

Metadata Errors not detected by Pinnacle 21 tool

SI05 – PHUSE Virtual EU Connect 2021

Caroline PELCKMANS (Statistical Programmer/Junior Biostatistician)



AGENDA

- **Introduction**
 - Pinnacle 21 (P21)
- **The flaws of P21**
 - False Positives
 - Issues not detected by P21
- **Conclusions**



#Data #DataDriven
#DataStrategy #DataIntegration
#DataViz #DataStandards
#DataScience #Datagovernance
#HealthCare #DataAnalysis
#ClinicalTrials
#ArtificialIntelligence
#AdvancedAnalytics
#MachineLearning #Digital
#Chatbot #Biometrics #CRO
#ClinicalData
#LifeSciencesExperts
#Whodrug #MedDRA
#Pharmacovigilance
#ClinicalResearchLifeCycle
#MedicalWriting
#MedicalInformation
#MedicalReview #SDTM
#ADaM #CDISC #Innovation
#RegulatorySubmissions
#Protocol #Coding

INTRODUCTION

PINNACLE²¹

- Designed by Programmers for Programmers
- Used to review the quality of submissions
 - Find issues that affect FDA/PMDA review
 - Fix quality issues at data cuts while studies are ongoing
 - Avoid costly delays in regulatory review
- Advocate for your data
 - Data Reviewer's Guide (SDRG and ADRG)

P21 – DATASET SUMMARY

Pinnacle 21 Validator Report								
Processed Sources								
Domain	Label	Class	Source	Records	Rejects	Errors	Warnings	Notices
GLOBAL	Global Metadata	--	--		0	1	0	0
ADAE	Adverse Events Analysis Dataset	ADAM OTHER	adae.xpt	414	0	10	0	0
ADEX	Exposure Analysis Dataset	ADAM OTHER	adex.xpt	977	0	12	0	0
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	adsl.xpt	240	0	42	3	0
ADSV	Subject Visits Analysis Dataset	ADAM OTHER	adsv.xpt	1693	0	0	0	0
AE	Adverse Events	EVENTS	ae.xpt	414	0	0	0	0
DM	Unrecognized	SYSTEM	dm.xpt	240	0	0	0	0
EX	Unrecognized	SYSTEM	ex.xpt	977	0	0	0	0
Total				4955	1	64	3	0
Unprocessed Sources								
Domain	Label	Class	Source	Records	Rejects	Errors	Warnings	Notices
Total					0	0	0	0
Grand Total				4955	1	64	3	0
Validation Summary Dataset Summary Issue Summary Details Rules (+)								

P21 – ISSUE SUMMARY

Pinnacle 21 Validator Report				
Issue Summary				
Source	Pinnacle 21 ID	Message	Severity	Found
GLOBAL				
	DD0101	Missing define.xml file	Reject	1
ADAE				
	AD0196	Required AEBODSYS value is null	Error	5
	AD0196	Required AEDECOD value is null	Error	5
ADEX				
	AD0099	EXSTDY is greater than EXENDY	Error	6
	AD0361	Value of ASTDT is greater than value of AENDT	Error	6
ADSL				
	AD0196	Required ARM value is null	Error	42
	CT2002	RACE value not found in 'Race' extensible codelist	Warning	3
Validation Summary Dataset Summary Issue Summary Details Rules ⋮ ⏪				

THE FLAWS OF PINNACLE 21



Does the report need to be completely issue-free?

Are we ready for regulatory submission, after we created an issue-free report?

P21 OUTPUT

Does the report need to be completely issue-free?

Reported issues
that need a fix
before submission

True Positive	False Positive
False Negative	True Negative

Reported issues that
are no real issues;
Report in xDRG

FALSE POSITIVES

- P21 Warnings are no real checks, but messages you need to review

- Reported warnings/issues can be False Positives



Report remaining
findings in xDRG

The report needs to be error-free, but **not** warning/issue free.

