

# ADaM Validation Update from OpenCDISC

Michael DiGiantomasso  
Webinar  
June 3<sup>rd</sup> and 4<sup>th</sup>, 2015

# Presenter

## Mike DiGiantomasso



Partner & Principal Consultant at Pinnacle 21  
Data Fitness Analyst for FDA's JumpStart Service  
CDISC ADaM Validation sub-team member

# Stay Connected


## Get updates about our events

 [linkedin.com/company/pinnacle-21](https://linkedin.com/company/pinnacle-21)

 [twitter.com/pinnacle\\_21](https://twitter.com/pinnacle_21)

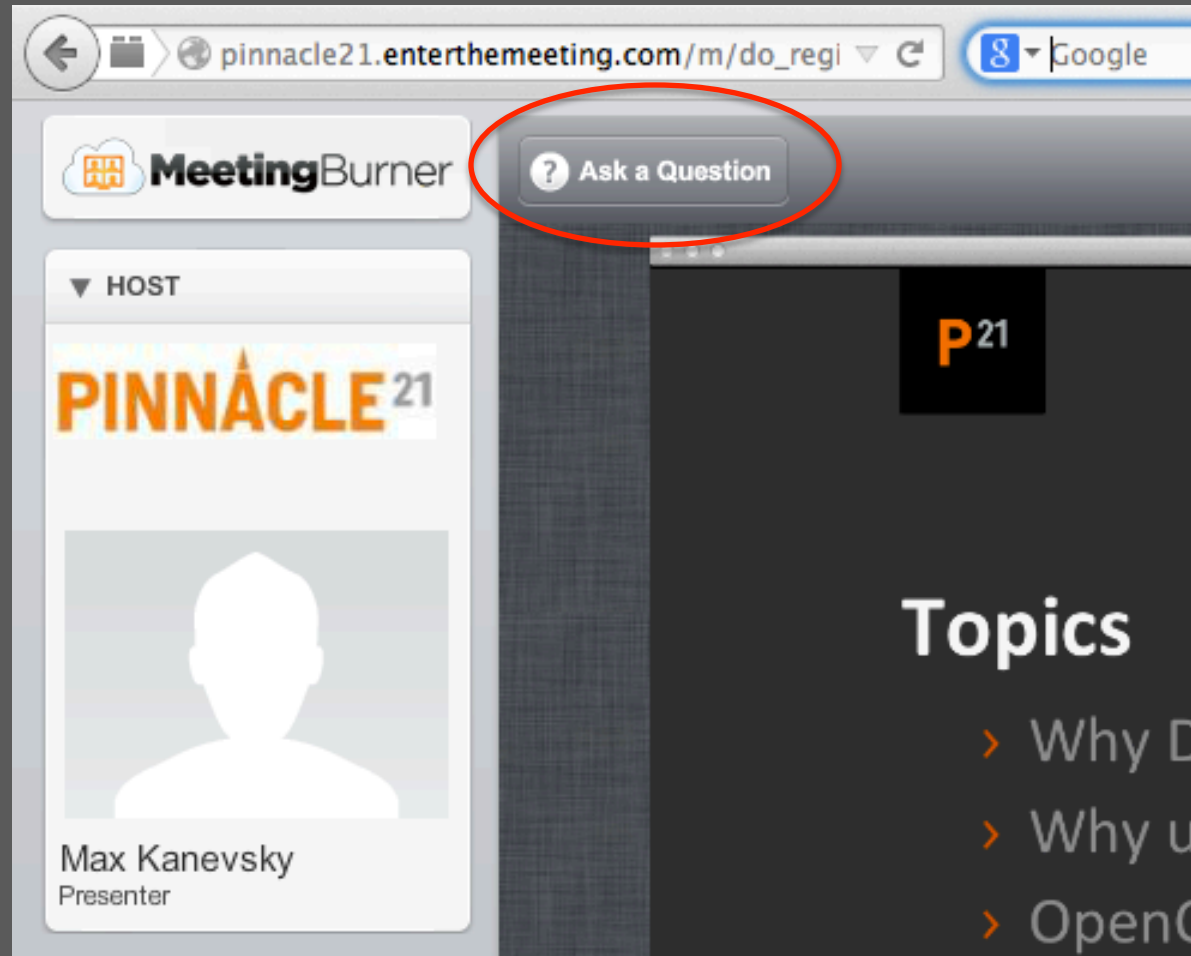
 [facebook.com/pinnacle21](https://facebook.com/pinnacle21)

 [pinnacle21.net/blog](https://pinnacle21.net/blog)

 [opencdisc.org](https://opencdisc.org)

# Q & A

- › click the “Ask a Question” button
- › Questions will be answered at the end



# Topics

- › ADaM Validation Rules <sup>(4)</sup>
  - › Brief History & Rule Basis
  - › Metrics & What's new
- › Datasets recognized by OpenCDISC <sup>(3)</sup>
- › ADaM Validation categories - **deep dive** <sup>(11)</sup>
- › Adding sponsor-specific CT and VLM checks <sup>(4)</sup>
  - › **Quick demo**
- › Common ADaM issues in submission data <sup>(9)</sup>
- › Rules for the upcoming release of ADaM IG v1.1 <sup>(1)</sup>
- › *Future rule considerations* <sup>(1)</sup>
- › *Questions*

Addendum for further reading

# CDISC ADaM Publication History: a Basis for Rules

Document	Version	Published	Note
ADaM	2.0	2006-08-11	
ADaM	2.1	2009-12-17	
ADaM IG	1.0	2009-12-17	Based on ADaM 2.1
ADaM BDS-TTE IG	1.0	2012-05-08	Based on ADaM IG 1.1
ADaM ADAE IG	1.0	2012-05-10	Occurrence Data Structure
* Analysis Results Metadata	1.0	2015-01-27	Extension for Define.xml 2.0
** ADaM Validation Checks	1.3	2015-03-16	304 Rules, 75 new for ADAE and TTE
ADaM IG	1.1	2015-05-23	Draft for Public Comment

