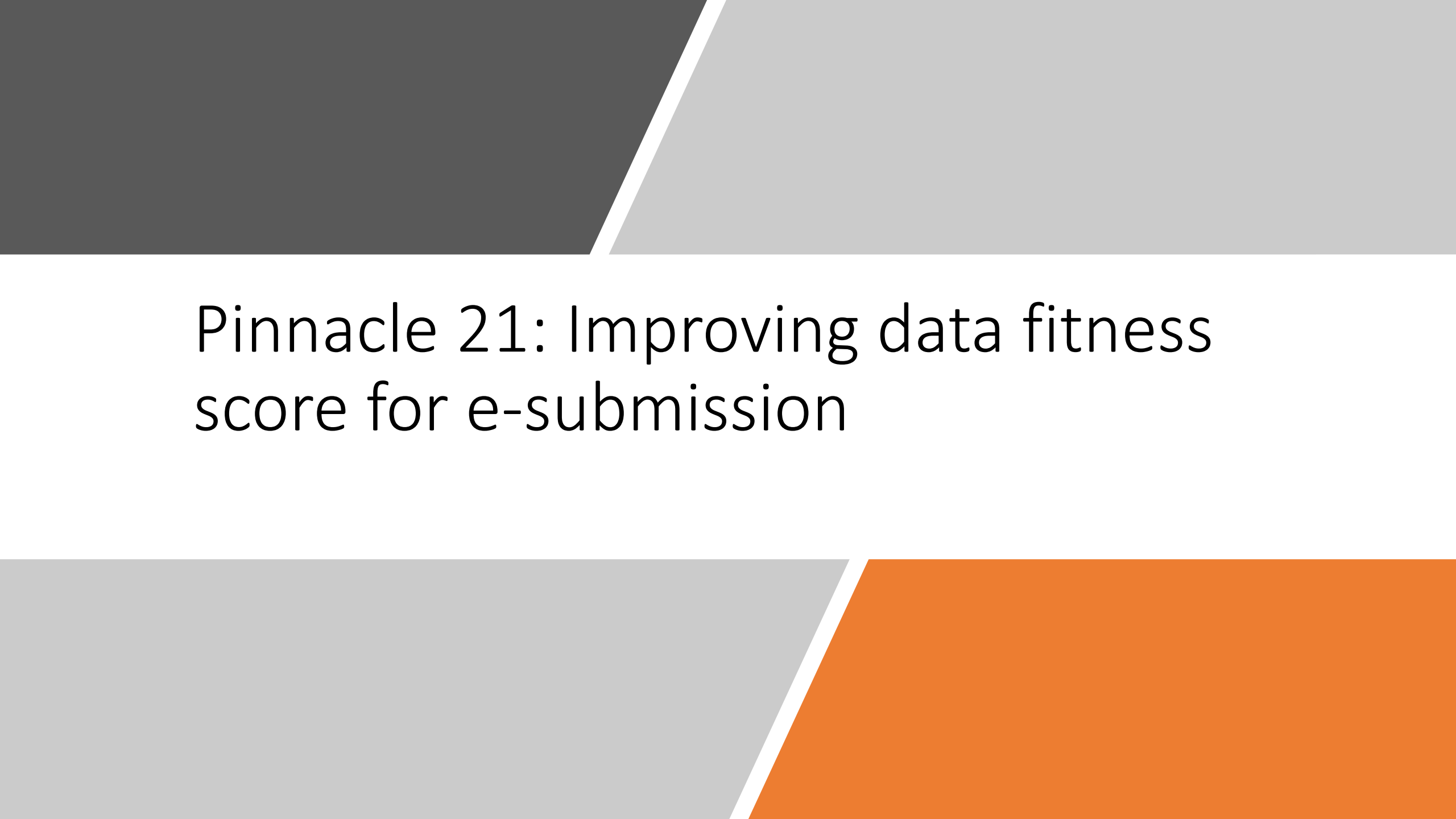




Pinnacle 21: Improving data fitness score for e-submission

Introduction

Today, Data Conformance is a standard and required part of regulatory submissions. Regulatory agencies expect sponsors to validate their study data before submission and either correct or explain discrepancies in the Reviewer's Guide.

