SUPPQUAL DATASETS: GOOD BAD AND UGLY

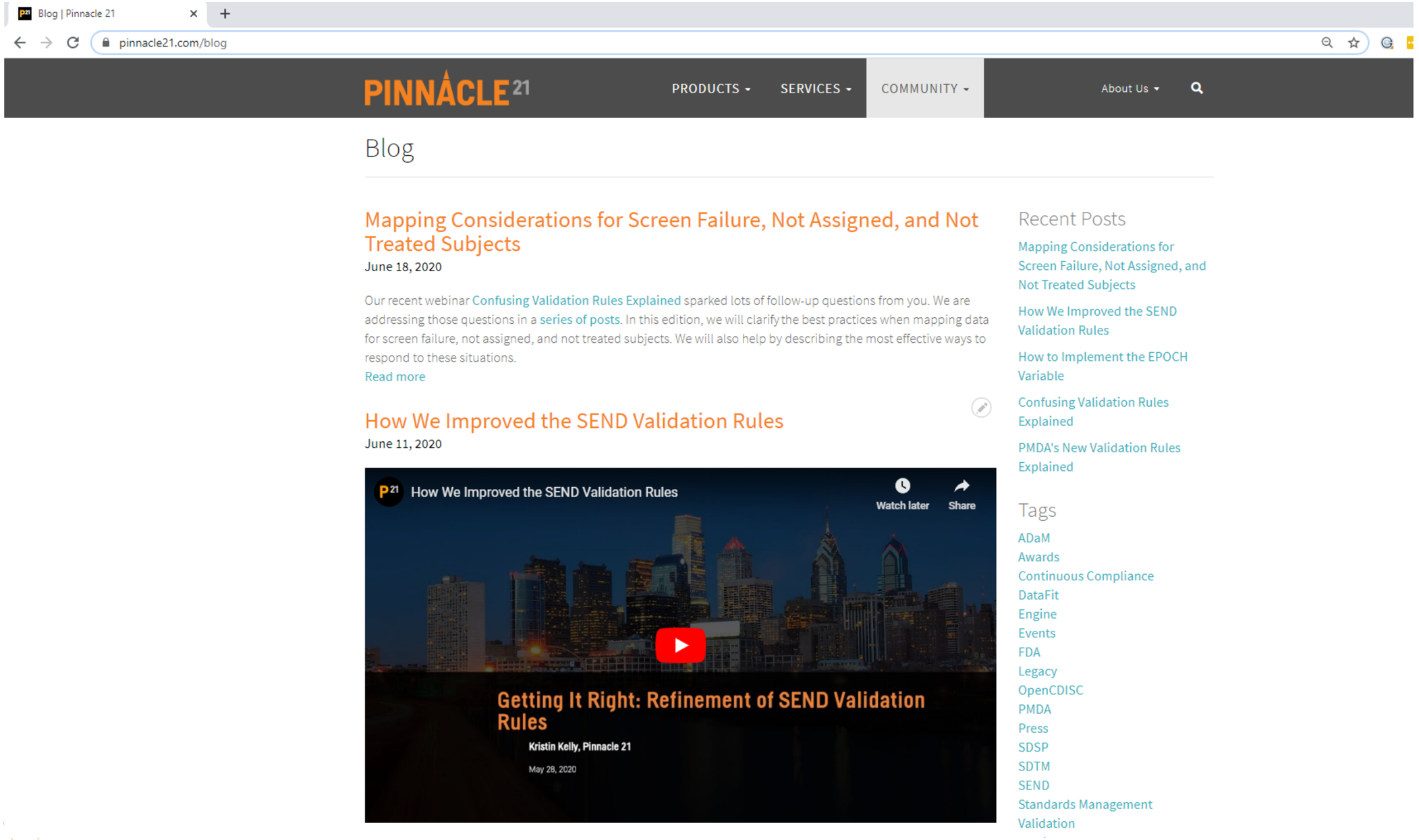
Sergiy Sirichenko, Pinnacle 21

July 2, 2020

PRESENTER – SERGIY SIRICHENKO

- Co-founder of OpenCDISC
- SME on FDA JumpStart, coreDF, and eDATA projects
- User advocate

