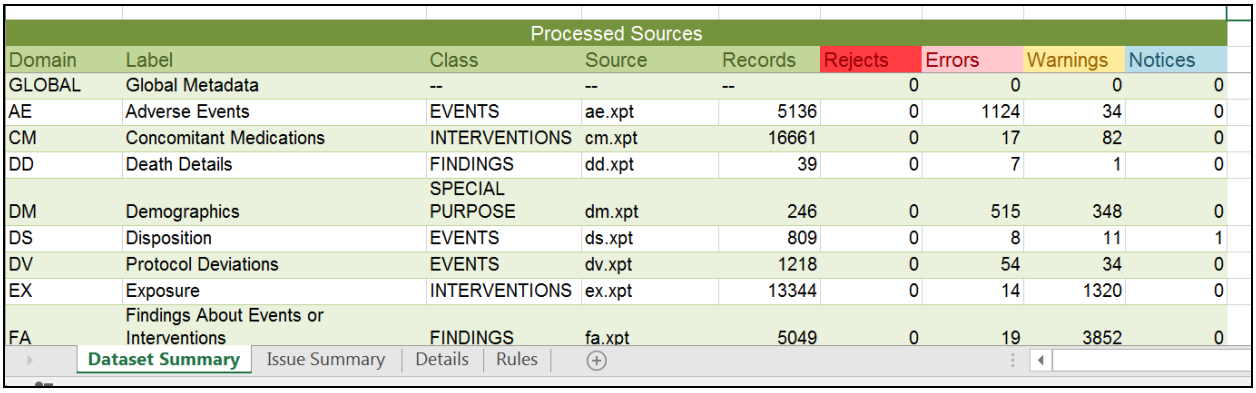
This paper will cover Pinnacle 21 Community validator tool which is widely use in the industry and user can download it at free of cost. However, using this application poses a real challenge with respect to understanding its concise error and warning messages. Some of the messages do not provide a lot of insight on how to resolve data issues. It's nice to know that a particular error occurred numerous times; however, it is more important to understand the error, where and why it happened. Is the problem systemic? For example, in the Pinnacle 21 report, the Subject Visits (SV) domain sometimes seems to have lot of issues, but only because it contains non-randomized subjects who don't belong there. Least of all, a lot of time is spent trying to decipher messages, some of which seem extraneous. In order to use Pinnacle 21 efficiently, it is necessary to realize the multi-disciplined nature of the validation process, which goes beyond the application, specifically: CDISC, SAS, and clinical data. Also, the user should understand how Pinnacle 21 functions with respect to validation checks and data/metadata issues. The report only points to the observation number of the data set that resulted in the issue. This leads to increased efforts required by the user to investigate the issue as the user needs to explore the data outside the report using secondary tools.

Admittedly, CDISC requirements for standardization are extensive, and always evolving. The validation process involves various types of checks to ensure compliance, including the metadata describing the clinical data. In fact, the clinical data, stored as SAS transport (XPT) data sets, must match that which is specified in the Case Report Tabulation – Data Definition (Define-XML) document. Pinnacle 21 uses both transport data sets and the Define-XML to perform the validation. Besides metadata checks, the application also checks for appropriate controlled terminology values (e.g. F, M, or U for the variable DM.SEX) and standard formats, such as using the ISO 8601 format for date/time values. Pinnacle 21 does not guarantee 100% compliance. However, it does a good job of detecting most data issues that would otherwise delay a submission. This application's important asset to conformance of the clinical trials data with the submission standards which is why I have decided to discuss some of the issues

surrounding interpreting the report and how to investigate and solve them in a timely fashion to produce quality submission data.