'06,09,12,15....18? – Hopefully sooner 😊

\*\* First 3 version of ADaM Validation Checks created between 2010 and 2012

\* Not evaluated:

- Analysis Results Metadata
- ADaM Examples in Commonly Used Statistical Analysis Methods

# OpenCDISC ADaM Rules History

OpenCDISC Release	ADaM 1.0 Spec	Packaged with Engine	Covers CDISC ADaM VC
2011-01-24	1.0	1.2	1.0
2012-03-29	1.1	1.3	1.1
2013-03-14	1.2	1.4	1.1
2013-09-12	1.3	1.5	1.2
2014-12-11	1.4	2.0.0	1.2
2015-03-12	1.5	2.0.1	1.2+
2015-03-18	1.6	OpenCDISC ENTERPRISE	1.3

Most free users fall in this range

Includes CDISC ADaM VC 1.3 rules for TTE, ADAE and all other enhancements mentioned today

\* Community typically lags behind Enterprise

# ADaM Validation Rule Metrics

## Business/Academic Rules

CDISC ADaM VC Version	Rules total	Rules Deleted	Rules Added	Rules Modified	Text Changes	Date
1.0	180					2010-09-20
1.1	174	9	3	5	30	2011-01-21
1.2	233	2	61	0	13	2012-07-05
1.3	304	4	75	0	27	2015-03-16

## Implementation Rules

CDISC ADaM VC Version	ADaM Rules	OpenCDISC ADaM Ver.	OpenCDISC Rules	ADSL *	BDS *	ADAE *	New/Enhanced
1.3	304	1.6	257	89	178	64	25

Outside of CDISC \*\*  
e.g., zz is not in [01-99]  
for ANLzzFL/FN

\* Some rules apply to multiple datasets

\*\* All enhancements have been suggested to CDISC ADaM Validation Subteam. Listed in addendum



# CDISC to OpenCDISC Mapping Cardinality

Why “less” rules ?

CDISC Business Rules 304	OpenCDISC Implementation 257
1	1
Many	1
1	Many

## OpenCDISC ENTERPRISE

Home > Design > Validation Specs > ADaM v1.0

### ADaM v1.0

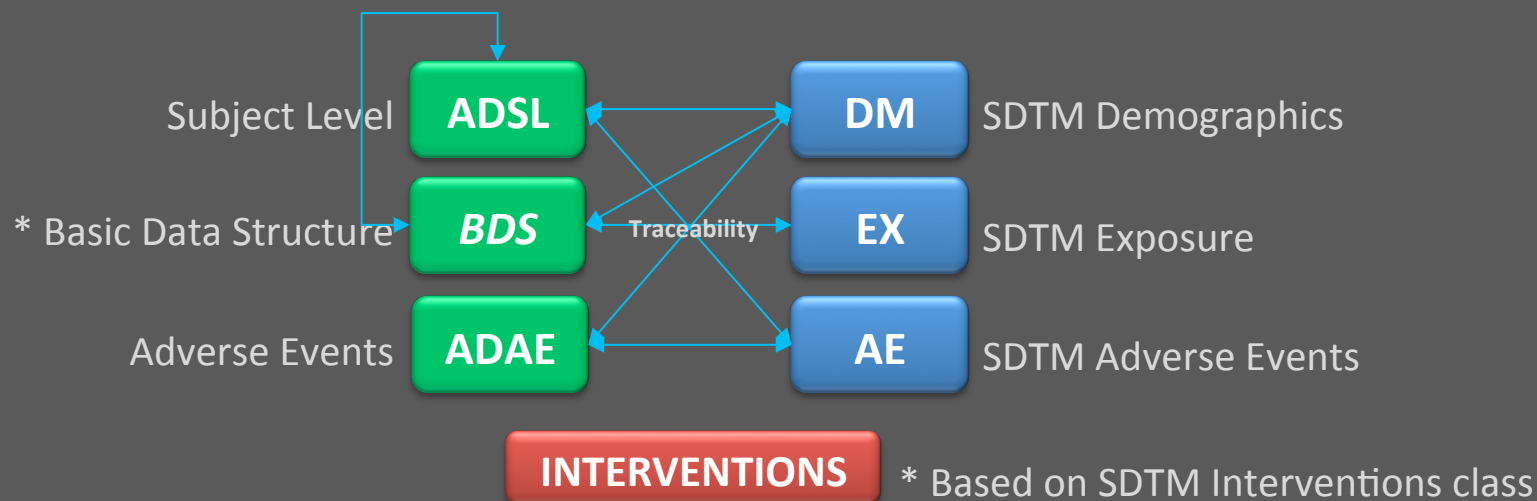
Properties Domains Variables Rules

Remove Selected

Search

ID	Message	FDA SDTM	CDISC
AD0303	illegal variable name: zz is not in [01-99] for CQzzNAM		303
AD0304	Only some of these variables are present and populated: SMQzzNAM,SMQzzCD,SMQzzSC		304,305,306,307,308,309
AD0314	illegal variable name: y is not in [1-9] for original or prior MedDRA coding variables of form *ORGy		314,315,316,317,318,319
AD0037	Inconsistent value for *GRyN		37,286,292,298
AD0038	Inconsistent value for *GRy		38,285,291,297
CT2001	Variable value not found in non-extensible codelist	FDAC340	39,40

# Datasets recognized by OpenCDISC



Dataset classification	Dataset contains 1 or more variables....
* BDS	PARAMCD, PARAM AVAL, AVALC ADT, ASTDT, ASTDTM CNSR, CNSDTDSC, EVNTDESC
* INTERVENTIONS	__TRT, CMTRT EXTRT, ECTRT SUTRT, PRTRT SGTRT, XCTRT