WWW.PINNACLE21.COM/BLOG



- Recording
- Slides
- Q&A

How We Improved the SEND Vali | Pinnacle 21 - YouTube

pinnacle21.com/blog/how-we-improved-send-validation-rules

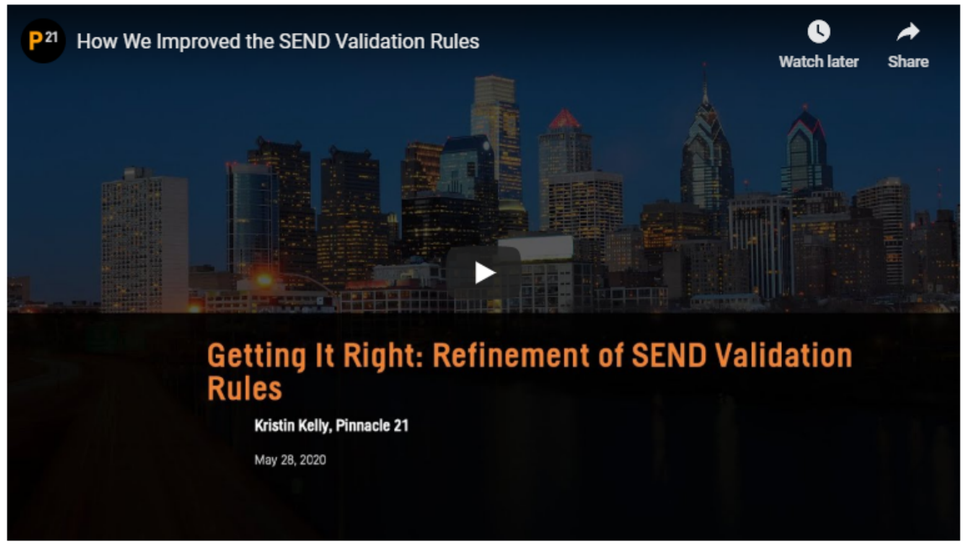
PINNACLE 21 | PRODUCTS | SERVICES | COMMUNITY | About Us

Blog / How We Improved the SEND Validation Rules

How We Improved the SEND Validation Rules

View | Edit

June 11, 2020



For years, the validation rules for nonclinical data were just an extension of published SDTM rules. However, through active collaboration of Pinnacle 21, FDA and the industry over the past year, the SEND rules have been refined. We have modified many of the existing rules, removed some and added others.

On May 28th, Pinnacle 21 hosted a webinar in which Kristin Kelly discussed the changes that have been made to get the rules right for SEND. Learn more by browsing the webinar resources.

Webinar Resources

- Download the slide deck
- Watch the video above
- Read the Q&A below

Answers to Your Questions

1. **Does the duplicate check refer to dataset keys in define.xml?**
   SD1117 does not currently check against the keys in the define.xml. It checks against a specific set of qualifier and timing variables with the exception of variables that can be sponsor-defined, such as --SPID, --REFID, etc.

Recent Posts

Mapping Considerations for Screen Failure, Not Assigned, and Not Treated Subjects