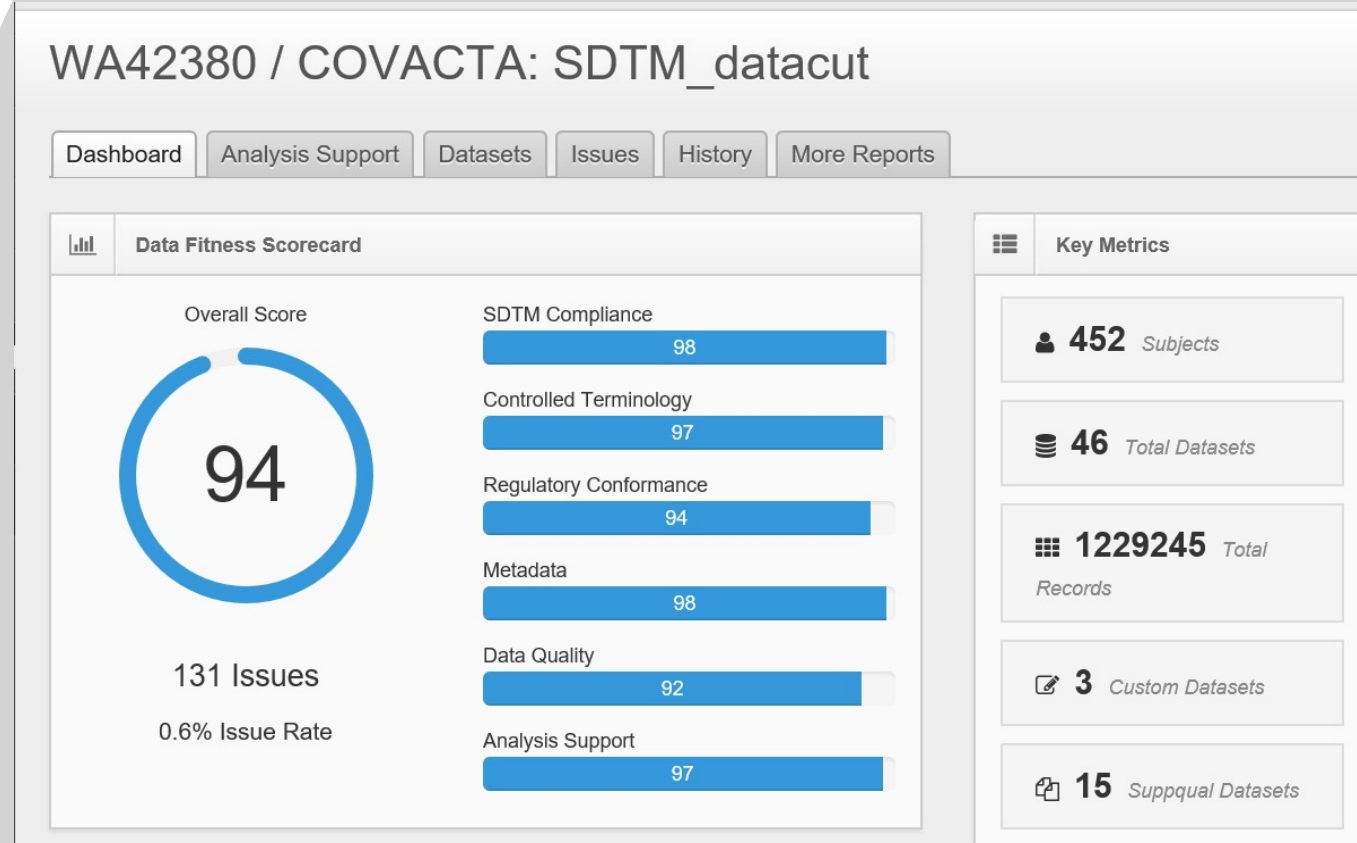
Pinnacle 21 Enterprise is designed to ease regulatory submission preparation, manage standards for datasets, and provide continuous compliance. The data compliance is assessed with respect to six quality dimensions in Pinnacle 21 Enterprise.

Pinnacle 21, dashboard gives an overview of where your data stands for a specific study. It gives details on issues and even standard rules to follow to make your data fit for submission. Fixing rejects/ errors/warnings helps a lot in increasing the data score. In addition, addressing the issues with define.xml, metadata and control terminology also helps in building the score. Last but not the least, providing explanations to issues not able to fix, and closing the issues within Pinnacle 21 does help a lot with data score. These issues explanations are submitted in Reviewer's guide with submission package.

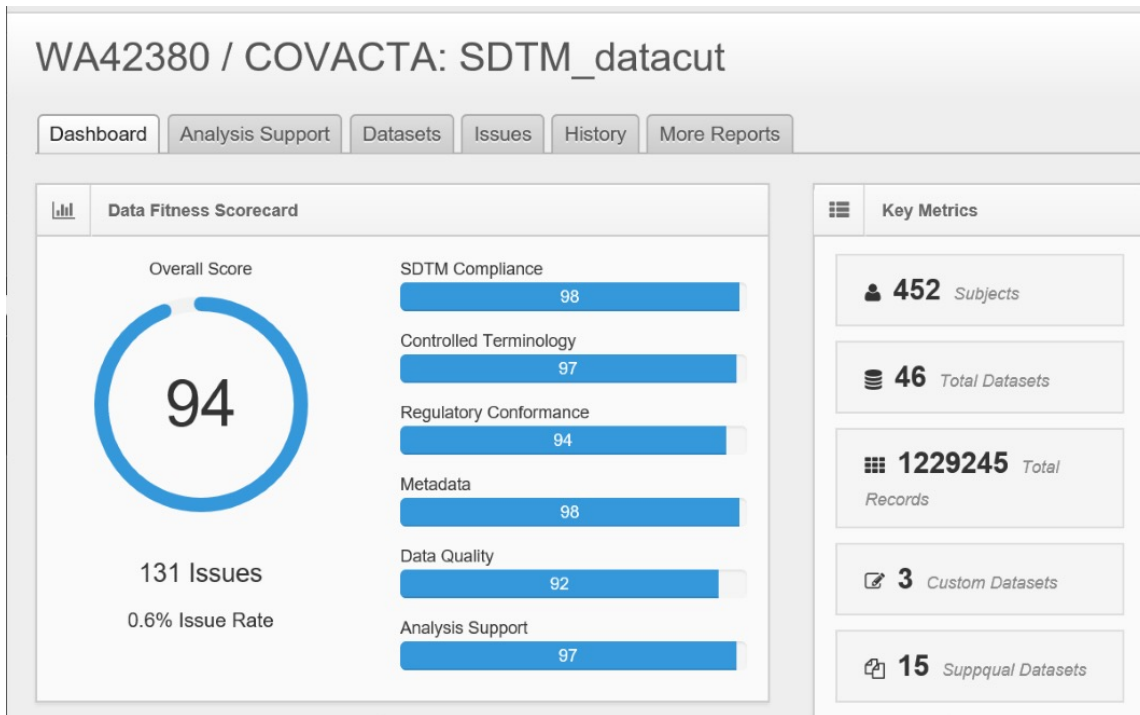
Six quality dimensions affecting data score

The fitness of data in terms of six quality dimensions includes

- SDTM compliance
- Data quality
- Control Terminology
- Metadata
- Regulatory conformance
- Analysis support