- Example: inconsistencies between guidelines for ARM/ACTARM

- **SDTM IG**

- recommended terms “Screen Failure”, “Not Assigned” and “Not Treated”

- **FDA Study Data Technical Conformance Guide**

- Missing value in ARM variable for all Screen Failure subjects
 - Missing value in ACTARM variable for all non-treated subjects
 - False Positive result in P21

- **PMDA** requirements (<-> FDA)

- Missing value in ARM or ACTARM variables is a violation

Does the report need to be completely issue-free?

No



P21 UNREPORTED ISSUES

Are we ready for submission, after we created an issue-free report?



FALSE NEGATIVE

- Due to missing validation messages:
 - False Negatives, that shouldn't be ignored, often go unnoticed.
- Minimize False Negative results
 - Non-reported violations of Rejection Regulatory rules will delay submission review
 - Low quality data may compromise results
- Reasons:
 - Incorrect algorithms for existing checks
 - No check available in current P21 version

FALSE NEGATIVE: EXAMPLE 1

- Complex Algorithms are not checked by P21
 - Algorithms are set-up based on the information presented in protocol/SAP
 - In metadata, they are explained in user-friendly language for reviewers
- Example of an issue that will not be detected

ADEG.AVALOUTL	If PARAMCD="AVQTCFAG" then do; Equals "Value of > 450 and ≤ 480 ms" when 450<AVAL≤480 and not present at baseline (before first dose intake); Equals "Value of > 480 and ≤ 500 ms" when 480<AVAL≤500 and not present at baseline (before first dose intake); Equals "Value of > 500 ms" when AVAL>500 and not present at baseline (before first dose intake); End;
---------------	--



Review by Biostatistician

ADEG.AVALOUTL	If PARAMCD="AVQTCFAG" then do; Equals "Treatment-emergent value of > 450 and ≤ 480 ms" when 450<AVAL≤480 and not present at baseline (before first dose intake); Equals "Treatment-emergent value of > 480 and ≤ 500 ms" when 480<AVAL≤500 and not present at baseline (before first dose intake); Equals "Treatment-emergent value of > 500 ms" when AVAL>500 and not present at baseline (before first dose intake); End;
---------------	---

FALSE NEGATIVE: EXAMPLE 2A

- Labels of variables:
 - Should be consistent no matter the data structure (BDS or OCCDS)
 - Inconsistent labels not detected by P21
- Example: Concomitant vs Prior Medications

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	ROLE
ADCM	PRCM	Analysis Category	char	14	Selection
ADPR	PRCM	Prior/Concomitant	char	14	Selection



Double QC by another
team member

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	ROLE
ADCM	PRCM	Prior/Concomitant	char	14	Selection
ADPR	PRCM	Prior/Concomitant	char	14	Selection

FALSE NEGATIVE: EXAMPLE 2B

- Check that SDTM variables copied in ADAM remain the same with all attributes.

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	ROLE	USER DEFINED FORMAT	CDISC CT	COMP. ALGORITHM	EXTERNAL DICTIONARY	ORIGIN (name of ADAM dataset or derived)
ADSL	AGE	Age	num	8	Analysis	20.				DM
ADSL	AGEU	Age Units	char	5	Analysis		AGEU			DM
ADSL	AGEGR1	Pooled Age Group 1	char	11	Analysis		AGEGR	ADSL.AGEGR1		Derived
ADSL	AGEGR1N	Pooled Age Group 1 (N)	num	8	Analysis		AGEGRN	ADSL.AGEGR1N		Derived
ADSL	COUNTRY	Country	char	10	Analysis		COUNTRY			DM
ADSL	ACOUNTRY	Analysis Country	char	3	Analysis			ADSL.ACOUNTRY		Derived