How We Improved the SEND Validation Rules

How to Implement the EPOCH Variable

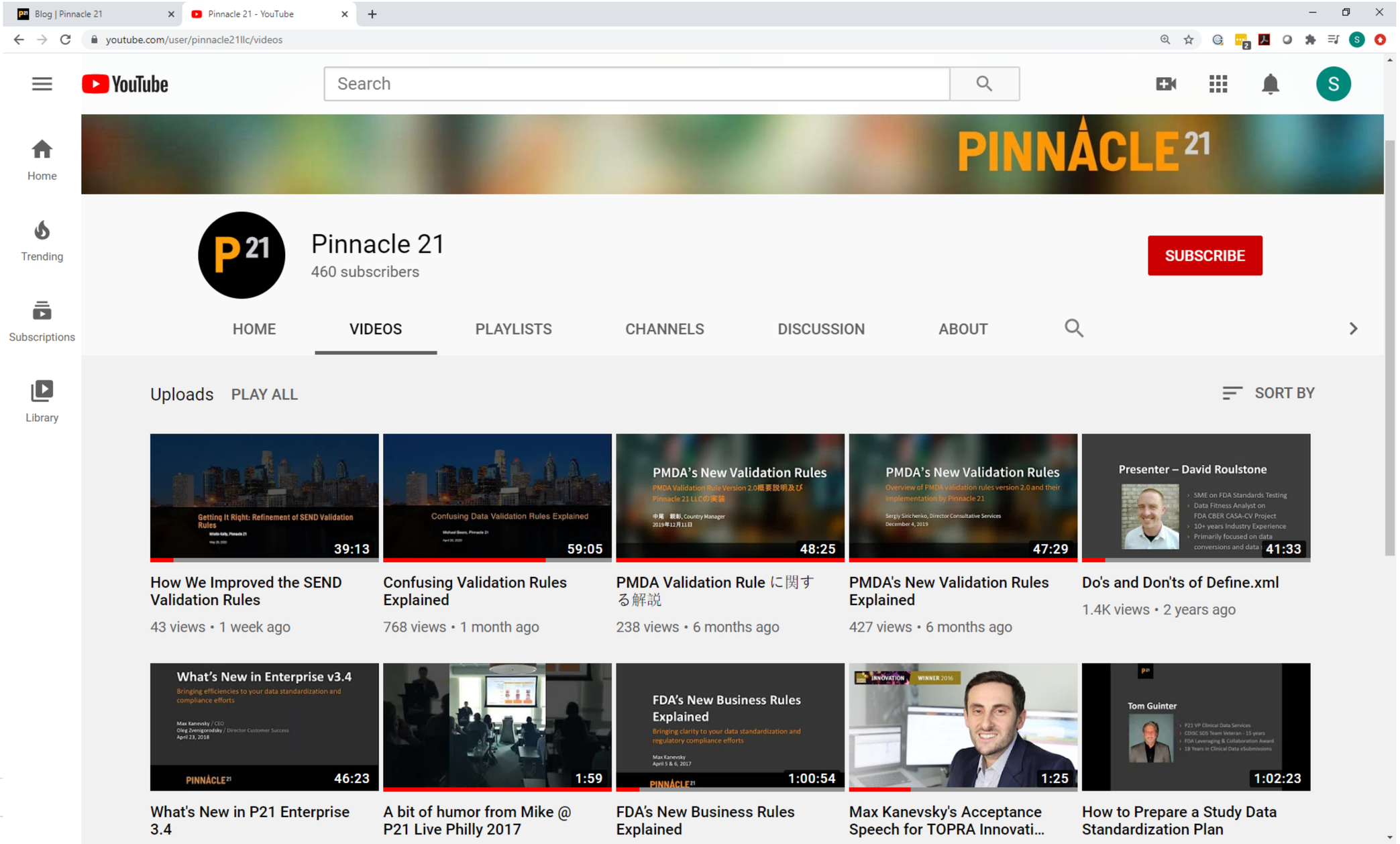
Confusing Validation Rules Explained

PMDA's New Validation Rules Explained

Tags

ADaM
Awards
Continuous Compliance
DataFit
Engine
Events
FDA
Legacy
OpenCDISC
PMDA
Press
SDSP
SDTM
SEND
Standards Management
Validation
Webinars
White Papers

WWW.YOUTUBE.COM/USER/PINNACLE21LLC/VIDEOS



INTRODUCTION

Standardized representation of sponsor’s non-SDTM variables

SUPPLEMENTAL QUALIFIERS

- Standardized representation of sponsor's non-SDTM variables (**NSV**)
- Intended to capture additional *Qualifiers*
- Other type of data should not be stored in SUPPQUAL
  - Separate observations
  - Subject Characteristics (SC) domain information
  - Interpretations
  - Information which required additional qualifiers like units
  - Timing information
  - Info about non-occurrence events
  - Comments

SDTM+

- SUPPQUAL datasets allows merging non-standard variables to their parent domains
  - QNAM – Name (8 chars)
  - QLABEL – Label (40 chars)
- SDTM+ structure
- CDISC team new proposal
  - Keep non-standard variables in their parent domains
  - **Pros:** simplifies review process
  - **Cons:** may encourage excessive use of non-standard variables with potential deviation from SDTM compliance

SDTM NON-STANDARD VARIABLES

- SDTM-IG “Appendix C2: Supplemental Qualifiers Name Codes”

QNAM	QLABEL	Applicable Domains
AESOSP	Other Medically Important SAE	AE
AETRTEM	Treatment Emergent Flag	AE
--CLSIG	Clinically Significant	Findings
--REAS	Reason	All general observation classes

- Therapeutic Area User Guides (TAUG)
  - *Provisional* standards waiting for adding new variables in SDTM Model
  - Pharmacogenomics/Genetics (PGx) TAUG example
  - ‘SDTM NSV Registry’ page on CDISC Wiki to keep track of non-standard variables

BEST PRACTICE ON CREATION OF NSV

- QNAM should start with <domain name> prefix
  - like names of standard variables in domains
  - For example, AETRTEM, AESOTH, EGCLSIG, etc.
  - Exceptions are sponsor-specific variables which are utilized across domains like VISIT or USUBJID.
- QNAM values cannot use variable names which already exist in SDTM model
- Utilization of SUPPQUAL variables should be consistent within a study and within a submission
- Users should try to use existing non-standard variables from CDISC documentation

INDUSTRY IMPLEMENTATION OF SUPPQUALS

- Driven by company and study-specific needs
- So far, there are no industry-wide metrics
  - To understand implementation of non-standard variables
  - To discover potential problems
  - To help developing CDISC SDTM standards
- P21 pilot project
  - To test methodology
  - To test potential use of industry metrics for improving standard management processes including data validation