Pinnacle 21 enterprise issues tab with filters based on data fitness category



Dashboard Analysis Support Datasets Issues History More Reports

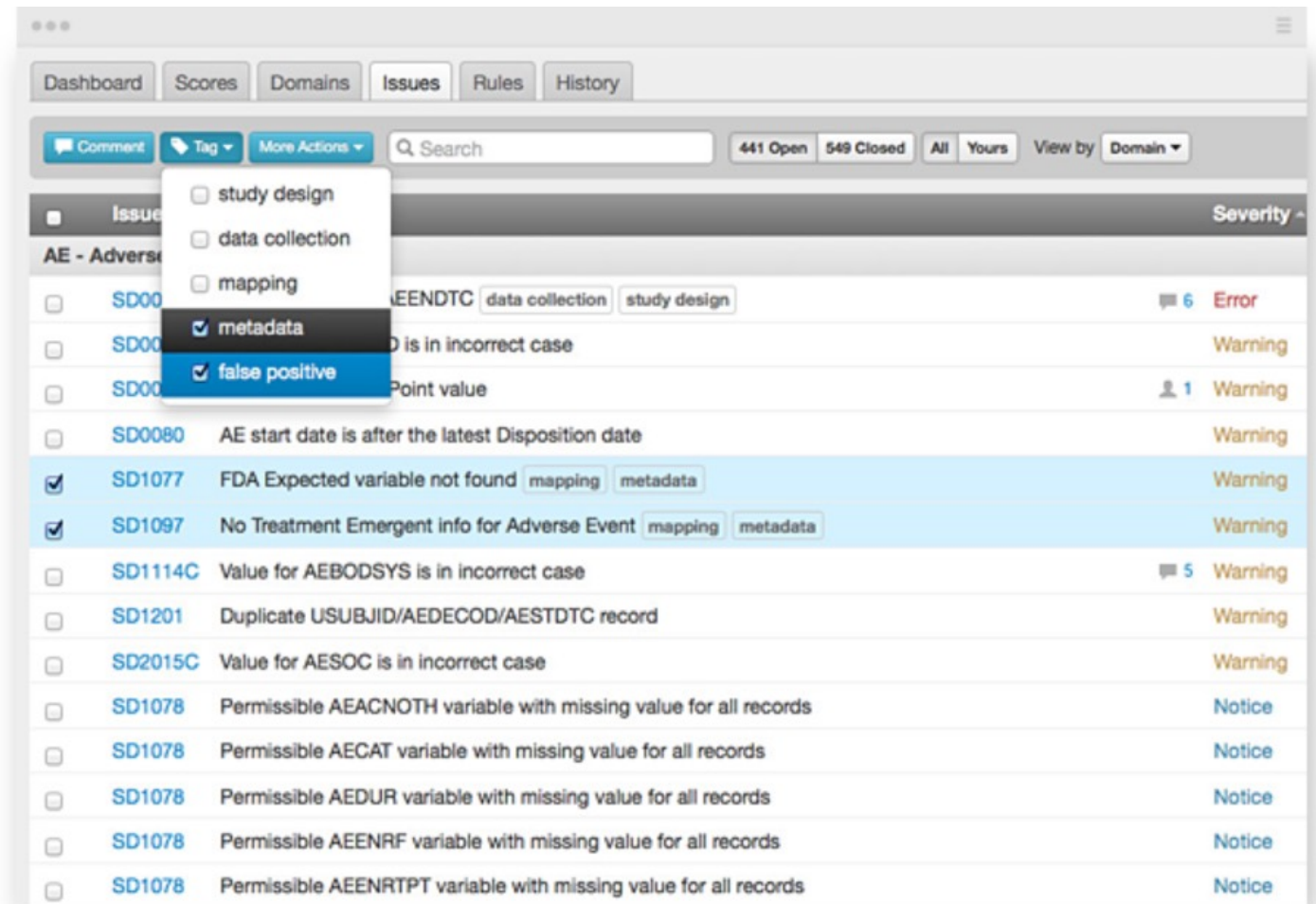
Actions Search Filters Status Assignee Tag Analysis Category View by Dataset

Rule	Message
AE - Adverse Events (1 Issues)	
☐ SD0037	Value for AEOUT not found in (AEOUT) user-defined codelist
DEFINE - Define.xml (38 Issues)	
☐ DD0001	Value '29.1' is not facet-valid with respect to pattern '-?[0-9]+' for type 'integer'.
☐ DD0001	Value '3.1' is not facet-valid with respect to pattern '-?[0-9]+' for type 'integer'.
☐ DD0001	Value '3.2' is not facet-valid with respect to pattern '-?[0-9]+' for type 'integer'.
☐ DD0001	Value '3.3' is not facet-valid with respect to pattern '-?[0-9]+' for type 'integer'.
☐ DD0001	Value '3.4' is not facet-valid with respect to pattern '-?[0-9]+' for type 'integer'.
☐ DD0001	Value '3.5' is not facet-valid with respect to pattern '-?[0-9]+' for type 'integer'.

- ☐ SDTM Compliance
- ☐ Controlled Terminology
- ☐ Regulatory Conformance
- ☒ Metadata
- ☐ Data Quality
- ☐ Analysis Support

RESTRICTED

Pinnacle 21 has an option of tagging various issues that can help in resolving or investigating or even provide an explanations in reviewers guide



#1 quality dimensions affecting data score: SDTM compliance

A very crucial part of the entire process is defining the SDTM mapping specification; it begins with annotations of the CRF according to the SDTM structures by analyzing the raw data. This process is time consuming, doing it first time right will save time & efforts.