# Datasets recognized by OpenCDISC (TTE)

## BDS

### Time To Event (TTE) Analysis

- › BDS dataset can have both TTE and non-TTE data mixed in.
- › TTE datasets have no standard naming conventions.
- › TTE is a BDS dataset with **9 rules** (focused on **CNSR**), which run on **all** BDS datasets when...
  - › **datasets** with TTE variables are *present*
  - › **records** with CNSR (censor) are *populated*

*OpenCDISC	CDISC	Functional Group	Machine-Testable Failure Criteria
AD0250	250	Present	EVNTDESC is present but <b>CNSR</b> is not present
AD0250	251	Present	CNSDTDSC is present but <b>CNSR</b> is not present
AD0170	170	Present	STARTDT is present but <b>CNSR</b> is not present
AD0245	245	Present	<b>CNSR</b> is present but STARTDT is not present
AD0245	246	Present	<b>CNSR</b> is present but ADT and ADTM are not present
AD0247	247	Populated	<b>CNSR</b> is populated but STARTDT is not populated
AD0247	248	Populated	<b>CNSR</b> is populated but ADT and ADTM are not populated
AD0247	249	Populated	<b>CNSR</b> is populated but AVAL is not populated
AD0169	169	Valid Value	<b>CNSR</b> is populated and the value is not an integer

\* Many to 1 cardinality. 9 CDISC rules translate to 5 OpenCDISC rules. AD0170 and AD0245 could be further combined into 1 rule

# Datasets recognized by OpenCDISC (Interv)

## INTERVENTIONS

- › no standard naming conventions. Some Examples seen in industry:
  - › ADCM – Concomitant meds
  - › ADCP – Concomitant Procedures
  - › ADEX - Exposure
  - › ADSURG – Surgery
  
- › Recognition of these domains eliminates
  - › MISSING\_CONFIG warning
  - › false positive SD0061 – domain referenced in define.xml was not submitted
  - › inability to assign rules (although none exist at the moment)
  
- › ADaM IG 1.1 states:
  - › “the ADaM team is working on specification for a more general structure supporting analysis of incidence data, such as concomitant medications and medical history”

# ADaM Rule Categories (10)

Consistency Pairs	
1-Way-Map	6
2-Way-Map	54

Equality of variables	
baseline	3
dates	4

Calculations	
arithmetic	5
missing operands	6
division by zero	4

Illegal Concepts	
value	5
value-order	5
variable-label	2
variable-name	25

Terminology	
basic	3
flags	14
controlled	6

Define.xml	
value Level Metadata	4
dataset metadata, codelists	6

Presence	
conditional	32
unconditional	6

Traceability	
ADaM-to-ADaM	21
ADaM-to-SDTM	17

Other	
logical duplicates	5
tech-conformance	6

Population	
conditional	8
conditional group	6
unconditional	4

- › CDISC ADaM Validation Team is working with OpenCDISC to publish similar categories
- › Some rules can be categorized into multiple buckets. Counts (sum to 257) based on the primary category
- › Categorization useful for implementation and book-keeping.
  - › In some cases may be useful for impact analysis (i.e., if all 80% of issues stem from calculations vs variable-labels, that could indicate a major problem)

# Consistency Pairs

- › 1-Way-Map: Ensures consistent categorization of data.
- › AD0221: Inconsistent value for CHGCAT1 (for a unique STUDY, PARAMCD, CHG)

CHG	CHGCAT1
0	LOW
0	LOW
29	LOW
30	MEDIUM
30	MEDIUM
60	HIGH
70	HIGH
70	VERY HIGH

Value of 70 can't have 2 categories

- › 2-Way-Map: Ensures consistent variable pair values.
  - › AD0141 Inconsistent value for PARAM within a unique PARAMCD
  - › AD0142 Inconsistent value for PARAMCD within a unique PARAM

PARAM	PARAMCD
Migraine Pain	MIGRAINE
Migraine Pain	MIGRAINE
Headache	MIGRAINE
Migraine Pain	Migraine

# Illegal Concepts

- › variable and label naming conventions \*\*
  - › AD1018: **zz** is not in **[01-99]** for ANLzzFL/FN
  - › AD0065: **xx** is not in **[01-99]** and **y** is not in **[1-9]** for TRxxPGy(N)
  - › AD0018: Variable label mismatch between dataset and ADaM standard
- › data values
  - › AD0046: \*DY = 0
  - › AD0169: CNSR is not an integer  $\geq 0$
- › data value order: *Low > High , Start > End*
  - › AD0162: \*LO > \*HI
  - › AD0121: \*SDT > \*EDT
  - › AD0122: \*SDTM > \*EDTM
  - › AD0098: \*SDY > \*EDY
  - › AD0099: \*STDY > \*ENDY

# Presence

- › Some concept is either conditionally or unconditionally present in the submission
- › typically requires a **primary variable** if its corresponding **secondary qualifying variable** is present.
- › **Variables** (*requiring or prohibiting*)
  - › AD0198: Neither AVAL nor AVALC are present in dataset
  - › AD0061: SDTM.EX is present but neither ADSL TRTSDT nor TRTSDTM are present
  - › AD0245: CNSR is present but not all of STARTDT, ADT and ADTM are present
  - › AD0252: Variable present in dataset is *prohibited* for use in ADaM
- › **Records**
  - › AD0127: BASE or BASEC is populated for a unique USUBJID,PARAMCD but no baseline record exists
- › **Datasets**
  - › AD0001: Missing ADSL dataset



# Population

- › A variable is either conditionally or unconditionally populated
- › typically requires a **primary variable** to be populated if its corresponding **secondary qualifying variable** is populated.
- › **Single Variables** *(requiring or prohibiting)*
  - › AD0137: CRITyFL is populated but CRITy is not populated
  - › AD225A: BASE = 0 but PCHG is populated (division by zero)
- › **Variables Groups**
  - › AD0304: Only some of these variables are present and populated: **SMQzzNAM,SMQzzCD,SMQzzSC**
  - › AD0247: CNSR is populated but not all of **STARTDT, ADT, ADTM** and **AVAL** are populated
  - › AD0159: AWTDIFF is populated but **AWTARGET** and/or one of [**ADY,ARELTM**] are not both present and populated

# Equality of Variables

values are often duplicated in the ADSL and BDS datasets.

