Implementing FDA Validation Rules by Pinnacle 21

PhUSE EU CSS

Sergiy Sirichenko
June 19, 2017

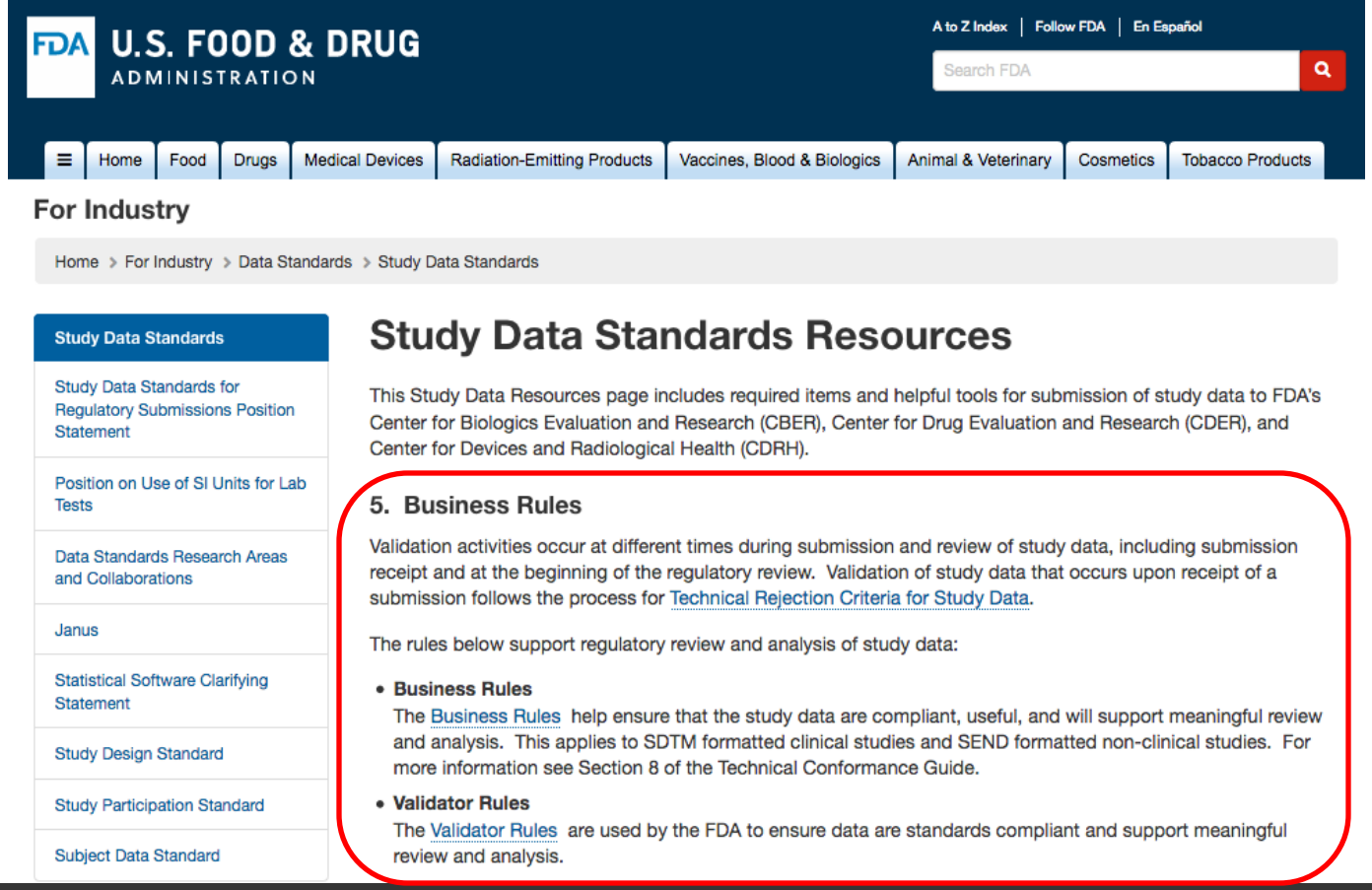
PINNACLE 21

Presenter – Sergiy Sirichenko

- Co-founder of OpenCDISC
- Principal Data Scientist at Pinnacle 21
- Validation rules implementer
- SME on FDA JumpStart and e-Data Support Services