INDUSTRY METRICS: IMPLEMENTATION OF SUPPQUALS

Pilot study

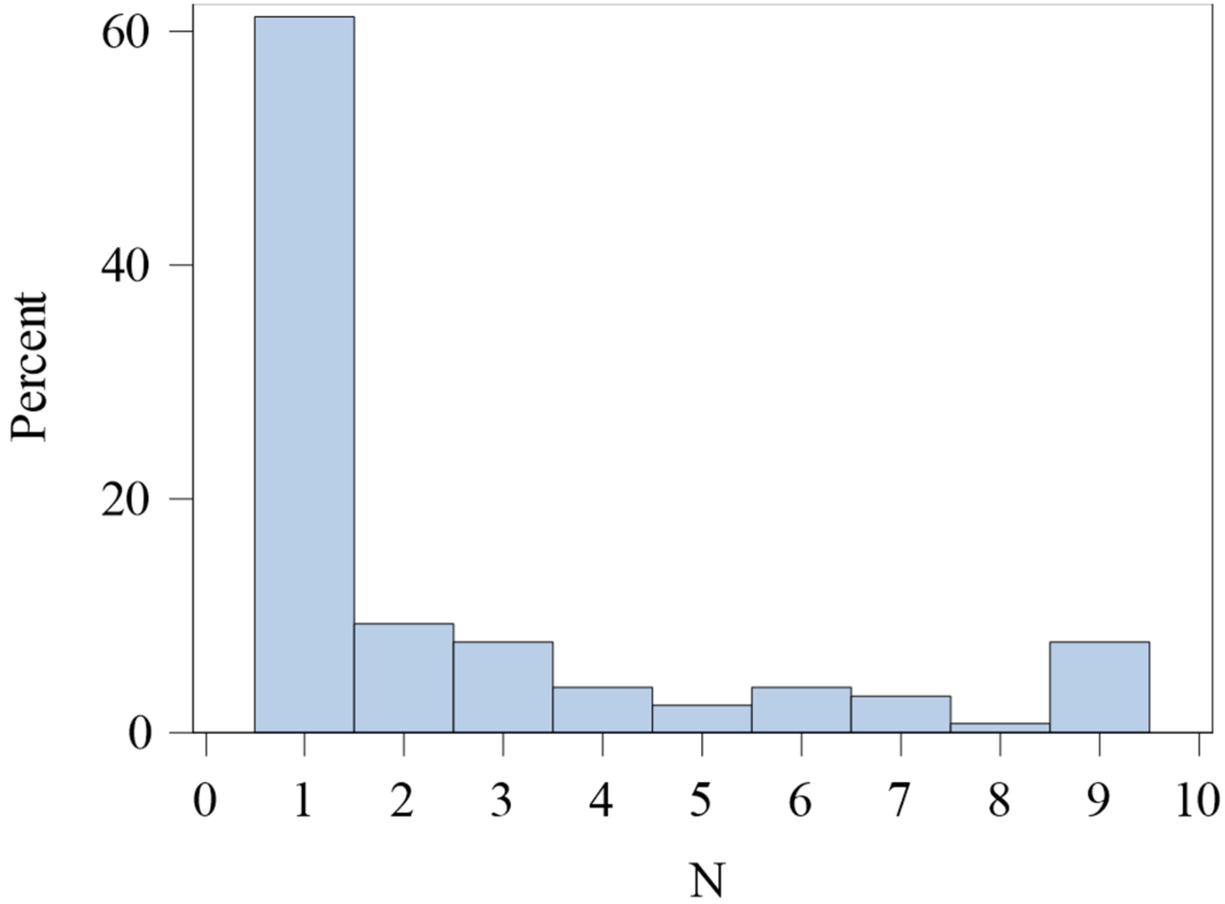
P21

METHODOLOGY

- Metrics collected by Pinnacle 21 Enterprise
- Finalized studies
- Diversity of collected data
  - One sponsor may be represented by up to 3 studies within each phase and each therapeutic area
  - For example, it could be up to 3 phase II studies with different indications like Oncology, Antiviral and Dermatology
- Content of collected data
  - List of SUPPQUAL variables
  - De-identified sponsor and study IDs
  - Study phase, start date, version of SDTM
  - Indication collapsed into Oncology/Non-oncology

ANALYZED METRICS

- 124 sponsors
- 325 studies
  - 28% Oncology
  - 82% started in 2015 or later
  - 76% based on SDTM-IG 3.2
  - 19% based on SDTM-IG 3.1.3
- 27,023 QNAMs
  - Sponsor/Study/Dataset/QNAM