- › Two variables in the same or different datasets represent the same date

AD0083	ADSL.TRTSDT != BDS.TR01SDT
AD0084	ADSL.TRTEDT != Maximum BDS.TRxxEDT on this record
AD0091	BDS.TRTP != at least one ADSL.TRTxxP
AD0244	BDS.TRTA != at least one ADSL.TRTxxA
AD0107	BDS.APERIOD xx has an APERSDT <> ADSL.APxxSDT
AD0100	BDS.APERIOD xx has an APEREDT <> ADSL.APxxEDT
AD0044	*TM and its corresponding *DTM variable have different time values
AD0045	*DT and its corresponding *DTM variable have different date values

- › Baseline Values should be the same as the analysis value for a baseline

AD0165	ABLFL = Y, but BTOXGR != ATOXGR
AD0168	ABLFL = Y, but BNRIND != ANRIND
AD0152	ABLFL = Y, but BASE != AVAL

# Terminology

- › **Flags** - Variables must have one of these values, where null is usually allowed
  - › Y
  - › 1
  - › 0,1
  - › N,Y
  
- › **Basic** - The IG specifies a list that is not controlled by NCI/CDISC
  - › AD0312: SMQzzSC not in [BROAD, NARROW, null]
  - › AD0313: SMQzzSCN not in [1, 2, null]
  - › AD0279: AESEVN,ASEVN not in [1, 2, 3]
  
- › **Controlled** - CT2001 to CT2006 to check consistency with NCI CDISC CT
  - › CL.C81224.DTYPE – *derivation type*
  - › CL.C81225.PARAMTYP – *parameter type*
  - › CL.C81223.DATEFL – *date imputation flag*
  - › CL.C81226.TIMEFL – *time imputation flag*

# Traceability

## › ADaM-to-SDTM

AD0053	ADaM USUBJID value does not exist in the SDTM DM domain
AD0204	For the same USUBJID, ADSL.AGE <> DM.AGE
AD0205	For the same USUBJID, ADSL.AGEU <> DM.AGEU
AD0210	For the same USUBJID, ADSL.ARM <> DM.ARM
AD0207	For the same USUBJID, ADSL.RACE <> DM.RACE
AD0206	For the same USUBJID, ADSL.SEX <> DM.SEX
AD0209	For the same USUBJID, ADSL.SITEID <> DM.SITED
AD0208	For the same USUBJID, ADSL.SUBJID <> DM.SUBJID
AD0258	Record key from ADaM ADAE is not traceable to SDTM AE ( <i>extra ADAE recs</i> )
AD0253	Record key from SDTM AE is not traceable to ADaM ADAE ( <i>not enough ADAE recs</i> )

## › ADaM-to-ADaM

### › Typical dataset cross-reference

AD0256	BDS or ADAE USUBJID value does not exist in the ADaM ADSL domain
AD0102	BDS.APERIOD xx does not have a corresponding ADSL.TRTxxP variable
AD0100	BDS.APERIOD xx has an APEREDT that does not equal ADSL.APxxEDT

### › Atypical definition, for Subject Treatment status (BDS to ADSL)

<b>off treatment</b> , but analysis date is	in treatment period (ADSL.TRTSDT <= BDS.ADT <=ADSL.TRTEDT)
(last value) <b>on treatment</b> , but analysis date is <b>NOT</b>	in treatment period !(ADSL.TRTSDT <= BDS.ADT <=ADSL.TRTEDT)

Rule	Variable Key	Treatment status	Treatment Period (ADSL.TRTSDT <= BDS.ADT <=ADSL.TRTEDT)
AD0173	BDS.ONTRTFL != Y	<b>off</b>	TRUE = Fail
AD0171,172	BDS.ONTRTFL = Y	<b>on</b>	FALSE = Fail
AD0174,179	BDS.LVOTFL = Y	<b>on</b>	FALSE = Fail

# Calculations

## › Arithmetic *(when all operands are present)*

AD0223	CHG        !=    AVAL - BASE
AD0132	R2BASE !=    AVAL / BASE
AD0225	PCHG       != (AVAL - BASE) / BASE * 100
AD0133	R2AyLO !=    AVAL / AyLO
AD0134	R2AyHI !=    AVAL / AyHI

## › Missing operands

AD0133B	R2AyLO is populated but AVAL and/or AyLO are not populated
AD0133A	R2AyLO is present    but AVAL and/or AyLO is    not present
AD0134B	R2AyHI is populated but AVAL and/or AyHI are not populated
AD0134A	R2AyHI is present    but AVAL and/or AyHI is    not present
AD0132B	CHG, PCHG or R2BASE is populated but AVAL and/or BASE are not present/populated

## › Division by zero

AD0133C	AyLO = 0 but R2AyLO is populated
AD0134C	AyHI = 0 but R2AyHI is populated
AD0132A	BASE = 0 but R2BASE is populated
AD0225A	BASE = 0 but PCHG    is populated

# Tech Conformance and Duplicates

## › Logical Duplicates

### › *Criteria restricting records in a dataset*

AD0054	USUBJID duplicate values are not allowed within a study
AD0153	Multiple baseline records exist for a unique USUBJID,PARAMCD with BASETYPE null in all records
AD0154	Multiple baseline records exist for a unique USUBJID,PARAMCD,BASETYPE
AD0175	There is more than 1 record with LVOTFL = Y for a unique USUBJID and PARAMCD