Disclaimer

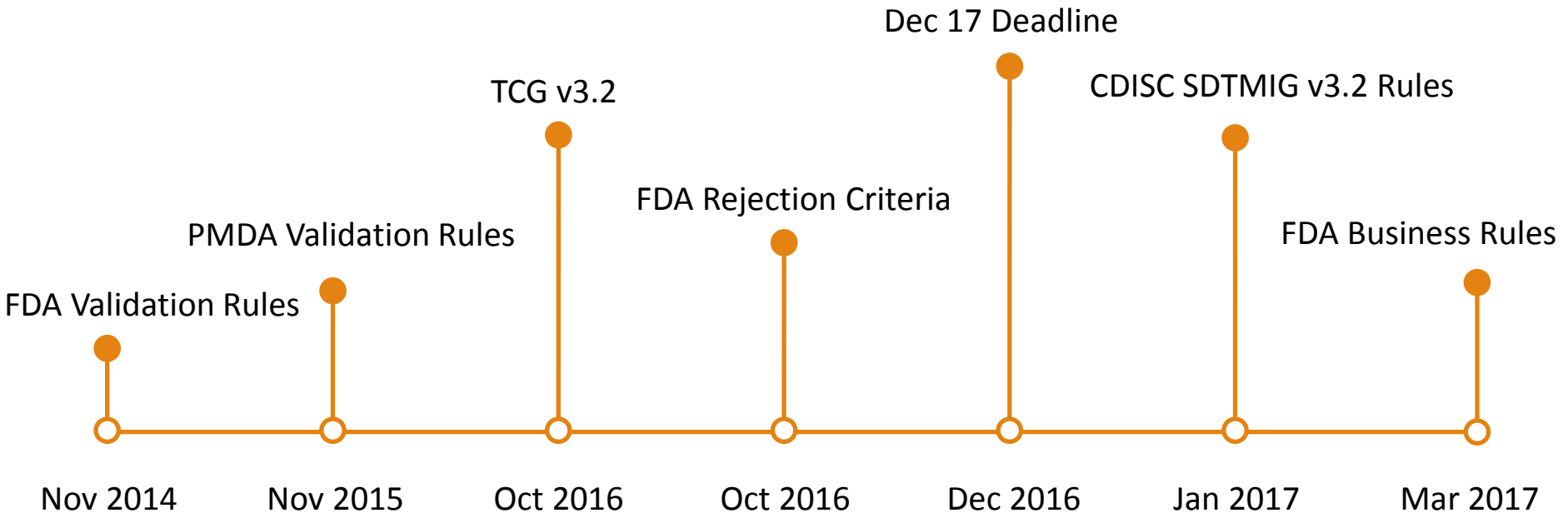
The views and opinions presented here represent those of the speaker and should not be considered to represent advice or guidance on behalf of the Food and Drug Administration and Pharmaceuticals and Medical Devices Agency