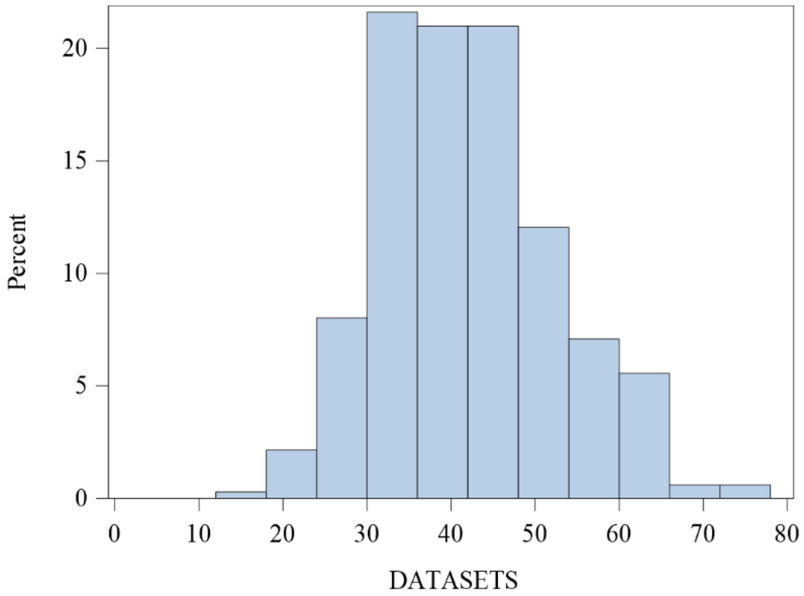


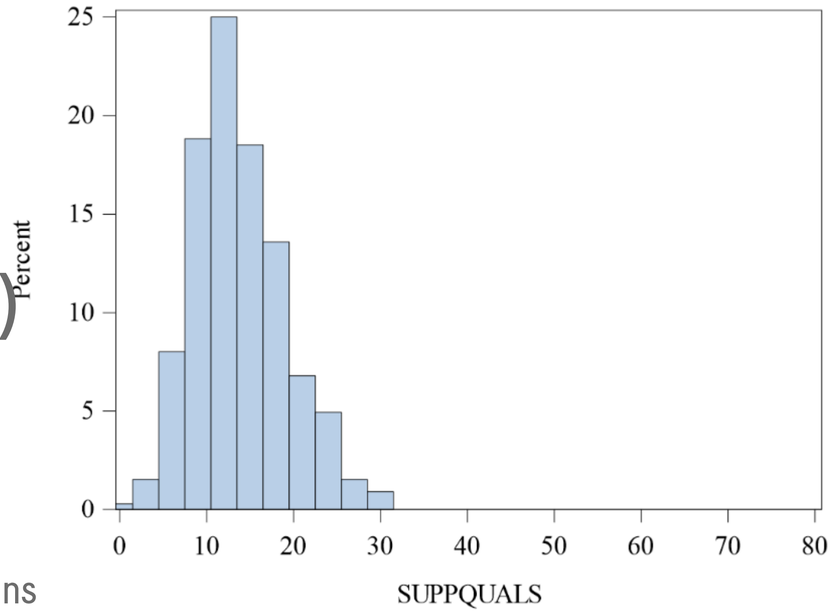
Distribution of number of sample studies (N) per sponsor

STATISTICS

- Datasets
  - 14-75 per study
  - Mean 41.5 (median 41)
    - Oncology – 50
    - Non-oncology – 38.2
- SUPPQAUL datasets
  - 0-30 per study
  - Mean 13.7 (median 13)
    - Oncology – 17.5 (72%*)
    - Non-oncology – 12.2 (67%)
- Non-standard variables (NSV)
  - 1-618 per study
  - Mean 84 (median 65.5)
  - Mean of NSV per dataset 5.8

* Number of SUPPQUAL datasets / number of qualifies domains





MOST COMMON SUPPQUAL DATASETS (1-13)

Dataset	Number of studies with SUPPxx dataset	% (all studies)	% (non-oncology studies)	% (oncology studies)
SUPPAE	313	96.3	94.9	100.0
SUPPCM	298	91.7	89.4	97.8
SUPPDM	289	88.9	87.7	92.1
SUPPLB	251	77.2	74.2	85.4
SUPPDS	246	75.7	71.6	86.5
SUPPEG	222	68.3	64.4	78.7
SUPPMH	206	63.4	62.3	66.3
SUPPEX	205	63.1	56.8	79.8
SUPPDV	190	58.5	58.9	57.3
SUPPPR	141	43.4	33.1	70.8
SUPPPC	139	42.8	39.8	50.6
SUPPPE	120	36.9	39.8	29.2
SUPPVS	115	35.4	35.2	36.0

MOST COMMON SUPPQUAL DATASETS (14-26)

Dataset	Number of studies with SUPPxx dataset	% (all studies)	% (non-oncology studies)	% (oncology studies)
SUPPEC	112	34.5	28.4	50.6
SUPPFA	106	32.6	29.2	41.6
SUPPQS	105	32.3	33.1	30.3
SUPPDA	86	26.5	27.5	23.6
SUPPSU	69	21.2	21.2	21.3
SUPPTU	68	20.9	0	76.4
SUPPIE	67	20.6	18.6	25.8
SUPPSV	61	18.8	20.8	13.5
SUPPCE	58	17.8	15.7	23.6
SUPPRS	48	14.8	1.7	49.4
SUPPHO	47	14.5	12.3	20.2
SUPPTR	46	14.2	0	51.7
SUPPSS	42	12.9	5.1	33.7