## › Tech-Conformance

### › *Only apply to SAS datasets*

AD0058	*DT is not a numeric variable
AD0059	*TM is not a numeric variable
AD0060	*DTM is not a numeric variable
AD0041	*DT does not have the ADaM required <b>SAS Date format</b>
AD0042	*TM does not have the ADaM required <b>SAS time format</b>
AD0043	*DTM does not have the ADaM required <b>SAS Datetime format</b>

# Define.xml

## › Dataset Metadata

SD0061	Domain referenced in define.xml but dataset is missing	Dataset
SD1063	Dataset is not present in define.xml	
SD0054	Variable in define.xml is not present in the dataset	Variable
SD0060	Variable in dataset is not present in define.xml	
SD0059	Define.xml/dataset variable type mismatch	Datatype Codelist
SD0037	Value for Variable not found in user-defined codelist	

## › Value Level Metadata (requires Define.xml 2.0)

SD1228	Variable value not found in user-defined value level codelist	Codelist
SD1229	Variable value is null when value level condition is met	Mandatory
SD1230	Value level datatype mismatch for Variable	Datatype
SD1231	Variable value longer than defined max length	Length

# Adding sponsor-specific CT and VLM checks

- › CT - Controlled Terminology
- › VLM – Value Level Metadata, a define.xml 2.0 construct
- › The prior rules do not provide a complete solution.
- › Comprehensive ADaM validation requires the ability to supplement these rules with sponsor/study/TA specific
  - › controlled terminology
  - › value level checks
  - › Others...
- › We do not know about your CT and VLM....
  - › However if you create it as part of define.xml... we will validate it
  - › Creating CT and VLM should be prescriptive, and doing so using a text editor in the language of define.xml 2.0 can be quite challenging. This demo highlight tools to make it easier



# Demo Input

- › CDISC Define.xml 2.0 release package data
  - › Define.xml
  - › ADaM CDISCPILLOT01 Datasets
    - › ADSL.xpt – Analysis Datasets Subject Level
    - › ADQSADAS.xpt – Analysis Questionnaire data for ADAS-Cog : Alzheimer's Disease Assessment Scale- Cognitive Subscale
- › Latest OpenCDISC 1.6 rules
- › Some working software to
  - › modify define.xml
  - › Validate ADaM datasets with define.xml

# Demo Data

## › ADQSADAS (BDS) dataset

<b>STUDYID</b>	CDISCPIL0T01
<b>USUBJID</b>	01-701-1294
<b>TRTP</b>	Xanomeline Low Dose
<b>TRTSDT</b>	19441
<b>TRTEDT</b>	19523
<b>AVISIT</b>	Week 24
<b>VISIT</b>	RETRIEVAL
<b>ADY</b>	199
<b>ADT</b>	19639

PARAMCD	PARAM	AVAL	DTYPE	QSSEQ
ACITM01	Word Recall Task	5		5046
ACITM02	Naming Objects And Fingers (Refer To 5 C	0		5047
ACITM03	Delayed Word Recall	8		5048
ACITM04	Commands	0		5049
ACITM05	Constructional Praxis	9.9		5050
ACITM06	Ideational Praxis	0.01		5051
ACITM07	Orientation	9999		5052
ACITM08	Word Recognition	2		5053
ACITM09	Attention/Visual Search Task	28		5054
ACITM10	Maze Solution	1		5055
ACITM11	Spoken Language Ability	0	THIN AIR	5056
ACITM12	Comprehension Of Spoken Language	-999		5057
ACITM13	Word Finding Difficulty In Spontaneous S	0		5058
ACITM14	Recall Of Test Instructions	0	LOCF	5059
ACTOT	Adas-Cog(11) Subscore	9.25		5060

- › VLM sample attributes which constraint
  - › Datatype,length,mandatory and codelist
  - › based on the Value of PARAMCD

Dataset	Variable	WhereClause	Data Type	Length	Mandatory	Codelist
ADQSADAS	AVAL	PARAMCD EQ ACTOT	float	8	Yes	
ADQSADAS	AVAL	PARAMCD NE ACTOT	integer	3	No	
ADQSADAS	DTYPE	PARAMCD EQ ACTOT	text	10	Yes	DTYPE
ADQSADAS	DTYPE	PARAMCD NE ACTOT	text	10	No	DTYPE



# Common ADaM issues in submission data

- › Methodology
  - › Data
    - › 28 submission that include 62 studies from 2013 to 2015
    - › 22 sponsors
    - › Range from 7 to 51 datasets per study
    - › Range from 6000 to 36,000,000 records per dataset per study
    - › 4 to 134 unique validation issues across all studies
  - › Process
    - › Gather validation diagnostics
    - › Eliminate false-positive messages from analysis
    - › Summarize and group
  - › Results (table)
    - › Unique Issue
    - › % of the 62 studies having the issue

# Metadata issues

## › Define.xml

Studies with Define.xml v1.0 which cannot support ADaM metadata	81%
Define.xml <b>Codelist</b> mismatched	65%
Define.xml/dataset <b>variable type</b> mismatch	27%
<b>Variable</b> in dataset is not present in define.xml	29%
<b>Variable</b> in define.xml is not present in the dataset	6%
<b>Dataset</b> is not present in define.xml	2%

## › Illegal Variable names, labels, values

<b>zz</b> is not in [01-99] for ANLzzFL/FN	21%
<b>zz</b> is not in [01-99] for AOCCzzFL	8%
<b>xx</b> is not in [01-99] for TRTxxA	5%
<b>xx</b> is not in [01-99] for TRTxxAN	5%
<b>y</b> is not in [1-9] for (R2)AyLO	24%
<b>y</b> is not in [1-9] for (R2)AyHI	15%
<b>y</b> is not in [1-9] for PARCATy(N)	15%
<b>y</b> is not in [1-9] for CHGCATy	13%
<b>y</b> is not in [1-9] for CRITy(FL/FN)	13%
<b>y</b> is not in [1-9] for AVALCATy	6%
Variable <b>label</b> mismatch	79%
Illegal <b>PARAMCD</b> value	8%
ADY (Analysis day) = 0	10%