Some typical SDTM mapping scenarios that have created data issues in Pinnacle 21

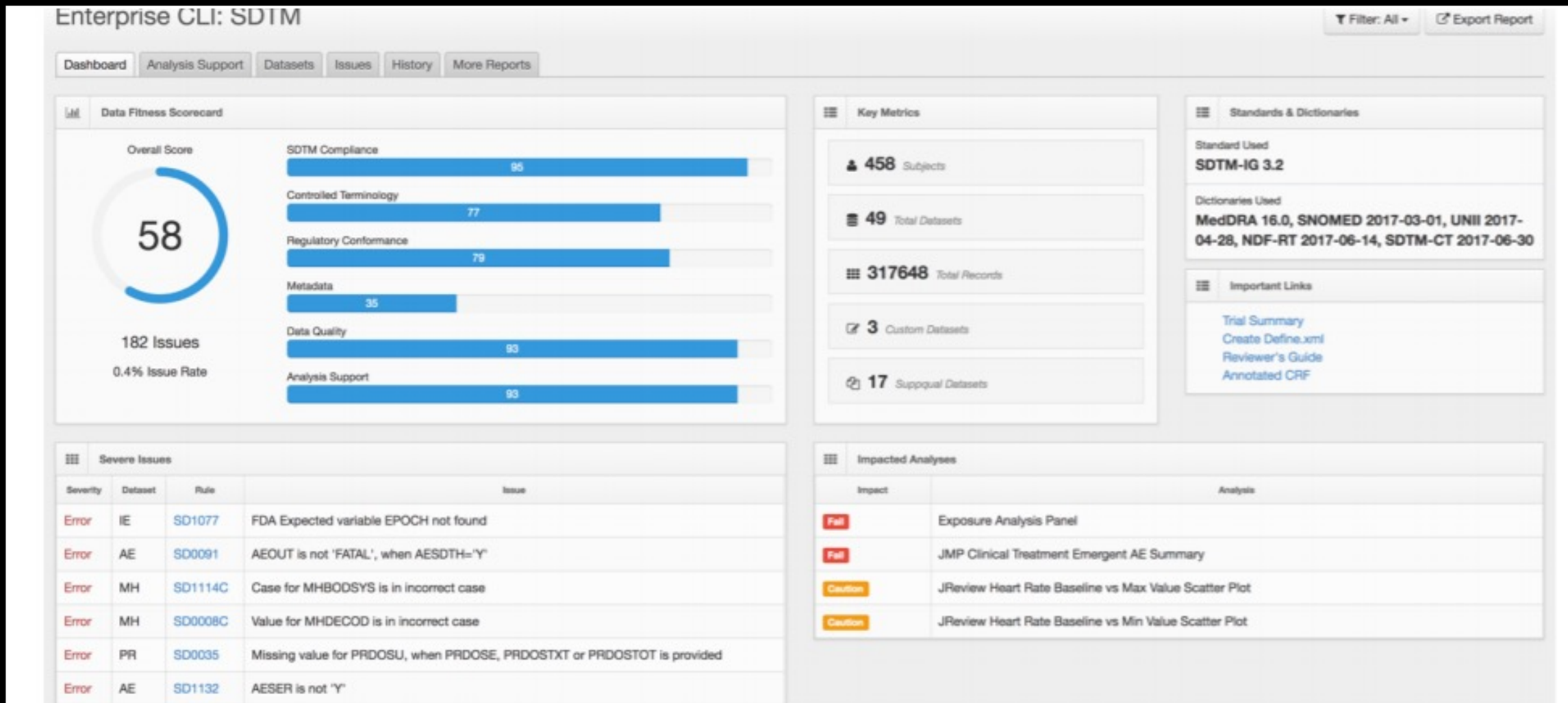
- You need to rename some variables to be able to map to the corresponding SDTM variable. For example, if the original variable is GENDER, it should be renamed SEX to comply with SDTM standards. Variable attributes must be mapped as well as variable names. Attributes like label, type, length and format must comply with the SDTM attributes.
- The format that a value is stored in is changed. However, the value itself does not change. For example, converting a SAS date to an ISO 8601 format character string.
- Some SDTM variables are obtained by deriving a result from data in the non-SDTM dataset. For example, instead of manually entering a patients age, using the date of birth and study start date to derive it instead
- Some variable values need to be recoded or mapped to match with the values of a corresponding SDTM variable. This mapping is recommended for variables with a code list attached that has control terminology that can't be extended. You should map all values in the controlled terminology, and not just the values present in the dataset.

Severe Issues			
Severity	Dataset	Rule	Issue
Reject	AE	SD0002	NULL value in AEDECOD variable marked as Required
Error	DM	TS0006	No Baseline (ALT) test results for Subject
Error	AE	TS0053	Neither AESEV or AETOXGR is populated
Error	AE	SD0013	AESTDTC is after AEENDTC
Error	DM	SD1343	Missing value for RFXSTDTC, when subject is treated
Error	AE	SD0080	AE start date is after the latest Disposition date
Error	AE	SD0009	No qualifiers set to 'Y', when AE is Serious
Error	AE	TS0012	Analysis Required variable AESEV not found
Error	DM	TS0008	No Baseline (AST) test results for Subject

Impacted Analyses	
Impact	Analysis
Fail	Disposition Analysis Panel
Fail	Exposure Analysis Panel
Fail	JMP Clinical Treatment Emergent AE Summary
Caution	JReview Heart Rate Baseline vs Max Value Scatter Plot
Caution	JReview Heart Rate Baseline vs Min Value Scatter Plot

#2 quality dimensions affecting data score: Data Quality

Once we receive the data, we have to make sure that required variables are populated and permissible values are included to make the study data compliant. There is definitely lot of communication initially with data managers to make sure we have data collected correctly. If further clarification or correctness needed with data, we need to get to communicate with sites to provide explanations or get the data corrected. It is very important that we analyze issues based on severity on Pinnacle 21 report: Reject, Error & Warning.



Pinnacle 21 dashboard: severe issues, impacted analysis and important links.