Double QC by another team member/
Use macros to automate checks

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	ROLE	USER DEFINED FORMAT	CDISC CT	COMP. ALGORITHM	EXTERNAL DICTIONARY	ORIGIN (name of ADAM dataset or derived)
ADSL	AGE	Age	num	8	Analysis	8.				DM
ADSL	AGEU	Age Units	char	5	Analysis		AGEU			DM
ADSL	AGEGR1	Pooled Age Group 1	char	11	Analysis		AGEGR	ADSL.AGEGR1		Derived
ADSL	AGEGR1N	Pooled Age Group 1 (N)	num	8	Analysis		AGEGRN	ADSL.AGEGR1N		Derived
ADSL	COUNTRY	Country	char	3	Analysis		COUNTRY			DM
ADSL	ACOUNTRY	Analysis Country	char	13	Analysis		ACOUNTRY	ADSL.ACOUNTRY		Derived

FALSE NEGATIVE: EXAMPLE 3

- Code lists that don't make sense will not always be detected by P21
 - For example, code lists that are not part of CT need to be checked manually

Sponsor value	Sponsor text	Format name	Value	Text	Subset	Flag (SDTM/ADaM)	Codelist Label
<18 years	<18 years	AGEGR	<18 years	<18 years	ALL	ADaM	Age group
≤ 65 years	≤ 65 years	AGEGR	≤ 65 years	≤ 65 years	ALL	ADaM	Age group
> 65 years	> 65 years	AGEGR	> 65 years	> 65 years	ALL	ADaM	Age group
1	<18 years	AGEGRN	1	<18 years	ALL	ADaM	Age group (N)
2	≤ 65 years	AGEGRN	2	≤ 65 years	ALL	ADaM	Age group (N)
3	> 65 years	AGEGRN	3	> 65 years	ALL	ADaM	Age group (N)

Are we ready for submission, after we created an issue-free report?



SOLUTION: CHECKLIST (USE MANUAL CHECKS AND MACROS)

		Description	ADSL	Adxx	Adxx	Adxx
		Structure Type	ADSL	BDS	OCC	TTE
1	All	Check that the number of analysis datasets is optimum	x			
2	All	Check that the analysis dataset has an informative name	x			
3	All	Check that the variables defined as Keys (B5) are sufficient to uniquely identify each record and that they are set in the correct order.	x			
4	All	Check that order of the variables is logical and follow the CDISC convention	x			
5	All	Check the name of variables. Note: before creating a new variable, first check if it does not exist in ADaM implementation guide or in SDTM	x			
6	All	Check that Study Data Tabulation Model (SDTM) variables copied in the Analysis Dataset (AD) remains with all attributes the same	x			
7	ADSL	Check that no Subject-Level Analysis Dataset (ADSL) variables are derived from another AD (except if Deviations Analysis Dataset (ADDV) should be created in order to define populations)	x			
8	ADSL	Check that all necessary populations are available in ADSL dataset (all populations defined in the Statistical Analysis Plan (SAP) have to be present in ADSL; if no specification in the SAP, at least one population flag has to be present in ADSL).	x			
9	All	Check the algorithms on correctness (by using SAP and Protocol as reference)	x			
10	All	Ensure variables are not truncated. Use macro to define minimum length to report in mapping	x			

CONCLUSION: P21 IS NOT FLAWLESS



Issues and challenges

- False Positive warnings
- Unreported Issues (False Negative results)



Solution

- Use xDRG to report False Positives
- Final checklist of items that cannot be validated by P21; manual check + use of macros



Results and Benefits

- Be time efficient
- Avoid costly delays in Regulatory Review

CONTACT



Caroline Pelckmans

Statistical Programmer/ Biostatistician

Caroline.pelckmans@businessdecision.be



Amine Khemiri

Statistical Programmer

amine.khemiri@businessdecision.be



Pamphile Beaujean

Statistical Programmer

pamphile.beaujean@businessdecision.be



THANK YOU FOR YOUR ATTENTION
Q&A