# Presence

## › Conditional

CNSR is present but not all of STARTDT, ADT and ADTM are present	48%
Secondary variable is present but its primary variable is not present	13%
APHASE is present but APERIOD is not present	11%

## › Unconditional

Required Variable not present	74%
-------------------------------	-----

- › TRTP
- › STUDYID
- › AESEQ
- › AESER
- › AGEU

# Categorization and pairing

## › Categories (1-way-map)

Inconsistent value for PARAMTYP	16%
Inconsistent value for AVALCAT1	29%
Inconsistent value for PARCAT1 within a unique PARAMCD	29%
Inconsistent value for PARCAT2 within a unique PARAMCD	5%
Inconsistent value for CHGCAT1	3%
Inconsistent value for BASECAT2	2%
Inconsistent value for PCHGCAT1	2%

## › Consistent Variable Pairs (2-way-maps)

Inconsistent value for BASEC within BASE	* 23%
Inconsistent value for BASE within BASEC	* 3%
Inconsistent value for PARAM within a unique PARAMCD	24%
Inconsistent value for PARAMCD within a unique PARAM	6%
Inconsistent value for PARAM	15%
Inconsistent value for PARAMN	10%
Inconsistent value for ATPT	5%

Inconsistent value for <b>AVAL</b> within AVALC	27%
Inconsistent value for <b>AVALC</b> within AVAL	48%

- › Some are logical false positives, but nevertheless inconsistencies based on a pure text match
  - › 6.2 vs 6.19865709
  - › 15.0 vs 15
  - › 0 vs null

# Baseline issues

- › span multiple rule categories:
  - › Presence
  - › Population
  - › Consistency pairs
  - › Equality of baselines
  - › Logical duplicates

Multiple baseline records exist for a unique USUBJID,PARAMCD,BASETYPE	31%
BASE or BASEC is populated for a unique USUBJID,PARAMCD but No baseline record exists	11%
BASE is present but ABLFL is not present	29%
BTOXGR is present but ABLFL is not present	18%
ABLFL = Y, but BASE != AVAL	23%
Inconsistent value for BASEC	23%
Inconsistent value for BASE	3%
Inconsistent value for BASECAT2	2%



# Calculation issues

CHG != AVAL - BASE	74%
PCHG != (AVAL - BASE) / BASE * 100	44%
BASE = 0 but PCHG is populated	11%

Some false positives due to SAS rounding issues with java engine

BASE	AVAL	CHG	Calculated	Exact()	Comment
146	175	146			Obvious errors
1.0485	1.121	0.0725	0.0725	TRUE	False-Positive
36.44444444	36.555556	0.11111111 <sup>1</sup>	0.11111111 <sup>2</sup>	FALSE	Accuracy issue - tool
0.21	0.24	0.04 <sup>4</sup>	0.03 <sup>3</sup>	FALSE	Accuracy issue - user
1.09786	1.16244	0.06458	0.06458	FALSE	?

# Terminology issues

## › Flag Coding

TRTEMFL value is not Y or null	26%
ANL02FL value is not Y or null	8%
ANL01FL value is not Y or null	3%
ABLFL value is not Y or null	2%
FUPFL value is not Y or null	2%

## › Controlled Terminology

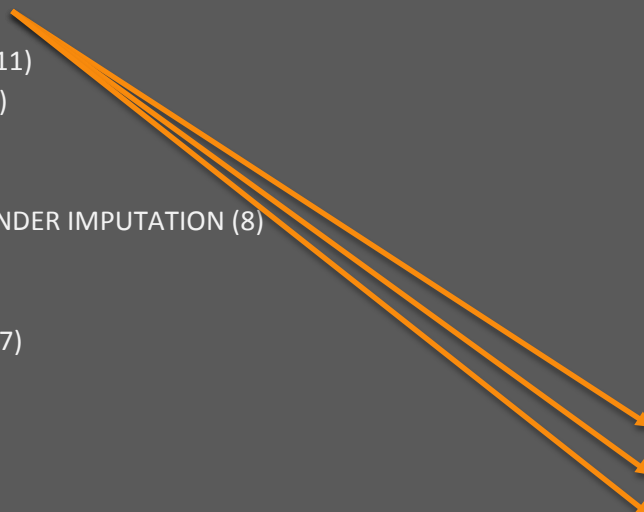
<b>DTYPE</b> value not found in 'Derivation Type' extensible codelist	71%
---	-----

## › Sample of terms outside codelist

- › WORST (17)
- › SUMMARY (11)
- › DERIVED (10)
- › COPY (9)
- › IMPUTED (9)
- › NON-RESPONDER IMPUTATION (8)
- › LPTCF (7)
- › SUM (7)
- › IMPUTED 0 (7)

## › CL.C81224.DTYPE

AVERAGE	
ENDPOINT	
INTERP	Interpolation
MAXIMUM	
MINIMUM	
ML	Maximum Likelihood
MOTH	Mean of Other Group
MOV	Mean Observed Value in a Group
BC	Best Case
BLOCF	Baseline Obs Carried Forward
BOCF	Best Observation Carried Forward
WC	Worst Case
WOFC	Worst Obs Carried Forward:
WOV	Worst Observed Value in a Group
SOCF	Screening Obs Carried Forward
POCF	Penultimate Obs Carried Forward
LOCF	Last Obs Carried Forward



# Traceability Issues

## › To SDTM DM

For the same USUBJID, ADSL.RACE <> DM.RACE	16%
For the same USUBJID, ADSL.AGEU <> DM.AGEU	15%
For the same USUBJID, ADSL.AGE <> DM.AGE	11%
For the same USUBJID, ADSL.ARM <> DM.ARM	8%
For the same USUBJID, ADSL.SUBJID <> DM.SUBJID	2%