RESTRICTED

Pinnacle 21 Enterprise Validation Report

Issue Summary								
Rule ID	Publisher ID	Message	Severity	Found	Comments	Assigned	Tags	Explanation
SD0037	FDAB035	Value for AEOUT not found in (AEOUT) user-defined codelist	Error	2				
SD0091		AEOUT is not 'FATAL', when AESDTH='Y'	Error	3	FATAL is populated from the CRF - need to check with DM (Amy, 1/9/2019 4:19 PM)	Brad	data collection	
SD1015		Invalid EPOCH	Error	147	EPOCH is null for these record - need to check programming logic (Amy, 1/9/2019 4:15 PM)	Amy	mapping	
SD1132	FDAB010	AESER is not 'Y'	Error	1				
TS0053		Neither AESEV or AETOXGR is populated	Error	22				
CT2002		EPOCH value not found in 'Epoch' extensible codelist	Warning	2376		Amy		
SD0008C	FDAB017	Value for AEDECOD is in incorrect case	Warning	1924		Brad	data collection	
SD0021	FDAB038	Missing End Time-Point value	Warning	576				Outcome is "UNKNO therefore no end tim blank)
SD0031	FDAB043	AEENRF or AEENRTPT is provided	Warning	1			study design	

Enhanced Validation Report contains 4 new columns – Comments, Assigned, Tags, and Explanation

Pinnacle 21 Enterprise excel issue summary tab

Pinnacle 21 ID	Publisher ID	Message	Description	Category
SD1061	FDAC012	Missing MB dataset, when MS dataset is present	Microbiology Specimen (MB) dataset should be included, when a Microbiology Susceptibility Test (MS) dataset is present.	Presence
SD1062	FDAC211	AESER is not 'Y', when AESOD equals 'Y'	Serious Event (AESER) variable value is expected to be 'Y', when Occurred with Overdose (AESOD) variable value equals 'Y'.	Consistency
SD1063	FDAC023	Dataset is not present in define.xml	Datasets included in study data must be described in the data definition document (define.xml).	Metadata
SD1064	FDAC219	Duplicate ETCD value	The value of Element Code (ETCD) variable must be unique within Trial Elements (TE) domain.	Consistency
SD1065	FDAC160	Missing RDOMAIN value, when IDVAR is populated	Value of Related Domain Abbreviation (RDOMAIN) variable must be populated, when value of Identifying Variable (IDVAR) variable is populated.	Presence
SD1066	FDAC157	IDVARVAL value is populated, when RELTYPE values is populated	Variable Value (IDVARVAL) variable value should not be populated, when Relationship Type (RELTYPE) variable value is populated.	Presence
SD1067	FDAC166	USUBJID value is populated, when RELTYPE values is populated	Unique Subject Identifier (USUBJID) variable value should not be populated, when Relationship Type (RELTYPE) variable value is populated.	Presence
SD1068	FDAC218	Duplicate ELEMENT value	The value of Element (Description of Element) variable must be unique within Trial Elements (TE) domain.	Consistency
SD1069	FDAC087	Inconsistent value for --PARM within --PARMCD	All values of a Parameter (--PARM) variables should be the same for a given value of a Parameter Short Name (--PARMCD) variables.	Consistency
SD1070	FDAC088	Inconsistent value for --PARMCD within --PARM	All values of a Parameter Short Name (--PARMCD) variables should be the same for a given value of a Parameter (--PARM) variables.	Consistency
SD1071	FDAC016	Dataset is greater than 1 GB in size	Large datasets should be split into smaller datasets no larger than 1 GB in size.	Metadata
SD1072	FDAC204	Missing IDVAR value, when RDOMAIN value is provided	Value of Identifying Variable (IDVAR) variable must be populated, when Related Domain Abbreviation (RDOMAIN) variable value is provided, with the only exception of 'DM' value for RDOMAIN.	Consistency
SD1073	FDAC028	Variable prohibited for use in SDTM	Variables described in IG as inappropriate for usage must be not included in the dataset.	Metadata
SD1074	FDAC029	Variable which can be used only in SEND	Variables designed only for SEND pre-clinical studies must be not included in the SDTM dataset.	Metadata
SD1075	FDAC030	Variable not recommended for use	Variables described in IG as not recommended for usage should be not included in the dataset.	Metadata
SD1076	FDAC031	Model permissible variable added into standard domain	SDTM model variable may be added into standard domains according its domain general class, if there are no restrictions on their usage specified in IG.	Metadata

Pinnacle 21 enterprise excel with SDTM rules

Rejects are the critical issues that prevent review and automation processes. These should be fixed to make the data compliant.

Below are current rejection rules from the FDA .

Rule	P21	FDA eCTD
Trial Summary (TS) dataset must be included in every submission	SD1115	1734
Demographics (DM) dataset must be included in every submission	SD1020	1736
ADaM Subject level (ADSL) dataset must be included in every submission	AD0001	1736
FDA eCTD submissions must include a define.xml file for each study in Module 4 (nonclinical) and Module 5 (clinical)	DD0101	1736

Errors and warnings in Pinnacle 21

Errors: These are the issues reported, which needs to be fixed to make the data compliant. If not, then to be addressed in reviewers guide explaining the reason for not fixing.

Some of the examples:

- -Start Date is after End Date
- -Missing SDTM required variable

Warnings: These are the potential issues requiring manual review. Some examples:

- -Terms added to Extensible CT
- - Missing Units on Results
- -False-Negative and False-Positive messages

-False-Negative means that an issue exists but is not reported. For example, Permissible variables with missing values for all records, downgraded from a 'warning' to 'notice' in SDTM.

False-Positive means that an issue is reported, but it is not a true issue. An example of a False Positive record is a duplicate records.