15 MOST COMMON QNAM VALUES

QNAM	QLABEL	Number of studies	% (all studies)
AETRTEM	Treatment Emergent Flag	204	63.0
EGCLSIG	Clinically Significant	148	45.7
RACEOTH	Race, Other	129	39.8
DVTERM1	Protocol Deviation Term 1	98	30.2
LBCLSIG	Clinically Significant	64	19.8
PECLSIG	Clinical Significance	62	19.1
PRLLT	Lowest Level Term	61	18.8
PRHLGT	High Level Group Term	60	18.5
PRHLT	High Level Term	60	18.5
DVTERM2	Protocol Deviation Term 2	57	17.6
PRHLGTCD	High Level Group Term Code	55	17.0
PRHLTCD	High Level Term Code	55	17.0
PRPTCD	Preferred Term Code	53	16.4
PRLLTCD	Lowest Level Term Code	52	16.0
ATC3	ATC Level 3 Text	49	15.1

MOST COMMON QNAM CONTINUED (16-30)

CMDECOD1	Standardized Medication Name 1	45	13.9
ATC2	ATC Level 2 Text	44	13.6
CMATC2	ATC2	43	13.3
CMATC3	ATC3	42	13.0
PROTVER	Protocol Version	42	13.0
ATC1	ATC Level 1 Text	41	12.7
PRSOC	System Organ Class	41	12.7
RACE1	Race 1	41	12.7
CMATC1	ATC1	40	12.3
CMDECOD2	Standardized Medication Name 2	39	12.0
CMCLAS1	Medication Class 1	38	11.7
RACE2	Race 2	38	11.7
CMATC4	ATC4	37	11.4
COHORT	Cohort	37	11.4
CMCLAS2	Medication Class 2	36	11.1

AE TREATMENT EMERGENT FLAG

- Requested by both FDA and PMDA
  - Special validation rules
- Only 63% studies are compliant
  - Often AETRTEM Flag is populated only in ADaM
- Most automated review tools use SDTM data
  - Predicted standardized structure

CLINICALLY SIGNIFICANT

- The second the most common SUPPQUAL
  - 46% of all studies

Dataset	QNAM	QLABEL	N of studies with QNAM	% (all studies)	N of studies with Dataset	% (studies with Dataset)
SUPPEG	EGCLSIG	Clinically Significant	148	45.7	224	66.1
SUPPPE	PECLSIG	Clinically Significant	62	19.1	121	51.2
SUPPLB	LBCLSIG	Clinically Significant	64	19.8	253	25.3
SUPPVS	VSCLSIG	Clinically Significant	24	7.4	116	20.7

- However, information in this table is not very accurate
  - Lack of CDISC conformance during the industry implementation of non-standard variables

EXAMPLES OF IMPLEMENTATIONS

QNAM	QLABEL
CLINSIG	CLINICALLY SIGNIFICANT
EGABN	Abnormal Clinically Significant
EGCHG	Clinically Significant Chg. from Screen?
EGCHGCS	ECG Changes Clinically Significant
EGCLIG	Clinical Significance
EGCLISG	Clinically Significant
EGCLISIG	Clinically Significant Result
EGCLSIG	Clinically Significant
EGCLSIG1	Abnormality 1 Clinically Significant
EGCLSIG1	Clinically Significant 1
EGCLSIG2	Clinically Significant 2
EGCS	ECG Clinically Significant

- Increases actual use of Clinically Significant Flag in SUPPEG
  - 50% of all studies
  - 73% of studies with SUPPEG dataset

LABELS FOR QNAM=EGCLSIG

QLABEL	N	QLABEL	N	QLABEL	N
Abnormality Clinically Significant	1	Clinically Significant for EG	1	EG: If Abnormal, is it Clin Significant?	1
Abnrml Interpretation Clin Significant?	1	Clinically Significant?	2	EGCLSIG	1
CLINICAL SIGNIFICANCE	1	Clinically significance	1	If Abnormal and clin. signf., specify	1
CLINICALLY SIGNIFICANT	1	Clinically significant	4	If Abnormal, Clinical Significance	1
CLINICALLY SIGNIFICANT OR NOT	2	ECG Res. Abnormal Clinically Significant	1	If abnormal, clinically significant?	1
CS/NCS	1	ECG Res. clinically significant	1	Interpretation Clinically Significant	2
Clinical Significance	19	ECG Result Abnormal Clin. Significant	1	Is the Result Clinically Significant?	1
Clinical Significance Flag	1	ECG Result clinically significant	2	SIGNIFICANCE OF ABNORMALITY	1
Clinically Siginificant	1	ECG Test Result Clinically Significant	1	Was Abnormality Clinically Significant?	2
Clinically Significant	94	EG Clinically Significant, Specify	1	Was Finding Clinically Significant?	1
Clinically Significant Abnormality	1	EG Clinically significant?	1		