FDA Business Rules

Published: March 14th, 2017

https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm2005545.htm

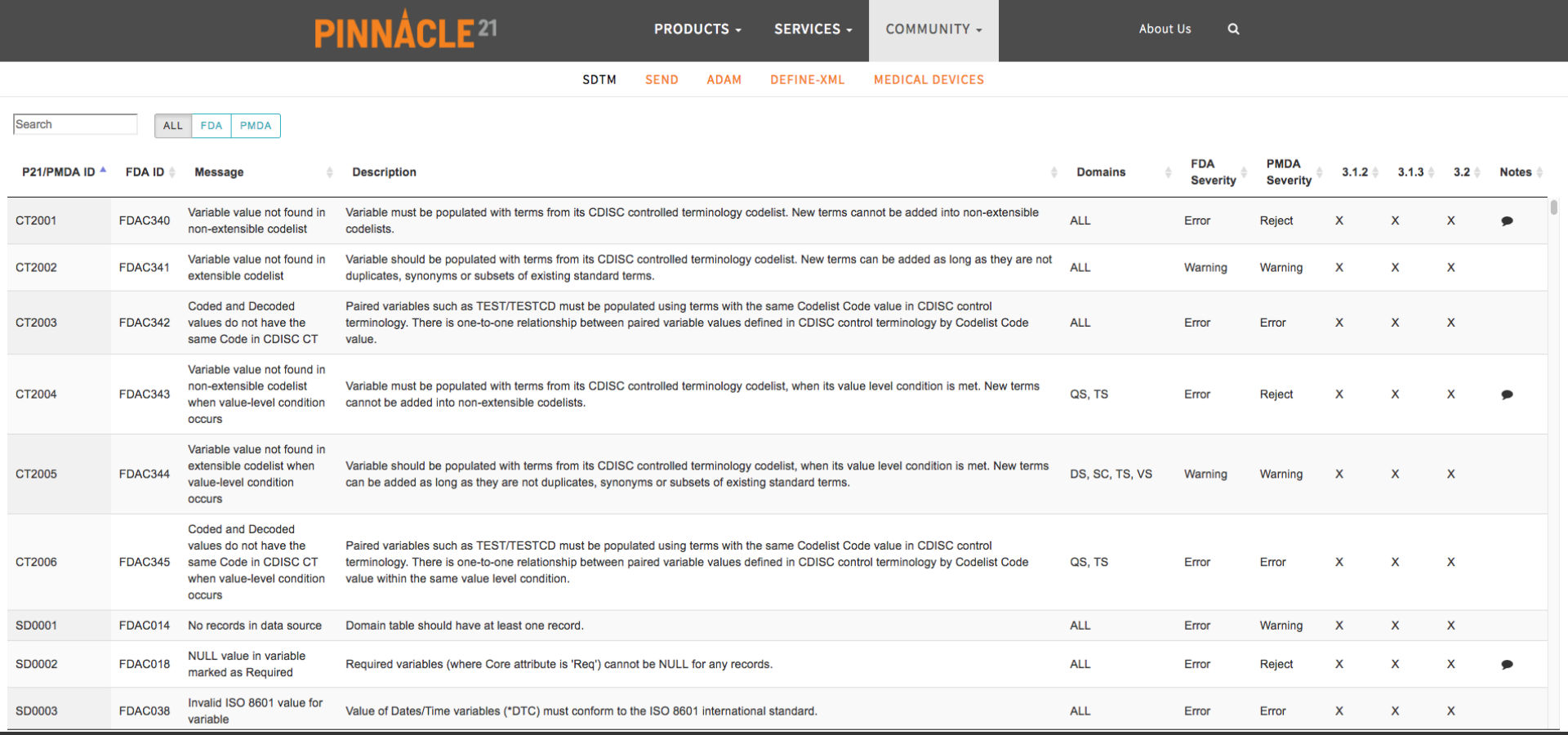


How did we get here?

FDA Business Rules is a result of multi-year effort to ensure that the study data are compliant, useful, and will support meaningful review and analysis

FDA and PMDA rules

- New FDA Validation Rules
  - Published: November 18, 2014
  - https://www.youtube.com/watch?v=pZ9YTdldQkk
- Original FDA Validation Rules
  - Published: November 18, 2014
  - https://www.pinnacle21.com/blog/fda-validation-rules-webinar-recap
- PMDA Validation Rules
  - Published: November 24, 2015
  - https://www.pinnacle21.com/blog/pmda-validation-rules-webinar-recap



P21/PMDA ID	FDA ID	Message	Description	Domains	FDA Severity	PMDA Severity	3.1.2	3.1.3	3.2	Notes
CT2001	FDAC340	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.	ALL	Error	Reject	X	X	X	
CT2002	FDAC341	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.	ALL	Warning	Warning	X	X	X	
CT2003	FDAC342	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.	ALL	Error	Error	X	X	X	
CT2004	FDAC343	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.	QS, TS	Error	Reject	X	X	X	
CT2005	FDAC344	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.	DS, SC, TS, VS	Warning	Warning	X	X	X	
CT2006	FDAC345	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.	QS, TS	Error	Error	X	X	X	
SD0001	FDAC014	No records in data source	Domain table should have at least one record.	ALL	Error	Warning	X	X	X	
SD0002	FDAC018	NULL value in variable marked as Required	Required variables (where Core attribute is 'Req') cannot be NULL for any records.	ALL	Error	Reject	X	X	X	
SD0003	FDAC038	Invalid ISO 8601 value for variable	Value of Dates/Time variables (*DTC) must conform to the ISO 8601 international standard.	ALL	Error	Error	X	X	X	