PINNACLE 21 COMMUNITY VALIDATOR TOOL:

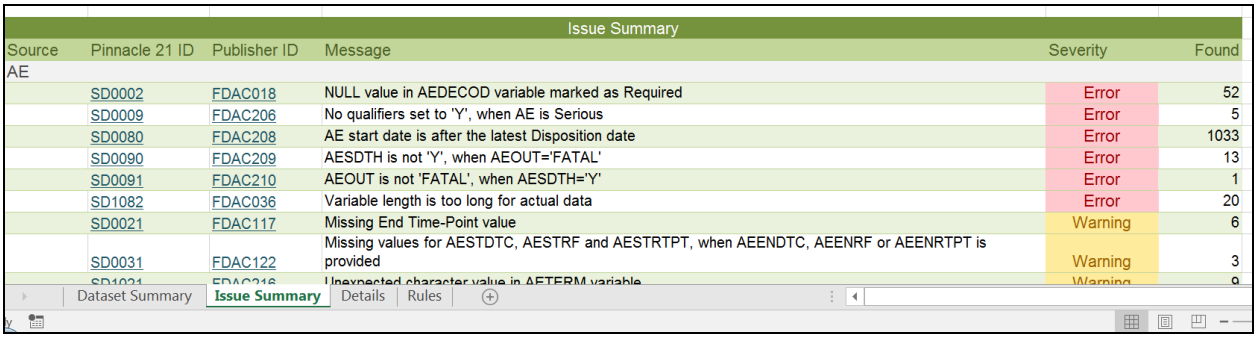
Pinnacle 21 Community Validator is the leading industry tool for validating SDTM data sets against CDISC standards (for more information about CDISC standards, please see cdisc.org). After the validator has finished checking the SDTM data sets, findings are made available to the user, typically in Excel format. The findings report consists of four tabs: Datasets Summary, Issue Summary, Details, and Rules. The Datasets Summary tab provides an overview of the contents for each input file and contains summary information about the total number of records, errors, warnings, and notices for each domain. The Issue Summary tab breaks down issues by severity (error, warning, and notice) and by type for each domain. Each issue type is categorized by FDA Publisher ID, which represents the FDA's published business rules. A description of each rule can be found on the Rules tab. The Details tab includes all issues in an expanded format and is presented on the record level. This tab includes the domain, record number, count, variables, values, rule ID, message, category, and severity for each issue.



Processed Sources

Domain	Label	Class	Source	Records	Rejects	Errors	Warnings	Notices
GLOBAL	Global Metadata	--	--	--	0	0	0	0
AE	Adverse Events	EVENTS	ae.xpt	5136	0	1124	34	0
CM	Concomitant Medications	INTERVENTIONS	cm.xpt	16661	0	17	82	0
DD	Death Details	FINDINGS	dd.xpt	39	0	7	1	0
DM	Demographics	SPECIAL PURPOSE	dm.xpt	246	0	515	348	0
DS	Disposition	EVENTS	ds.xpt	809	0	8	11	1
DV	Protocol Deviations	EVENTS	dv.xpt	1218	0	54	34	0
EX	Exposure	INTERVENTIONS	ex.xpt	13344	0	14	1320	0
FA	Findings About Events or Interventions	FINDINGS	fa.xpt	5049	0	19	3852	0

Display 1. Dataset Summary Tab View



Issue Summary

Source	Pinnacle 21 ID	Publisher ID	Message	Severity	Found
AE					
	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Error	52
	SD0009	FDAC206	No qualifiers set to 'Y', when AE is Serious	Error	5
	SD0080	FDAC208	AE start date is after the latest Disposition date	Error	1033
	SD0090	FDAC209	AESDTH is not 'Y', when AEOUT='FATAL'	Error	13
	SD0091	FDAC210	AEOUT is not 'FATAL', when AESDTH='Y'	Error	1
	SD1082	FDAC036	Variable length is too long for actual data	Error	20
	SD0021	FDAC117	Missing End Time-Point value	Warning	6
	SD0031	FDAC122	Missing values for AESTDTC, AESTRF and AESTRTPT, when AEENDTC, AEENRF or AEENRTPT is provided	Warning	3
	SD1021	FDAC216	Unexpected character value in AETERM variable	Warning	9

Display 2. Issue Summary Tab View

Domain	Record	Count	Variables	Values	Pinnacle 21 ID	Publisher ID	Message	Category	Severity
AE	8		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	208		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	292		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	303		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	809		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1021		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1028		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1032		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1262		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1459		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1466		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1467		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1475		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1478		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1489		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1492		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1497		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1498		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1501		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1502		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1509		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error
AE	1547		AEDECOD	null	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Presence	Error