- Correct implementation is only in 64% cases
- Questions about correct utilization of EGCLSIG
  - '*EG Clinically Significant, **Specify***'
  - *'If Abnormal and clin. signf., **specify***'

NON-STANDARD IMPLEMENTATION

- ‘Interpretation’ variable in SUPPEG
  - ‘*Normal*’
  - ‘*Abnormal, not clinically significant*’
  - ‘*Abnormal, clinically significant*’
- Mix of two potentially different types of information
  - Normal/Abnormal Result Interpretation
  - Clinically Significance Flag
- CDISC is planning to add --CLSIG (Clinically Significant) variable to SDTM model

WHO DRUG CODING

- WHO Drug dictionary has been added to FDA Data Standards Catalog
- No special SDTM variables for WHO Drug
- SDTM-IG suggests utilization of SUPPCM
  - However, no details or examples are provided

- Huge diversity of implementation by the industry
  - 298 studies with SUPPCM
  - 1,023 different QNAM/QLABEL or 667 unique QNAM for WHO Drug ATC coding
  - 128 variations of QNAM values which include text '*ATC1*'
    - 194 QNAM/QLABEL

EXAMPLES OF ATC1 VARIABLES

QNAM	QLABEL
ATC1	ATC Level 1 Text
ATC1_C	ATC 1 Class Code
ATC1_T	ATC 1 Class Text
ATC1C	ATC Level 1 Code
ATC1C_1	ATC1 Code 1
ATC1C_15	ATC1 Code 15
ATC1C03	ATC 1 Code
ATC1CD	ATC Level 1 Code
ATC1CODE	ATC 1 CODE
ATC1M10	ATC Level 1 Term for 10th Multiple Term
ATC1P	ATC Level 1 Term for Primary Term

QNAM	QLABEL
ATC1P15	WHO-DDE ATC1-MAIN GROUP-15
ATC1T	ATC Level 1 Text
ATC1T	ATC 1 NAME
ATC1TERM	ATC 1 NAME
ATC1TEXT	Level 1 Term
ATC1TM	ATC1 TERM
CMATC1	ATC 1 Term
CMATC115	ATC Level1 2015Jun
CMATC1TX	ATC 1 Text
ORATC1	Original ATC Level 1 Term
WHOATC1	WHO-DDE ATC1-MAIN GROUP

EXAMPLES OF QLABEL FOR QNAM='CMATC1'

QLABEL
ATC Chemical Subgroup 1st Level
ATC Classification Level 1
ATC Level 1
ATC Level 1 Code
ATC Level 1 Decode
ATC Level 1 Term
ATC Level 1 Text
ATC1 Name

QLABEL
ATC1 Term
ATC1 TERM
Comed ATC1 Term
Level 1 ATC
Medication Class 1
Medication/Therapyatc
WHO-DD, ATC Code, Level 1
WHODrug ATC1

MAJOR VIOLATIONS OF SDTM CONFORMANCE

Looking across 325 analyzed studies you can find many possible examples of violation of CDISC SDTM conformance for implementation of SUPPQUAL datasets.

COMMENTS IN SUPPQUALS

- Widespread violation

Dataset	N	% (all studies)	% (studies with Dataset)
SUPPAE	1	0.3	0.3
SUPPCM	4	1.2	1.3
SUPPDM	0	0.0	0.0
SUPPDV	8	2.5	4.2
SUPPEX	3	0.9	1.5
SUPPLB	51	15.7	20.2
SUPPPC	22	6.8	15.8
SUPPVS	0	0.0	0.0

TIMING INFORMATION

- >1000 SUPPQUAL variables with Timing info
- 966 unique QNAM with QLABEL which includes text '*date*'
  - *Date of Best Response, Subject Date of Birth, Data Entry Date, Last Contact Date, Report Date, Randomization Date,* etc.
- 492 unique QNAM with QLABEL which include text '*time*'
  - *Randomization Time, Time of onset, Time of blood draw, Actual Time*, etc.
  - some of these non-standard variables are overlapped with 'date' variables
  - few of them do not represent Timing info (e.g., *Ongoing at Time of Death*)
- Visit variables
  - In datasets like SUPPEX, SUPPDV, SUPPCO, SUPPTR, SUPPLB, SUPPPC, etc.