Validation Rules Browser

Search across all available validation rules, locate rule-related information, and see what's required by FDA and PMDA

https://www.pinnacle21.com/validation-rules

Study Data Technical Conformance Guide v3.2

- Published: October 2016
- Expanded validation rules to 3 types
- Introduced Technical Rejection Criteria
- Introduced FDA Business Rules
- Clarified the role of SDOs in validation

Types of Validation Rules

- Technical Rejection Criteria
- Standards Conformance
- FDA Business

Technical Rejection Criteria

- Published: October 2016
- First step in enforcing the Dec 17th Study Data Standards Guidance
- Added to existing eCTD validation criteria
- FDA will give industry 30 days notice

https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM523539.pdf

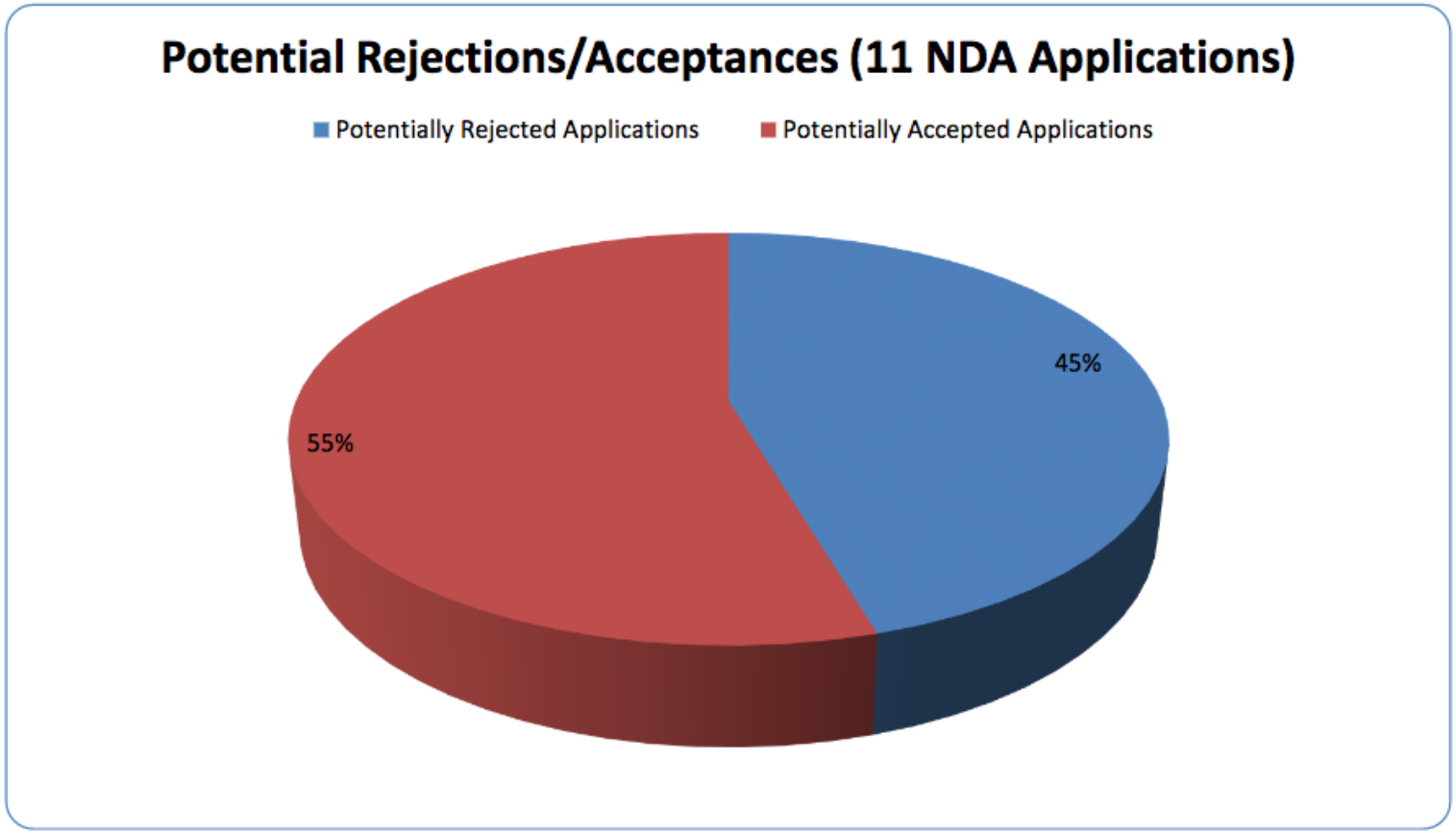
There are just 2 rules – So far

- Rule # 1734:
  - Trial Summary (TS) dataset must be present for each study in Module 4 and 5
- Rule # 1736:
  - Demographic (DM) dataset and define.xml must be submitted in Module 4 for nonclinical data
  - DM dataset, Subject level analysis dataset (ADSL) and define.xml must be submitted in Module 5 for clinical data