Dataset Summary | Issue Summary | Details | Rules

Display 3. DetailsTab View

Message	Description	Category	Severity
Variable value not found in non-extensible codelist	Variable must be populated with terms from its CDISC controlled terminology codelist. New terms cannot be added into non-extensible codelists.	Terminology	Error
Variable value not found in extensible codelist	Variable should be populated with terms from its CDISC controlled terminology codelist. New terms can be added as long as they are not duplicates, synonyms or subsets of existing standard terms.	Terminology	Warning
Coded and Decoded values do not have the same Code in CDISC CT	Paired variables such as TEST/TESTCD must be populated using terms with the same Codelist Code value in CDISC control terminology. There is one-to-one relationship between paired variable values defined in CDISC control terminology by Codelist Code value.	Terminology	Error
Variable value not found in non-extensible codelist when value-level condition occurs	Variable must be populated with terms from its CDISC controlled terminology codelist, when its value level condition is met. New terms cannot be added into non-extensible codelists.	Terminology	Error
Variable value not found in extensible codelist when value-level condition occurs	Variable should be populated with terms from its CDISC controlled terminology codelist, when its value level condition is met. New terms can be added as long as they are not duplicates, synonyms or subsets of existing standard terms.	Terminology	Warning
Coded and Decoded values do not have the same Code in CDISC CT when value-level condition occurs	Paired variables such as TEST/TESTCD must be populated using terms with the same Codelist Code value in CDISC control terminology. There is one-to-one relationship between paired variable values defined in CDISC control terminology by Codelist Code value within the same value level condition.	Terminology	Error
No records in data source	Domain table should have at least one record.	Presence	Error
NULL value in variable marked as Required	Required variables (where Core attribute is 'Req') cannot be NULL for any records.	Presence	Error
Invalid ISO 8601 value for variable	Value of Dates/Time variables (*DTC) must conform to the ISO 8601 international standard.	Format	Error
Inconsistent value for DOMAIN	Domain Abbreviation (DOMAIN) variable should be consistent with the name of the dataset.	Consistency	Error
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) , Pool Identifier (POOLID) or Sponsor Device Identifier (SPDEVID) variables value when they are present in the domain.	Consistency	Error

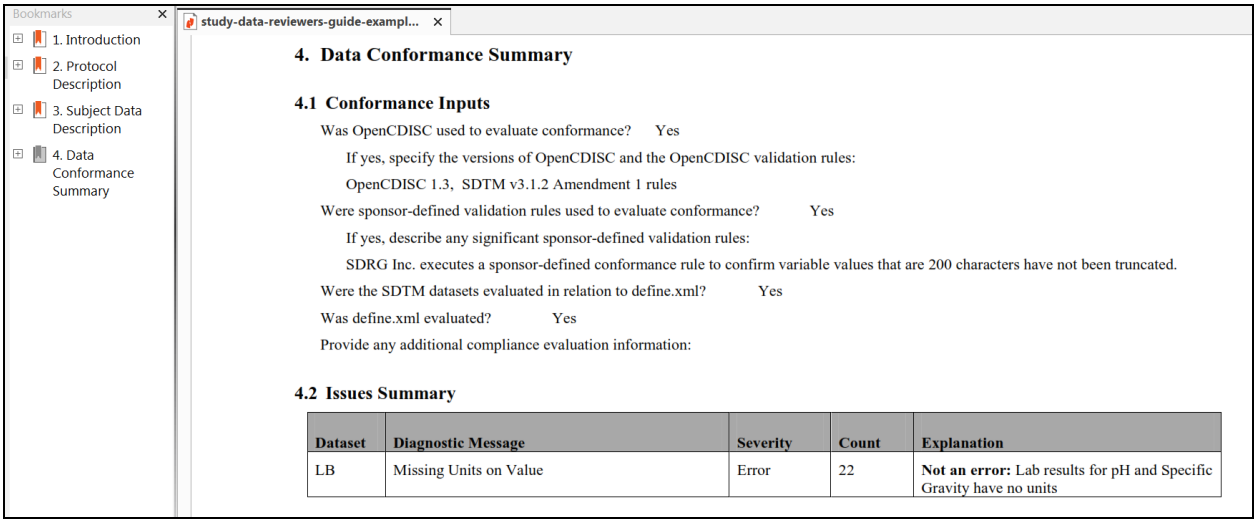
Dataset Summary | Issue Summary | Details | Rules