ADDITIONAL RESULTS AND UNITS

- Normal Range information
  - Rare cases
  - Examples:
    - SUPPLB.AGE_HIGH (*Normal Range Upper Limit-Age*), SUPPEG.EGORNRHI (*UPPER NORMAL RANGE VALUE*), SUPPLB.SINORMHI (*SI upper limit of normal range*), SUPPVS.SYSBPHI (*Sys BP Normal Range High*), etc.
- Original, previous or supplemental results in Conventional or SI units
  - Usually in SUPPLB datasets
  - Examples:
    - SIRESN (*SI Numeric Result*), CNVRESC (*Conventional Text Result*), LBORRES4 (*Result or Finding in Original Units*), LBSTRSCN (*Char. Result/Finding in Std Format (N)*), PSTRESC (*Previous Character Result in Std Format*), etc.

NOT APPLICABLE INFO IN SUPPDM

QNAM	QLABEL
BLOODONR	Blood Donor
DEMEDU	6.Highest level of edu?
DEMMAR	4.Current marital status
DMBLWT	Baseline Weight g
DMEMPLO	Current employment situation
DTH_D	Day of Death
EXCONC	Final Study Drug Concentration
IEORRES	Did subject meet eligibility criteria?
INITDOSU	Dose Units
P85BMI	85th Percentile BMI (kg/m2)

INVALID SUPPQUAL DATASETS

- 61 (19%) studies with SUPPSV
  - All information collected on Subject Visit CRF
    - SUBJID1 (*Subject Identifier 1 for the study*), TVISYN (*Is This a Treatment Visit?*), SVASSESS (*Assessments Performed*), SVUPDES1 (*Description of Unplanned Visit*), VISLB (*Lab Collection*), DOVDTC (*Date of Visit*), OTHERSP (*If Other, specify*), etc.

- 11 (3%) studies with SUPPCO

EXAMPLES OF OTHER VIOLATIONS

- Data management and tracking info
  - Not applicable for regulatory submissions
    - One study with 618 non-standard variables which represent raw data collected in EDC
      - SUPPAE.AESERN (*Serious Event (N)*), SUPPAE.AEST_Y (*Start Year of Adverse Event*), SUPPAE. EPOCHN (*Epoch (N)*) SUPPDM.RACEN (*Race (N)*), etc.

- Only 44% of non-standard variables have a name with prefix corresponding to domain value
  - Like SUPP**AE**.**AE**TRTEM, SUPP**EG**.**EG**CLSIG, SUPP**CM**.**CM**ATC1, SUPP**XY**.**XY**ABCDEF

UTILIZATION OF CDISC TAUGS

NON-STANDARD VARIABLES FROM TAUGS

- ‘SDTM NSV Registry’ page on CDISC Wiki
  - 173 variables used as new non-standard variables across 40+ existing CDISC TAUGs
  - 142 unique variables

- Only 24 (17%) of CDISC NSV were found in 324 analyzed studies
  - Low utilization of existing CDISC TAUG by the industry

CONSISTENCY WITH CDISC

- In many cases NSV in SUPPQUALs are not consistent with CDISC

- CDISC: --SPEC is '*Specimen Type*'
- Industry:
  - --SPEC NSV was implemented in 24 SUPPQUAL datasets
  - 11 of them have different interpretation of --SPEC variable
    - *'Other, **Spec**ify',*
    - *'Other Symptom'*
    - *'AE of **Spec**ial Interest'*
    - *'Disposition **Spec**ifications'*
    - *'AE **Spec**ify'*
    - *'Abnormal, **Spec**ify'*

CONCLUSION

This study was run as a pilot to understand the potential use of industry metrics for improving standards management practices and to test methodology

SUGGESTIONS FOR ADDITIONAL ANALYSIS

- How consistent is the implementation of SUPPQUAL within each company?
  - We saw both cases
    - Consistent across organization
    - Inconsistent within study
- Is there any correlation of SUPPQUAL implementation with version of SDTM-IG, Data Fitness Score, etc.?
- Implementation of CDISC TAUGs
- Implementation of non-standard domains

HOW CAN WE IMPROVE STANDARDS MANAGEMENT PRACTICES?

- Existing standards and regulatory guidance documents are underutilized or ignored
  - AETRTEM, SDTM-IG Appendix C2
  - Additional educational efforts in promotion of data standards and regulatory requirements are expected
- Some information is utilized in almost every study but is not represented by standard SDTM variables yet
  - --CLSIG, WHO Drug coding, MedDRA in PR domain, etc.
- There is still a common practice of misuse and incorrect mapping of collected data into SUPPQUAL datasets
  - Education efforts are expected to promote good SDTM mapping practices
  - New validation rules may also help

CONCLUSION

- Collection and analysis of the industry implementation metrics can
  - Help identify global implementation issues
  - Help with their eventual resolution

Sergiy@Pinnacle21.com


THANK YOU:)