## › To SDTM EX, AE

ADaM ADAE record key is not traceable to SDTM.AE	15%
SDTM.EX is present but neither ADSL TRTEDT nor TRTEDTM are present	10%
SDTM.EX is present but neither ADSL TRTSDT nor TRTSDTM are present	10%
USUBJID value does not exist in the SDTM DM domain	10%

- › Integrated data. E.g., DB + OL-extension studies
- › Wrong study

## › To ADaM

Subject is off treatment, but analysis date (ADT) is within treatment period	19%
Subject is on treatment, but analysis date (ADT) is after treatment end date	18%
BDS.APERIOD xx does not have a corresponding ADSL.TRxxEDT variable	15%

# Tech Conformance issues

*TM is not a numeric variable	23%
*DTM is not a numeric variable	18%
*DT is not a numeric variable	8%

- › All date and time fields are numbers in SAS
  - › Date: days between 1960-01-01 and a specified date
  - › Time: seconds since midnight of current day (0 to 86400)
  - › Datetime: seconds between 1960-01-01 and an h/m/s within a specified date
- › Usually due to incorrect usage of variables
  - › AESTENDT="2015-01-01:2015-04-30" (text)
- › ARELTM (Analysis Relative Time) is excluded

# Rules for the upcoming release of ADaM IG v1.1

- › Occurrence dataset rules
- › Enhancements to BDS and ADSL since many new variables have been added
- › Corrections to existing CDISC Validation Checks v 1.3 due to either
  - › Human error
  - › Requirements clarification
  - › New requirements to existing variables (i.e. y can now be in 1-99 instead of 1-9)
- › Rules that have always exists in ADaM 1.0 but may have been missed (examples...)

ADaM IG 1.0 page	ADaM IG text, paraphrased or interpreted	note
24	AVISIT should be <b>unique</b> for a given analysis visit window.	
31	<b>AWTDIFF</b> = ABS (ADY - AWTARGET) or ABS(ARELTM - AWTARGET)	ABS is the absolute value
31	<b>AWRANGE</b> = CONCAT(AWLO + <anything> + AWHI + AWU	ignoring all spaces and dashes
31	AWRANGE, AWLO , AWHI, AWU must be <b>present and populated together</b> .	

# Future rule considerations

- › Usage/adoption of Analysis Results Metadata
- › How to standardize Incidence/Occurrence data (Intervention ,Events, subject visits)
- › Therapeutic Area (TA) specific ADaM checks for safety and efficacy ?
- › More standardization on CT and VLM
- › Rules to validate if data can support Commonly Used Statistical Analysis Methods ??????
- › examples document published on 2011-12-16 by CDISC references

ANOVA - Analysis of Variance	Mixed Models Analysis
Chi-squared	Regression, Cox
Chi-squared, Corrected	Regression, Linear
Cochran-Mantel-Haenszel	Regression, Logistic
Mantel Haenszel	Sign Test
Fisher's Exact	t-Test, 1-sided
Kruskal-Wallis	T-Test, 2-sided
Log Rank	Wilcoxon (Mann-Whitney)
McNemar	

# Questions

Michael DiGiantomasso

mike.digian@pinnacle21.net *(mike@opencdisc.org)*

Sergiy Sirichenko

ssirichenko@pinnacle21.net *(sergiy@opencdisc.org)*

# Addendum

- › Rule Code Updates
- › Additional/Enhanced rules not published by CDISC Validation Checks v1.3
- › Links to useful Resources



# Rule Code Updates

In order to better align with CDISC Published rules ids and versions, the following OpenCDISC rule codes were modified. The rule descriptions also note this change. Enterprise customers can filter on “*formerly*” to produce the list on demand

1. AD1006 -> AD0013
2. AD1003 -> AD0019
3. AD1004 -> AD0026
4. AD1001 -> AD0047
5. AD1010 -> AD0061A
6. AD1013 -> AD0127A
7. AD1023 -> AD0132A
8. AD1020 -> AD0132B
9. AD1016 -> AD0133A
10. AD1025 -> AD0133B
11. AD1027 -> AD0133C
12. AD1017 -> AD0134A
13. AD1026 -> AD0134B
14. AD1028 -> AD0134C
15. AD1007 -> AD0143
16. AD0177 -> AD0154
17. AD1008 -> AD0196
18. AD1022 -> AD0225A
19. AD1021 -> AD0239A
20. AD1024 -> AD0282A

ADaM v1.0

Properties Domains Variables Rules

Remove Selected Q, formerly

ID ^	Message ^	CDISC ^
AD0013	illegal variable name: basic format is violated	13,14,15
AD0019	Variable subject-population flag value is null	19,20,21,22,23,24,25
AD0026	Variable subject-population flag value is null	26,27,28,29,30,31,32
AD0047	Required variable is not present	47,49,50,51,52,55,71,72,88,89,90,194,195,257,260,261,262,278
AD0061A	SDTM.EX is present but neither ADSL TRTEDT nor TRTEDTM are present	61+
AD0127A	BASE or BASEC is present but ABLFL is not present	127+,128+
AD0132A	calculation issue: BASE = 0 but R2BASE is populated	132+
AD0132B	calculation issue: CHG, PCHG or R2BASE is populated but AVAL and/or BASE are not present and populated	132+,223+,225+
AD0133A	calculation issue: R2AyLO is present but AVAL and/or AyLO is not present	133+
AD0133B	calculation issue: R2AyLO is populated but AVAL and/or AyLO are not populated	133+
AD0133C	calculation issue: AyLO = 0 but R2AyLO is populated	133+
AD0134A	calculation issue: R2AyHI is present but AVAL and/or AyHI is not present	134+
AD0134B	calculation issue: R2AyHI is populated but AVAL and/or AyHI are not populated	134+
AD0134C	calculation issue: AyHI = 0 but R2AyHI is populated	134+
AD0143	Illegal PARAMCD value	143,144,145
AD0154	Multiple baseline records exist for a unique USUBJID,PARAMCD,BASETYPE	154,155
AD0196	Required Variable value is null	196,197
AD0225A	calculation issue: BASE = 0 but PCHG is populated	225+
AD0239A	TRxxPGy and TRTxXA are present but TRxxAGy is not present	239+
AD0282A	Analysis severity Variable value is null	282+