#3 quality dimensions affecting data score: Control terminology

- Controlled Terminology validation rules identify discrepancies between the values a sponsor used in their data compared to allowable values of controlled terminology lists.
- Incorrect or missing codelists create errors/warnings which should be either resolved or provide valid explanations in reviewers guide. There are instances where sponsors populate codelists only for variables that have standard CDISC Control Terminology (AEACN), but do not create study specific codelists. For example, for Category (--CAT), Subcategory (--SCAT), or EPOCH variables.
- Highly recommended in creating codelists only for variables where data was collected, derived or assigned based on a list of pre-specified terms Include only terms and all terms which were used as options during data collection process
- Do not include full CT codelist. E.g., for EXDOSU ◦ Remember to include terms from CRF, which are not present in actual collected data.
- Ensure to populate study specific codelits! ◦ --CAT, --SCAT ◦ ARM, EPOCH, ELEMENT ◦ --GRID, --SPID if applicable

Control terminology cont...



Sometimes the issue is reported for valid standard terms and their NCI codes, the possible explanation is an incorrect version of CDISC Control Terminology used for validation.



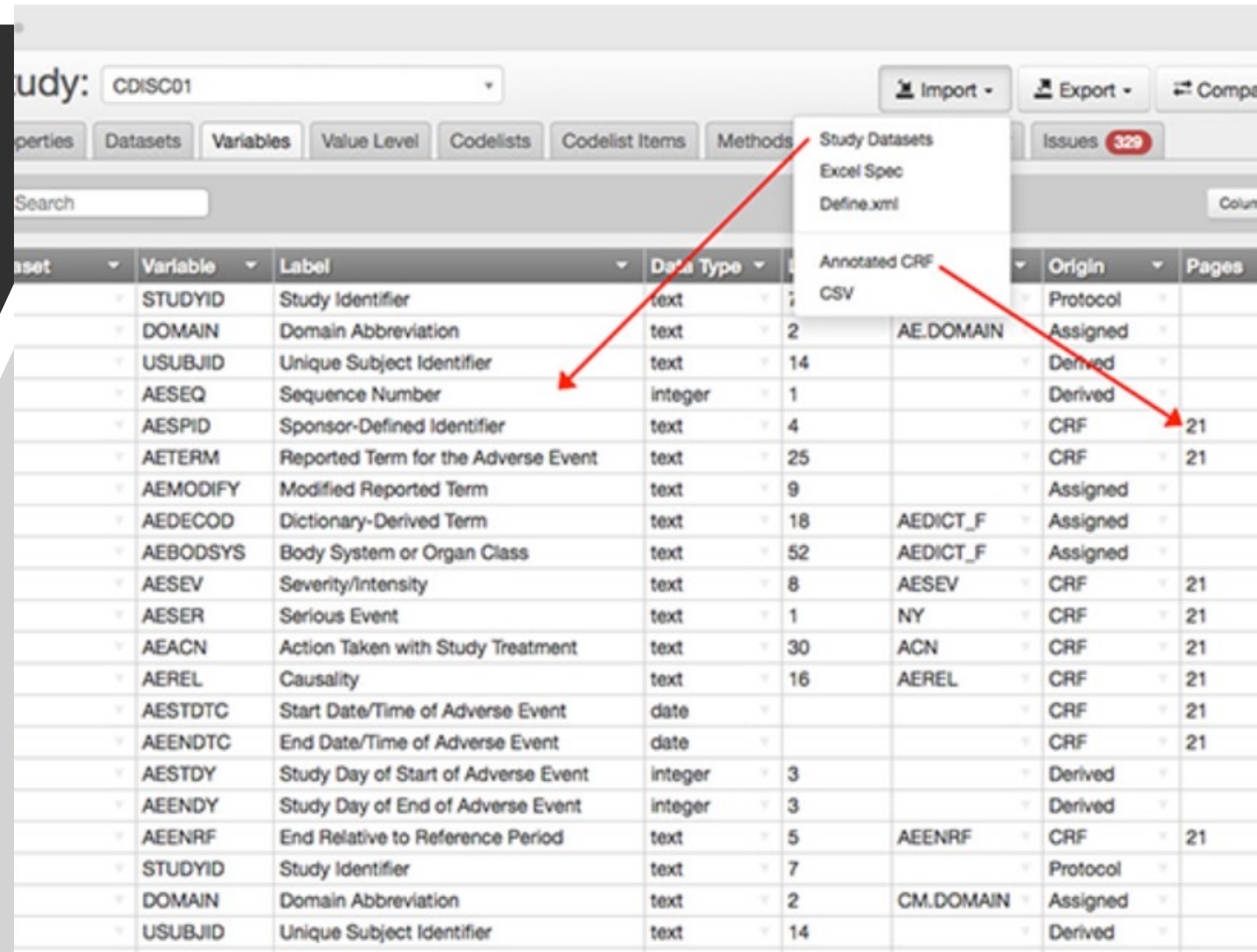
For example, Error DD0028: Term/NCI Code mismatch in Codelist 'Laboratory Test Code'. A reported term 'CALB' and its NCI Code 'C125942' are correct standard term and its corresponding NCI Code in CDISC Control Terminology.

Control terminology cont..

- Another example, DD0024: Invalid Term in Codelist 'No Yes Response' reporting 'N', 'U' or 'NA' terms which are standard terms in CDISC CT. Actual issue is due to invalid utilization of non-relevant terms for Flag variables like DTHFL, --BLFL or --PRESF where these 'N', 'U' and 'NA' terms are not applicable according to SDTM IG documentation. SDTM Flag variables are assigned to CDISC SDTM CT codelist (NY) 'No Yes Response' which includes 4 terms. However, SDTM IG specifies that these variables may only have either 'Y' or a missing value. For validation purpose, original NCI files of CDISC CT are extended by Pinnacle 21 with additional codelists including a subset of (NY) codelist limited to a single term 'Y'.

#4 quality dimensions affecting data score: Metadata

- Metadata validation rules look for issues with the define.xml, or inconsistencies between the define.xml and the study datasets.
- Pinnacle 21 extract metadata and merge with standards to get the best results.
- Uploading annotated CRF to automatically populate origin page numbers.