Covered by Pinnacle 21 rules

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

How many FDA applications since December 17th were compliant?



FDA ran a simulation on new NDA applications from January to March 1st, 2017 with results above

45% would have been rejected if criteria was in force today

http://www.phusewiki.org/docs/2017_CSS_US/Introduction_CrystalAllard.pdf

FDA vs. PMDA Rejection Rules

4 FDA **vs** **34** PMDA

Types of Validation Rules

- Technical Rejection Criteria
- **Standards Conformance**
- FDA Business

Standards Conformance Rules

- Provided by CDISC teams
- FDA will no longer publish standards conformance rules

P21

Available CDISC Rules

- SDTMIG v3.2
- ADaMIG v1.0
- ADaMIG v1.1 (planned for Q2)
- SENDIG v3.1 (in the works, ≈2018)
- Define-XML (?)
- TAUGs (???)

SDTMIG v3.2 Conformance Rules

- Published: January 27, 2017
- 410 total rules
- 325 are “Programmable”
  - Can be implemented as automated check
- 85 are “Conditional”
  - May be programmable based on conditional factors
- 17 of Conditional rules are actually Programmable for FDA and PMDA

P21 support for SDTMIG rules

- cover 212 of 325 programmable rules
- 104 need to be implemented
- 9 need clarification
  - Conflicts with FDA business rules
  - Misclassified as programmable

Example of conflict in rules

- CDISC
  - ARM/ARMCD and ACTARM/ACTARMCD are Required variables
  - Required variables must always be included in the dataset and cannot be null for any record
- FDA
  - Screen failures, when provided, should be included as a record in DM with the ARM field left blank. For subjects who are randomized in treatment group but not treated, the planned arm variables (ARM and ARMCD) should be populated, but actual treatment arm variables (ACTARM and ACTARMCD) should be left blank.

ADaMIG v1.0 Conformance Rules

- Published: September 20, 2010
- Last updated: January 20, 2015
- Pinnacle 21's implementation:
  - https://www.pinnacle21.com/blog/how-does-pinnacle-21-implement-adam-validation-checks

P21

Types of Validation Rules

- Technical Rejection Criteria
- Standards Conformance
- **FDA Business**

FDA Business Rules

- Is study data useful and supports meaningful review and analysis?
- Complements Technical Rejection and Standards Conformance rules
- Supersedes previously published FDA validation rules

FDA Business Rule ID	FDA Business Rule
Clinical and Nonclinical	
FDAB005	Age or age range should be provided for all subjects, except for Screen Failures.
FDAB008	All exposure records should occur between First and Last Study Treatment dates.
FDAB009	All paired variables should have a one-to-one relationship. Examples include Short Name and Name of Test; Parameter Name and Parameter Code or Number; Variable Name and Variable Label, etc.
FDAB011	All Trial Design data should be submitted as specified in the Technical Conformance Guide (TCG).
FDAB012	Assessment results should include units whenever a unit of measure is available.
FDAB013	Baseline flags for Laboratory results, Vital Signs, ECG, Pharmacokinetic Concentrations, and Microbiology results should be submitted if the data was collected or can be derived.
FDAB015	Character values should not have leading spaces or only have a period character.
FDAB016	Collection Study Day should be populated when Date/Time of Collection is available.
FDAB017	Controlled terms should use the exact same case used by the terminology maintenance organizations (e.g., MedDRA, CDISC controlled terminology).
FDAB018	A variable's length across a study should be no longer than the maximum length of the actual data (except for SUPPQUAL).
FDAB019	SUPPQUAL variable length should be no longer than the maximum length of the actual data within the dataset.
FDAB020	Demographics (DM) and Trial Summary (TS) domains must be submitted.
FDAB021	Duplicate records should not be submitted (as constrained by the unique key in the underlying standard).
FDAB024	Large datasets should be split into smaller datasets no larger than 5 GB in size.