Display 4. Rules Tab View

REVIWERS GUIDES

Reviewer's Guides are relatively new type of study metadata developed by Association Programming Pharmaceutical Users Software Exchange (PhUSE). Study Data Reviewer's Guide (SDRG) was introduced in 2013 to provide FDA reviewers with a high-level summary and additional context for the submission data package. It purposefully duplicates information found in other submission documentation (protocol, clinical study report, annotated CRFs, define.xml, etc.) in order to provide FDA reviewers with a single point of orientation to the submission data. Reviewer's Guide communicates additional information about mapping decisions, sponsor-defined domains, and sponsor extensions to CDISC controlled terminology. It also captures sponsor's explanations of data validation issues, specifically the reason why those issues were not addressed during study conduct, mapping, and submission preparation. There is a rapid adoption of Study Data Reviewer's Guide by the industry, primarily due to its popularity with FDA reviewers, but also for its usability. On average, a Reviewer's Guide has only about 30 pages, which is a lot less than hundreds of pages across protocol, define.xml, and other documents.

The Data Conformance Summary section of the Reviewer's Guide provides an opportunity for sponsors to identify and explain in detail why some of the data issues were not fixed. This helps reviewers navigate around the data issues during analysis and preempts the need for additional question and clarifications.



Display 5. Data Conformance Summary section in SDRG

GENERAL APPROACH TO REVIEW PINNACLE 21 REPORTS:

The general approach to review Pinnacle 21 reports are as follows.

1. Review errors and warnings carefully. Understanding P21C findings can be difficult and it is the responsibility of the programmer to discern which issues can or cannot be resolved. It is common to have issues come from multiple sources such as dirty data, incorrect mapping, and programming errors. Additionally, some issues are present simply because the study is ongoing (this is especially true with issues related to the Disposition domain). Before the programmer can communicate issues to Data Management, the programmer must discern which issues are suspected to be data related.
2. Notify data management (DM) about any data related and try to fix any programming errors, mapping, and CT related issues. But, strictly avoid any workaround programming to get rid of errors and warnings e.g. no date imputation on SDTM level to avoid any warning related to date. General perspective, Errors always have High severity; whereas Warnings have either Low or Medium severity. Errors must be corrected; however, Warnings should be corrected in order to assist with the submission, even though some warning and errors may be acceptable depending on the study.
3. Give special attention to notices it might be related to configuration issue.
4. If the issue is irresolvable then provide a proper justification in reviewer's guide. Issue explanations should be detailed and study specific. Unfortunately, there are many cases when provided explanations are generic and invalid.

ISSUES AND RESOLUTIONS

Now I would like to discuss the various messages and how to determine a resolution. This will not be all inconclusive but hopefully will have most of issues that you may come across when reviewing your Pinnacle 21 reports specially for SDTM.

Message / Issue	Proposed Solution
XXX value not found in 'XXX' non-extensible codelist	Please check the respective code list in "SDTM Terminology.xlsx" located on cancer.gov. Later, try to replace the value with most suitable match from code list. For e.g. if the value for variable race is "BLACK/AFRICAN AMERICAN" then replace it with "BLACK OR AFRICAN AMERICAN" as per the code list race.
XXX value not found in 'Frequency' extensible codelist	Please check the respective code list in SDTM Terminology.xlsx located on cancer.gov. Later, try to replace the value with most suitable match from respective code list. Since, this code list is extensible which indicates that if suitable match is not found in the code list then leave the value as is and document it in reviewer's guide. For e.g. frequency variable with value "UNK" can be replace by "UNKNOWN".
NULL value in XXX variable marked as Required	This warning indicates that missing values for any records is not permitted for respective variable. For e.g. AEDECOD in AE domain is a required variable. Please follow the steps below to resolve the issue: 1. Identify the row with missing AEDECOD in AE dataset. data temp; set AE; where missing(AEDECOD); run; 2. Notify Data management with necessary information. 3. There is a possibility that data is partial and coding will be provided at later stage.
Permissible variable with missing value for all records	As per the SDTM IG V3.2, permissible variables are added only when there is a data collected for respective variable. In other words, it is ok to drop the permissible variable with missing value for all records. If the data is 100% then permissible variable with missing value for all records can be drop from the dataset.
Invalid ISO 8601 value for variable	Date and time should be populated in ISO8601 format. e.g.: YYYY-MM-DD (Date) or YYYY-MM-DDTHH:MM:SS (Date and time). Check SDTM IG for more details on ISO8601 format. Also, please reach to data management for any data issue. If the data issue is not resolved in final transfer then justify the issue in reviewer's guide.
No baseline result in Domain for subject	Please follow the steps below to resolve the issue: 1) Check if DM.RFSTDTC is missing for the subjects that have missing baseline flag. 2) Discuss with Biostats to make sure the baseline