The screenshot displays the Pinnacle 21 software interface. At the top, a search bar shows 'Study: CDISC01'. Below it, a navigation bar includes tabs for 'Properties', 'Datasets', 'Variables', 'Value Level', 'Codelist', 'Codelist Items', and 'Methods'. A dropdown menu is open under the 'Methods' tab, showing options: 'Study Datasets', 'Excel Spec', 'Define.xml', 'Annotated CRF', and 'CSV'. A red arrow points from the 'Annotated CRF' option to the 'AESEQ' row in the table below. Another red arrow points from the 'CSV' option to the 'Origin' column header. The table lists various study variables with their labels, data types, and origin information.

Variable	Label	Data Type	Origin	Pages
STUDYID	Study Identifier	text	Protocol	
DOMAIN	Domain Abbreviation	text	Assigned	
USUBJID	Unique Subject Identifier	text	Derived	
AESEQ	Sequence Number	integer	Derived	
AESPID	Sponsor-Defined Identifier	text	CRF	21
AETERM	Reported Term for the Adverse Event	text	CRF	21
AEMODIFY	Modified Reported Term	text	Assigned	
AEDECOD	Dictionary-Derived Term	text	Assigned	
AEBODSYS	Body System or Organ Class	text	Assigned	
AESEV	Severity/Intensity	text	CRF	21
AESER	Serious Event	text	CRF	21
AEACN	Action Taken with Study Treatment	text	CRF	21
AEREL	Causality	text	CRF	21
AESTDTC	Start Date/Time of Adverse Event	date	CRF	21
AEENDTC	End Date/Time of Adverse Event	date	CRF	21
AESTDY	Study Day of Start of Adverse Event	integer	Derived	
AEENDY	Study Day of End of Adverse Event	integer	Derived	
AEENRF	End Relative to Reference Period	text	CRF	21
STUDYID	Study Identifier	text	Protocol	
DOMAIN	Domain Abbreviation	text	Assigned	
USUBJID	Unique Subject Identifier	text	Derived	

Dataset	Variable	Label	Data Type	Length	Code list	Origin	Pages
AE	STUDYID	Study Identifier	text	7		Protocol	
AE	DOMAIN	Domain Abbreviation	text	2	AE.DOMAIN	Assigned	
AE	USUBJID	Unique Subject Identifier	text	14		Derived	
AE	AESEQ	Sequence Number	integer	1		Derived	
AE	AESPID	Sponsor-Defined Identifier	text	4		CRF	21
AE	AETERM	Reported Term for the Adverse Event	text	25		CRF	21
AE	AEMODIFY	Modified Reported Term	text	9		Assigned	
AE	AEDECOD	Dictionary-Derived Term	text	Invalid 'LENGTH' value	AEDICT_F	Assigned	
AE	AEBODSYS	Body System or Organ Class	text	52	AEDICT_F	Assigned	
AE	AESEV	Severity/Intensity	text	8	AESEV	CRF	21
AE	AESER	Serious Event	text	1	NYX	CRF	
AE	AEACN	Action Taken with Study Treatment	text	30	ACN	CRF	
AE	AEREL	Causality	text	16	AEREL	CRF	21
AE	AESTDTC	Start Date/Time of Adverse Event	date			CRF	21
AE	AEENDTC	End Date/Time of Adverse Event	date			CRF	21
AE	AESTDY	Study Day of Start of Adverse Event	integer	3		Derived	
AE	AEENDY	Study Day of End of Adverse Event	integer	3		Derived	
AE	AEENRF	End Reason for the Adverse Event	text		AEENRF	CRF	21
CM	STUDYID	Study Identifier	text	7		Protocol	
CM	DOMAIN	Domain Abbreviation	text	2	CM.DOMAIN	Assigned	
CM	CMTRT	Reported Name of Drug, Med, or Therapy	text	23		CRF	9 22

Define.xml issues to be resolved to get clear report and improve data score. Highlighted in yellow or red are errors or invalid or missing

#5 quality dimensions affecting data score: Regulatory conformance

Regulatory Conformance validation rules look for violations of the guidance provided by regulatory agencies (FDA and PMDA) in their Technical Conformance Guides. These include missing datasets or variables requested by the regulatory agencies, and implementations inconsistent with the regulatory guidance.

According to the regulatory requirement, project team must ensure the Pinnacle 21 report 'clean' before e-submission. 'Clean' means no error in the PINNACLE 21 report, and the remaining warnings can be explained clearly in the Reviewers' Guide section 4.2 Issues Summary. Addressing most of the warnings and errors caused by mapping issue or the raw data issue. For mapping issues, programmers can update mapping logic and mapping program to correct them. That is under programmers' control. However, for raw data issues, it requires cross function collaboration to solve them. Some may need site correction and database to be updated. It normally takes quite long time to have the issue solved.

FDA Business Rule ID	FDA Business Rule	FDA Validator Rule	Domains	SDTM 3.1.2	SDTM 3.1.3	SDTM 3.2
FDAB001	A treatment-emergent flag must be submitted.	A treatment-emergent flag should be included in SUPPAE according to SDTM IG v3.1.2 #8.4.3.	SUPPAE	X	X	X
FDAB022	EPOCH should be included for clinical subject-level observations (e.g., adverse events, laboratory, concomitant medications, exposure, and vital signs).	Variables requested by FDA in policy documents should be included in the dataset. E.g., EPOCH and ELEMENT.	ALL	X	X	X

P21 has following rules that checks against these issues and reports them as warnings.