# Additional/Enhanced rules

not published by CDISC Validation Checks v1.3

For some rules, the enhancement is subtle. For example, AD0042 excludes AERELTM. AD0127A checks presence, whereas published rule only checks population. AD0188 allows the numeric version of the variable. The enhancements can be easily queried with a text search in enterprise as seen in next slide. Some of these rules may be in community

ID	CDISC	Message
AD0042	42	*TM does not have the ADaM required SAS time format
AD0061A	61+	SDTM.EX is present but neither ADSL TRTEDT nor TRTEDTM are present
AD0127A	127+, 128+	BASE or BASEC is present but ABLFL is not present
AD0132A	132+	calculation issue: BASE = 0 but R2BASE is populated
AD0132B	132+, 223+, 225+	calculation issue: CHG, PCHG or R2BASE is populated but AVAL and/or BASE are not present and populated
AD0133A	133+	calculation issue: R2AyLO is present but AVAL and/or AyLO is not present
AD0133B	133+	calculation issue: R2AyLO is populated but AVAL and/or AyLO are not populated
AD0133C	133+	calculation issue: AyLO = 0 but R2AyLO is populated
AD0134A	134+	calculation issue: R2AyHI is present but AVAL and/or AyHI is not present
AD0134B	134+	calculation issue: R2AyHI is populated but AVAL and/or AyHI are not populated
AD0134C	134+	calculation issue: AyHI = 0 but R2AyHI is populated
AD0159	159+	calculation issue: AWTDIFF is populated but AWTARGET and/or one of [ADY,ARELTM] are not both present and populated
AD0188	188+	illegal variable name: y is not in [1-9] for CHGCATy(N)
AD0225A	225+	calculation issue: BASE = 0 but PCHG is populated
AD0239A	239+	TRxxPGy and TRTxxA are present but TRxxAGy is not present
AD0282A	282+	Analysis severity Variable value is null
AD1011	TBD+	Secondary variable Variable is populated but its primary variable * is not populated
AD1012	TBD+	Secondary variable is present but its primary variable is not present
AD1015	TBD+	APHASE is present but APERIOD is not present
AD1018	TBD+	illegal variable name: zz is not in [01-99] for ANLzzFL/FN
AD1019	TBD+	illegal variable name: xx is not in [01-99] for an ADSL trial date
SD1228	TBD+	Variable value not found in user-defined codelist
SD1229	TBD+	Variable value is null when value level condition occurs
SD1230	TBD+	Value level datatype mismatch for Variable when value level condition occurs
SD1231	TBD+	Variable value longer than defined max length when value level condition occurs

# Additional/Enhanced rules

not published by CDISC Validation Checks v1.3

A user can quickly see that AD0127 was published with only population in mind, whereas AD0127A checks for presence of BASE or BASEC if the baseline flag is present

The screenshot shows the OpenCDISC ENTERPRISE interface. The breadcrumb navigation is Home > Design > Validation Specs > ADaM v1.0. The main heading is ADaM v1.0. Below it are tabs for Properties, Domains, Variables, and Rules. The Rules tab is active, showing a search bar with 'AD0127' and a 'Remove Selected' button. A table lists the rules:

ID	Message	CDISC	FDA
AD0127	BASE or BASEC is populated for a unique USUBJID,PARAMCD but no baseline record exists	127, 128	
AD0127A	BASE or BASEC is present but ABLFL is not present	127+, 128+	

In this case, the BDS version (AD0239A) of the published ADSL check was forgotten, but clearly documented on ADaM IG p 16.

The screenshot shows the OpenCDISC ENTERPRISE interface with the search bar containing 'ad0239'. The table lists the rules:

ID	Message	CDISC
AD0239	TRTPGy and TRTA are present but TRTAGy is not present	239
AD0239A	TRxxPGy and TRTxxA are present but TRxxAGy is not present	239+

# Links to useful Resources

These are essentially what the rules are and will be based upon. Interesting to note are the newly published **OCCDS** and **IADSL Considerations** on June 9<sup>th</sup>

- ✓ <http://cdisc.org/adam>
- ✓ <http://cdisc.org/adam-future>
- ✓ <http://www.opencdisc.org/>
- ✓ <http://www.opencdisc.org/forums/adam>
- ✓ <http://www.pinnacle21.net/>
- \* *links require a CDISC login id*
- ✓ \* [CDISC ADaM Validation Checks](#)
- ✓ \* [CDISC ADaM Examples in Commonly Used Statistical Analysis Methods](#)
- ✓ \* [CDISC ADaM v2.1](#)
- ✓ \* [CDISC ADaM IG v1.0](#)
- ✓ \* [CDISC ADaM IG v1.1 \(coming soon\)](#)
- ✓ \* [CDISC Define.xml](#)
- ✓ \* [CDISC ARM for Define.xml 2.0](#)
- ✓ \* [CDISC ADaM ADAE v1.0](#)
- ✓ \* [CDISC ADaM BDS TTE v1.0](#)
- ✓ \* [CDISC ADaM OCCDS v1.0](#)
- ✓ \* [CDISC ADaM IADSL Considerations](#)

ARM - Analysis Results Metadata  
ADAE - Analysis Adverse Events  
BDS - Basic Data Structure  
TTE - Time to Event Analysis  
OCCDS - Occurrence Data Structure  
IADSL - integrated dataset for ADSL