	derivation algorithm is correct, if not fix it. 3) If DM.RFSTDTC date is missing due to data not available in the raw dataset or data issue, inform DM to query it. 4) If there are missing baseline flag records after DBL lock, ask DM to provide justification and save it in the reviewers' guide.
Inconsistent value for Standard Units	Standard unit should be consistent for Test name, category, specimen, and method. Please follow the steps below to resolve the issue: 1) Check the source data, if standard unit is directly collected in the source, and if that is inconsistent, inform to DM or Vendor to fix it. Doing a frequency table will be helpful here. 2) If standard units are derived in program then check the logic to derive standard unit for each Test name, category, specimen, and method. If standard values are derived using any factor sheet, make sure the source file is correct. 3) This issue has to be resolved, consult with DM and Sponsor to find out a solution.
No qualifiers set to 'Y', when AE is Serious	Check Adverse Event page, when serious event is collected as Yes then any one of the Involves Cancer, Congenital Anomaly or Birth Defect , Persist or Signify Disability/Incapacity , Results in Death , Requires or Prolongs Hospitalization , Is Life Threatening , or Other Medically Important Serious Event data must be collected on same page. If information not collected on adverse event page for any of these variables then check the associated raw datasets for the subject (E.g.: Serious event is Yes and death page is entered for that subject then Result in Death will be mapped to yes). If you notice data issue, inform DM to query it. If issue doesn't resolve after database lock, ask DM to provide justification to document it in the reviewer's guide.
--STDTC is after --ENDTC	Check Start and End dates in the raw data for the affected subject(s), if start date is after end date then, please follow the steps below to resolve the issue: 1) Check the algorithm in your DM program. 2) Inform DM to query it, if it's a data issue.
Invalid value for –TEST or -TESTCD variable	Please follow the steps below to resolve the issue: 1) Test name and Test code should be only 40 and 8 characters respectively in all finding datasets expect IE or TI were IETEST can be 200 characters. 2) If test name is more than 40 characters then set it to 40

	characters with appropriate meaning. If you're not sure, ask study statistician to provide the cut-short text for SDTM programming. 3) Align test code and name with SDTM controlled terminology and make sure they have one-to-one mapping.
Missing End Time-Point value	Check --ENDTC mapping in specification. If --ENDTC is derived from multiple source dates then check all date variables in source data. If they are missing, inform DM to query it. If --ENDTC is missing and --OCCUR is not collected then --ENRF should be populated. If --ENDTC is missing and --OCCUR is collected Yes then --ENRF should be populated. Please consider using the newer relative timing variables introduced in SDTM IG v3.1.2: --STRTPT, --STTPT, --ENRTPT, --ENTPT. These variables can be used in the exact same manner as --STRF and --ENRF, but the big plus is they can provide a lot more precision if needed.
Missing value for --STAT, when --REASND is provided	Please follow the steps below to resolve the issue: 1) Check source data for completion status (--STAT), reason not done (--REASND) and result(--ORRES). 2) If reason not done (--REASND) and result (--ORRES) are collected then inform to DM team to query it. 3) If reason not done (--REASND) has collected and completion status (--STAT) is not collected and result (--ORRES) not collected then inform to DM team, and populate --STAT with "NOT DONE". 4) If reason not done (--REASND) and result (--ORRES) are not collected then inform to DM team, and populate --STAT with "NOT DONE".
Missing value for --ORRESU, when --ORRES is provided	Check source data for collected original results unit. If result is not missing and unit is missing then inform to DM to query it.
Value for variable not found in user-defined codelist	When Define.xml is also utilized in SDTM datasets validation, this check validates variables custom codelist provided in define.xml vs actual data. If any values that were populated in data but not present in custom codelist then this check will have populated in the Validation report. In order to resolve, make sure to present values in custom codelist that are expected in actual data.
Inconsistent value for --TEST within --TESTCD	Name of Measurement, Test or Examination (--TEST) and Short Name of Measurement, Test or Examination (--TESTCD) are one to one matching. Check the derivation of --TEST and --TESTCD with controlled terminology and update accordingly.
Inconsistent value for QLABEL within QNAM	Check SDTM mapping specs for SUPPQUAL Out variable and Out labels. Please follow the steps below to resolve the issue:

	1) Make sure they are one to one mapping. 2) Make sure the maximum characters for QNAM and QLABEL are 8 and 40 characters respectively.
Inconsistent value for VISIT within VISITNUM	Please follow the steps below to resolve the issue: 1) VISIT and VISITNUM should be one to one matching. 2) Check derivation of visit and visit number for unplanned visits and compare with trial visits (TV) dataset for mapping of scheduled visits. 3) If the error is associated with the Unscheduled visits, check the corresponding values in SV domain and fix the logic, so that it reflects the true scenario for any given patient and make sure it's consistent across all domains where visit variables are present.
SDTM Required or Expected variable not found	Check SDTM specification on why the required or expected variable is not kept in the specs, and add it to the specs to regenerate datasets.
Variable appears in dataset, but is not in SDTM model	If the non-SDTM variable were kept by mistake, then drop it, if not move it to SUPP-- dataset.
SDTM/dataset variable label mismatch	Compare SDTM specification against dataset and SDTM IG version used to make sure variable label is same for the given variable. If it is same, document it in the reviewer's guide.
Subject is not present in DM domain	Please follow the steps below to resolve the issue: 1) Check source for Demographic (DM) and Screen Failure datasets. If the subject present in other datasets but not in DM datasets then inform to Data Management to query it. Do not remove records through programmatically until DM take care of it. If the issue exists after DBL, then ask DM to provide documentation for reviewer's guide.
USUBJID/VISIT/VISITNUM values do not match SV domain data	Please follow the steps below to resolve the issue: 1) The SV domain should be derived using VISIT panel (if available) and all the planned visit domains. 2) Compare subject and visit combination in other datasets with Subject visit (SV) dataset. If subject and visit is not present then add it to subject visit (SV) dataset. 3) If these errors are associated with the Unscheduled visits, please check the corresponding values in SV domain and fix the logic, so that it reflects the true scenario for any given patient and make sure it's consistent across all domains where visit variables are present. 4) If data not entered, then inform DM to take care of it.
No Disposition record found for subject	Disposition should have at least one record for subject present in DM dataset. If any subject not present in Disposition (DS) dataset which is present in Demographic (DM) then check the derivation of records in disposition (DS) dataset, if subject not present in source dataset to derive record in disposition (DS) then inform to DM to query