FDA Business Rules

45 business rules for clinical data

41 business rules for nonclinical data

FDA Validator Rules

- Accompany Business Rules
- Describe implementation of Business Rules in DataFit (a.k.a. P21 Enterprise)
- Published for transparency

FDA Business Rule ID	FDA Business Rule	FDA Validator Rule	Domains	SDTM 3.1.2	SDTM 3.1.3	SDTM 3.2	SEND 3.0 Coming Soon
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	
FDAB002	A value for a Toxicity (--TOX) variable should be provided, when a Toxicity Grade (--TOXGR) variable value is greater than 0.	A value for a Toxicity (--TOX) variable should be provided, when a Toxicity Grade (--TOXGR) variable value is populated and greater than 0.	AE, MH, CE, EG, LB, PC, PP	X	X	X	
FDAB003	Adverse Events must be coded using 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).	AE	X	X	X	
		Value for Body System or Organ Class (--BODSYS) variable must be populated using a System Organ Class of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	AE, MH, CE	X	X	X	
		Value for Preferred Term Code (--PTCD) variable must be populated using a Preferred Term Code of the MedDRA dictionary of a version specified in the define.xml.	AE, MH, CE		X	X	
		Value for Lowest Level Term (--LLT) variable must be populated using a Lowest Level Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	AE, MH, CE		X	X	
		Value for High Level Term (--HLT) variable must be populated using a High Level Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	AE, MH, CE		X	X	
		Value for High Level Term Code (--HLTCD) variable must be populated using a High Level Term Code of the MedDRA dictionary of a version specified in the define.xml.	AE, MH, CE		X	X	
		Value for High Level Group Term (--HLGT) variable must be populated using a High Level Group Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	AE, MH, CE		X	X	
		Value for High Level Group Term Code (--HLGTCD) variable must be populated using a High Level Group Term Code of the MedDRA dictionary of a version specified in the define.xml.	AE, MH, CE		X	X	
		Value for Body System or Organ Class Code (--BDSYCD) variable should be populated using a System Organ Class Code of the MedDRA dictionary of a version specified in the define.xml.	AE, MH, CE		X	X	

FDA Validator Rules

115 validator rules for clinical data

? validator rules for nonclinical data (coming soon)

Where is Severity?

- FDA will no longer publish Severity
- It was confusing with many sponsors ignoring Warnings
- All rules, with the exception of Technical Rejection Criteria, are now equal

Action Plan

Where do we go from here?

What to expect from Pinnacle 21?

- Working on implementing new rules for
  - SDTMIG v3.2 and ADaMIG v1.1 – final rules
  - Define-XML v2.1 and SENDIG v3.1 – draft rules
- Working with FDA, PMDA, and CDISC to clarify rules and resolve any conflicts
- Launching public BETA in August to solicit feedback from users
- Production release in November

What to expect from Pinnacle 21?

- Papers, blogs, webinars, and training on "Good Data Validation Practices"
  - What is data validation?
  - How to configure Validator?
  - How to perform data validation?
  - How to interpret validation results?
  - How to evaluate risk of data issues?
  - How to fix data errors?
  - How to explain data issues?

What to expect from FDA?

- Updated Validator rules for SENDIG v3.0
- Updated Business and Validator rules to
  - Clarify rules and correct any issues and inconsistencies
  - Take advantage of growing experience and cross-center collaboration

What to expect from PMDA?

- New release of validation rules that incorporates
  - New SDTMIG v3.2 and ADaM v1.1 rules
  - FDA Business Rules
  - Pilot experience