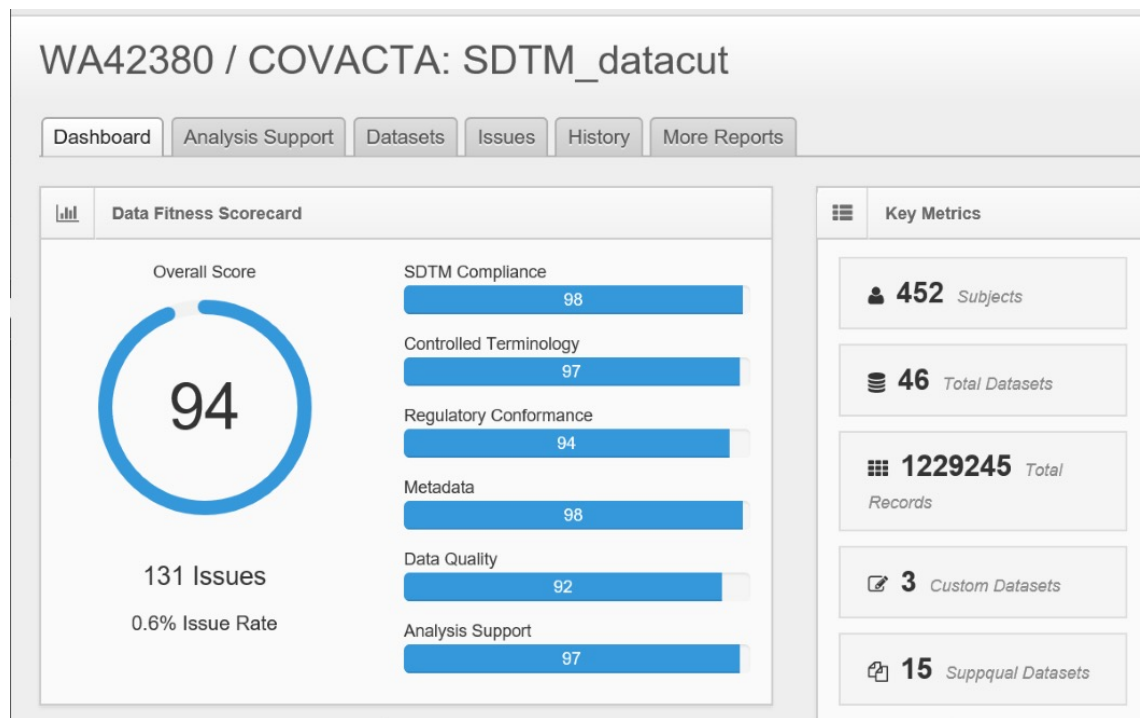
SD1077	FDAC021	FDA Expected variable EPOCH not found	Warning	1
SD1097	FDAC022	No Treatment Emergent info for Adverse Event	Warning	6794

Regulatory Conformance cont....:

It is also mentioned in regulatory conformance check, to make sure that the EPOCH variable is provided in the appropriate domains. The confusion associated with this validation rule is due to seemingly conflicting information between the regulatory agencies (i.e, FDA and PMDA) and CDISC guidance. While regulatory agency guidance states that the EPOCH variable should be provided, CDISC guidance lists this as a Permissible variable, which some sponsors interpret as meaning that it is not necessary to include the variable.

#6 quality dimensions affecting data score: Analysis support -gives summary look at which domains needs attention and look into to fix the issues. It provides various analysis topics that have issues and needs to be resolved.

Validation Spec	Score	Pass/Fail	Domains	Issues	Failures	Errors	Warnings
General Data Quality (3 Profiles)							
CDER Common Data Standards Issues	0	Fail	41	570	570	236	211
SDTM v3.1.2 General Data Quality	92	Pass	41	81	345,552	3,258	342,294
SDTM v3.1.3 General Data Quality	95	Pass	41	47	345,549	3,258	342,291
Laboratory Findings Analysis (6 Profiles)							
JReview Labs Baseline vs Max/Min Scatter Plots	0	Fail	3	2	784	392	392
JReview Hy's Law Patient Listing	0	Fail	3	2	784	392	392
JReview Hy's Law Plots	0	Fail	3	2	784	392	392
JReview Liver Function Baseline Box Whiskers	0	Fail	3	2	784	392	392
eDISH	96	Pass	6	7	17,993	0	17,601
Liver Function Analysis Panel	97	Pass	3	7	17,283	29	17,254
Overall Survival Analysis (1 Profiles)							
Overall Survival Analysis Panel	97	Pass	7	3	349	0	349
Standards Compliance (3 Profiles)							
SDTM v3.1.3 Compliance	77	Pass	41	185	175,996	408	175,444
SDTM Controlled Terminology	88	Pass	41	30	208,466	19,166	0
SDTM v3.1.2 Compliance	90	Pass	41	165	66,485	411	65,939
Subject Disposition Analysis (1 Profiles)							
Disposition Analysis Panel	59	Pass	4	2	629	392	237



RESTRICTED

Conclusion

DEVELOPMENT OF VALIDATION CHECKS IS A CONTINUOUS PROCESS. THERE ARE ALWAYS NEW STANDARDS, VERSIONS OF STANDARDS, AND BUSINESS RULES TO BE IMPLEMENTED.

TO INCREASE DATA SCORE, WE NEED TO MAKE SURE WE HAVE STUDY SPECIFICATION CREATED CORRECTLY, USE CORRECT VERSION OF CONTROL TERMINOLOGY AND MEDDRA. THE SEVERITY OF ISSUES WILL REDUCE YOUR SCORE, FIXING REJECTS WILL GREATLY HELP INCREASE YOUR DATA SCORE. RUNNING VALIDATING DATA AND SPECIFICATION (DEFINE.XML) BRINGS SCORES UP. ENTERING CORRECT CODELIST, UPDATING ISSUES WITH METADATA ON DEFINE.XML AND FOLLOWING REGULATORY CONFORMANCE RULES INCREASE DATA SCORE.



Is your data "fit for use"? Measure your Data
Fitness Score with Pinnacle 21 Enterprise

THANK YOU

RESTRICTED