	it. If the missing subject belong to SCRNFAIL or Treatment not assigned, it's acceptable.
AE start date is after the latest Disposition date	Check the respective Subject's AE and DS records, and make sure the logic in DS domain is correct. In most cases, data may be not entered in to the EDC yet, inform Data Management to take care of it, and watch it after next data extraction. If the issue still exists after database lock, then ask Data Management to provide justifications to put it in the reviewer's guide.
Exposure end date is after the latest Disposition date	Check the respective Subject's EX and DS records, and make sure the logic in DS domain is correct. In most cases, data may be not entered in to the EDC yet, inform Data Management to take care of it, and watch it after next data extraction. If the issue still exists after database lock, then ask Data Management to provide justifications to put it in the reviewer's guide.
RFSTDTC is after RFENDTC	Check first exposure date (EXSTDTC) with last disposition date (DSSTDTC). If EXSTDTC is after DSSTDTC and it is not a programming issue then inform to Data Management.
Invalid ETCD/ELEMENT	Compare Subject Element (SE) dataset with Trial Element (TE) dataset, ELEMENT and ELEMENT CODE (ETCD) always match with TE dataset. ELEMENT and ELEMENT CODE (ETCD) should always one to one matching, if both are different then check derivation of each element.
Invalid EPOCH	Compare EPOCH in all dataset with TA dataset. All planned Visit EPOCH should match with EPOCH in TA dataset.
Unexpected character value in variable	Remove leading and trailing spaces from the character values.
Redundancy in paired variables values	Redundancy values are not expected in SDTM. If any paired values are same as other then set one to missing.
Model permissible variable added into standard domain	Contact Sponsor whether to keep the permissible variables in parent domain or map it to SUPP domain.
FDA Expected variable not found	According to FDA expectations, EPOCH should be added into SDTM domains
No Treatment Emergent info for Adverse Event	Create a record called QNAM = "AETRTEM" and QLABEL = "Treatment Emergent Flag" to populate TEAE flag. Obtain algorithm from the Stats. See SDTM IG v3.2 section 8.4.3 SUPP-- Examples for details.
Duplicate records	Check source data for result and date variables for subject and test name. Please follow the steps below to resolve the issue: 1) If result is collected more than once to test name on same date for a subject then consult with Data Management to make sure data is correct. Then, see if adding any other variable can make the records unique. e.g.: Time, otherwise document it in the reviewers guide. 2) If the duplicate is due to Programming issue, then check raw and SDTM records for the same subjects to find out root cause of the issue, to fix it. 3) If the duplicate is due to true data issues, inform to Data Management to query it.

	4) Also, make sure if the variable required in sort order, to make the duplicate records unique in the dataset is being mapped to SUPPXX, then have a plan to map such information to correct variable of main parent domain.

Table 1 Pinnacle 21 Report Issues and Resolutions Table

CONCLUSION

Pinnacle 21 is an excellent validation tool. However, using the application poses a real challenge with respect to understanding its concise error and warning messages. Therefore, it's critical that Programmers avoid common mapping and programming errors, which can reduce the overall quality of submissions. Programmers can follow the examples and recommendations in this paper to detect, understand, and fix common issues to avoid impacting regulatory review process.

REFERENCES

Gupta Ajay, 2016. Enhanced OpenCDISC Validator Report for Quick Quality Review Proceedings of the PharmaSUG 2016 Conference, paper AD07.

Redner Virginia and Gerlach John, 2011. Resolving OpenCDISC Error Messages Using SAS®. Proceedings of the PharmaSUG 2011 Conference, paper CD07.

Amy Garrett and Chris Whalen, 2016. The Devil is in the Details – Reporting from Pinnacle 21 (OpenCDISC) Validation Report Proceedings of the PharmaSUG 2016 Conference, paper AD09.

Sergiy Sirichenko, 2017. Common Programming Errors in CDISC Data Proceedings of the PharmaSUG 2017 Conference, paper DS15.

Sergiy Sirichenko and Max Kanevsky, 2016. What is high quality study metadata? Proceedings of the PhUSE 2016, Conference, paper RG01.

https://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide

https://www.phusewiki.org/wiki/index.php?title=Analysis_Data_Reviewer%27s_Guide

Pinnacle 21 Community. Available at www.pinnacle21.net/download

ACKNOWLEDGMENTS

Thanks to Ryan Wilkins, Lindsay Dean, Ken Borowiak, Richard DAmato, Lynn Clipstone, and PPD Management for their reviews and comments. Thanks to my family for their support.

CONTACT INFORMATION

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

Ajay Gupta, M.S.
PPD
3900 Paramount Parkway
Morrisville, NC 27560
Work Phone: (919)-456-6461
Fax: (919) 654-9990

E-mail: Ajay.Gupta@ppdi.com,
Ajaykailasgupta@aol.com

DISCLAIMER

The content of this paper are the works of the authors and do not necessarily represent the opinions, recommendations, or practices of PPD.