Actions for P21 Enterprise users

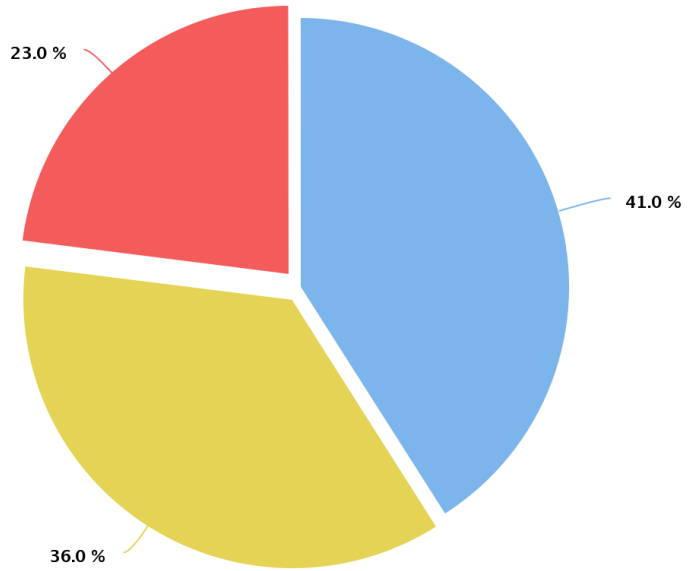
- Your environment is always in sync with FDA and PMDA, so no actions here
- Provide feedback during BETA
- Learn and follow "Good Data Validation Practices"

Actions for P21 Community users

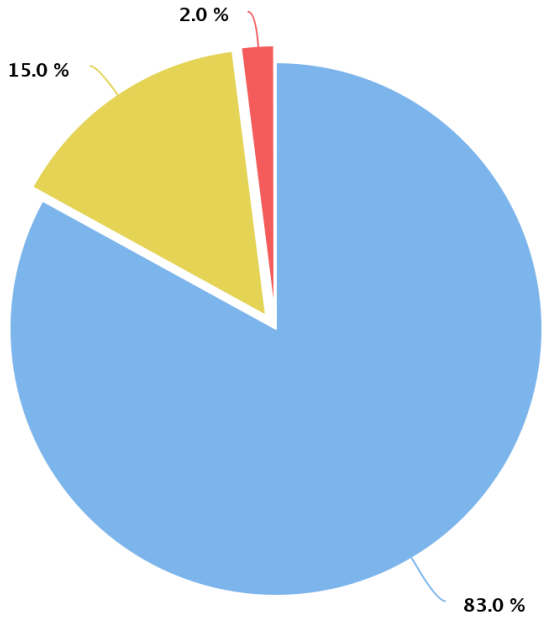
- Get your environment in sync with FDA and PMDA, **ASAP**

- Compatible versions of Community
  - FDA - 2.2.0
  - PMDA - 2.1.3 or above

US/EU Users



Japan Users



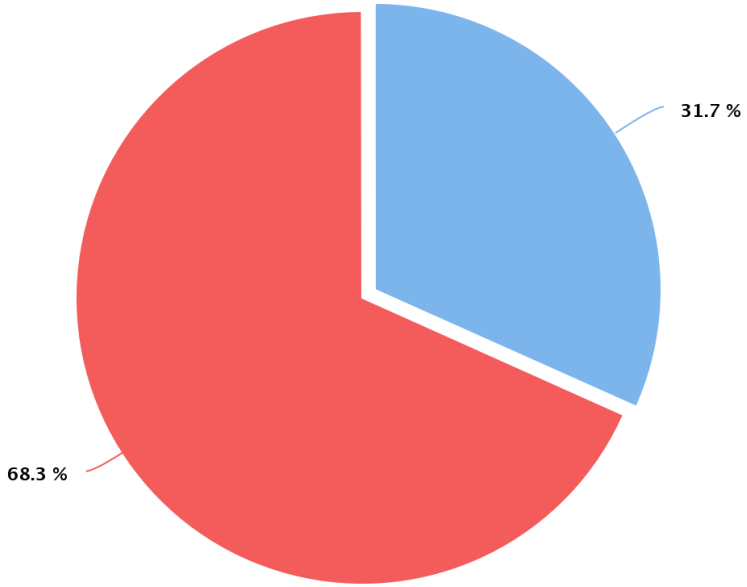
Compatible Not compatible In need of intervention (below 2.0.0)

How many users are compatible?

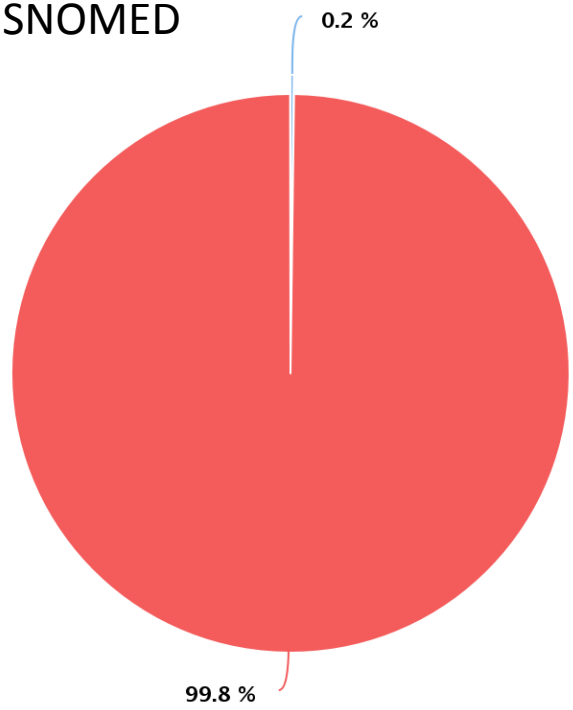
59% of US users are not compatible with FDA

17% of Japan users are not compatible with PMDA

MedDRA



SNOMED



Using Not using

How many are using dictionaries?

68.3% of users have not configured MedDRA

99.8% of users have not configured SNOMED

Common issues with compatibility

- Using older versions of Community
- Not configuring MedDRA and SNOMED
- Customizing rule configs
- Customizing CDISC CT files
- Running validations on a server or shared environment

Implementation of Data Validation

Data Quality is What Matters

- Data quality is defined by intended use
- All users are different

Data Quality by Intended Use

- FDA Clinical Trial Repository (Janus)
  - Can I load data?
- FDA JumpStart analyst
  - Can I run tools?
- Reviewer
  - Can data support meaningful review and analysis?

Message and description are user-specific

- P21 checks are designed for Programmers
  - *"Inconsistent LBSTRESU"*
  - Assuming good knowledge of CDISC standards
- Example of Reviewer-friendly messages:
  - *"Conversion of collected results into the same units was not done for all records"*
- Example of message for Data Managers:
  - *"Subject ABC-001 has potentially invalid Visit Dates: VISIT 2: 2017-02-19, VISIT 3: 2016-02-27"*

PhUSE Guidelines for Validation Rule Developers

- *"**3. Message** (required)*

Message is a simple and short statement specifying what went wrong. Messages need to be concise, complete and consistent. The messages should be simple statement capturing the error (ex: invalid, missing, and inconsistent) Syntax: – Element name Should/SHALL – Variable Label [Variable Name] SHOULD/SHALL be something, Condition (when or IF something) Example: The value of Age (AGE) SHOULD be greater than 0."

- http://www.phusewiki.org/wiki/index.php?title=Guidelines_for_Validation_Rule_Developers

Rules implementation

- There is no exact implementation of all business rules due to
  - Multiple stakeholders
  - Limitations in programing algorithms
- P21 provides traceability by *Publisher ID*

Pinnacle 21 Interpretation Errors and Warnings

- Errors are issues reported with 100% confidence
  - *Start Date is after End Date*
  - *Missing SDTM required variable*
- Warnings are potential issues requiring manual review
  - *Terms added to Extensible CT*
  - *Missing Units on Results*

FDA/PMDA Interpretation of Severity

- Error/Warning – issues with more/less severe impact on review process
- Reject – critical issues that prevent review and automation processes
- Since Mar 14th, FDA no longer publishes Severity, except Rejection criteria

False-Negative and False-Positive messages

- Most diagnostics tests have FP and FN results
- Validation is diagnostics test for quality of study data
- Rule tuning can change detection sensitivity of real data issues
- However, a decrease of false-positive issues may also increase the number of non-reported actual issues

What to do?

- Minimizing false-positives when manual review is expensive
  - for example, extra work for data vendors or when missing planned timelines
- Minimizing false-negatives when impact of non-detected issue is high
  - for example, non-detected issue could delay regulatory review

A few notes

- Almost nobody complains about false-negatives
  - Issue is still the issue regardless if it is reported or not!
- Programming bug is a bug, so report it to P21 team for fast fix
- Invalid business rule will result in invalid P21 rule, so report it to rule publisher

Summary & Questions

- Sponsors should be ready to comply with 3 types of validation rules
  - FDA Reject → FDA Business → Standards Conformance
- You can rely on P21 to provide tools/education
- Community users need to ensure they are compatible
- Data quality is defined by intended use
- All users are different

Sergiy Sirichenko
ssirichenko@pinnacle21.net