



CD09

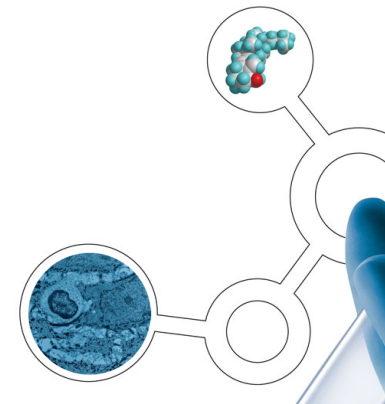
Interpretation of the output from the OpenCDISC validator

Jørgen Mangor Iversen
LEO Pharma A/S



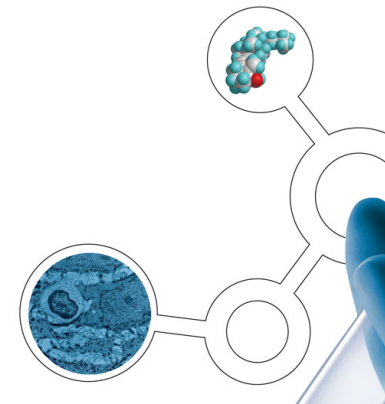
About LEO Pharma A/S

- Independent, research-based pharmaceutical company
- Founded in 1908
- Employs around 5,000 employees worldwide
- Headquartered in Denmark
- Fully owned by the LEO Foundation
- Drugs to dermatologic and thrombotic patients in more than 100 countries
- Vision of becoming the preferred dermatology care partner



What Is OpenCDISC Validator?

- Validating clinical data complying to CDISC standards
 - SDTM
 - SEND
 - ADaM
- Free <http://www.opencdisc.org/>
- Fully configurable
- Open source (sort of)
- New versions ~6 months interval
 - Tool as well as standards
- Version 1.2 covered
 - Version 1.4 available
- Commercial version available
 - Supported



What Does It Look Like?

The screenshot shows the OpenCDISC Validator application window. The title bar reads "OpenCDISC Validator". The menu bar contains "File" and "Help". The main interface is divided into several sections:

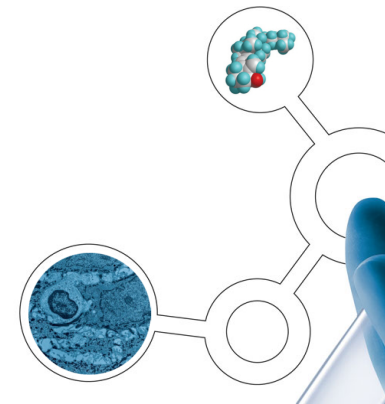
- What would you like to do?**: Two radio buttons are present: "Validate Data" (selected) and "Generate Define.xml".
- Standard:**: A dropdown menu showing "SDTM".
- Source Format:**: A dropdown menu showing "SAS® Transport (XPORT)".
- Source Data:**: A large empty text area for listing source files. To its right are three buttons: "Browse", "Remove", and "Clear".
- Configuration:**: A dropdown menu showing "config-sdtm-3.1.2.xml".
- Define.xml:**: An empty text input field with a "Browse" button to its right.
- Optional**: A label indicating that the following options are optional.
- Report Format:**: A dropdown menu showing "Excel" with a link to "Report Settings".
- MedDRA:**: A dropdown menu showing "13.1".
- Progress:**: An empty progress bar with the text "Waiting to begin." below it.
- Start/Stop:**: Two buttons at the bottom right, "Start" and "Stop".





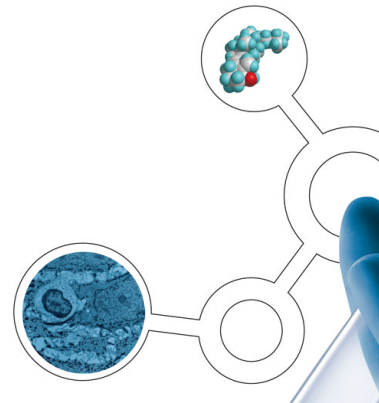
Who is Behind OpenCDISC Validator?

- Vendor is Pinnacle 21
 - Pennsylvania, USA
- Claims to be independent consultancy company
- Ownership and dependencies unknown
- Trust in the product must come from
 - Open source nature
 - Experienced quality of the product
 - Acceptance by Pharma community



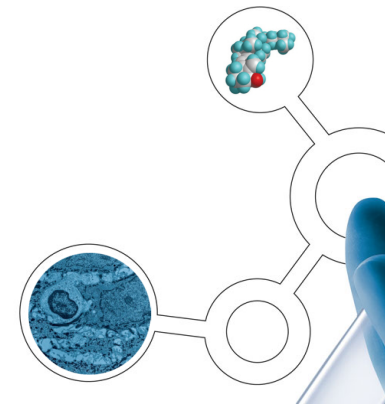
Why Do We Need OpenCDISC Validator?

- Imaginative interpretations of CDISC standards
 - CROs
 - In-house Data Management
- Bend standards to encompass company specific flavour
- Several vendor products exist
- OpenCDISC Validator advantages
 - Everybody can afford it (**FREE**)
 - Vendor and data format independent



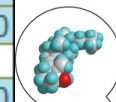
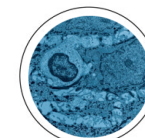
What Does OpenCDISC Validator Do?

- Validates the format and contents of data
 - SDTM, SEND, ADaM, define.xml, custom data
 - With or without define.xml
 - Create define.xml
 - MedDRA version
- Input data
 - SAS V5 transport files
 - Delimited ASCII files
 - Configuration metadata
- Output report
 - Excel spread sheets
 - CSV files



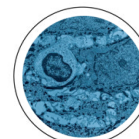
Dataset Summary

OpenCDISC Validator Report								
Configuration: P:\business\opencdisc-validator\config\config-sdtm-3.1.1.xml								
Define.xml: Not provided								
Generated: 30-Jun-2010 15:34:55								
Processed Sources								
Name	Label	Class	Source	Records	Errors	Warnings	Infos	
AE	Adverse Events	Events	AE.xpt	88	0	4	0	
CM	Concomitant Medications	Interventions	CM.xpt	507	0	1	0	
DM	Demographics	Special Purpose	DM.xpt	102	0	0	4	
DS	Disposition	Events	DS.xpt	1126	820	1	2	
LB	Laboratory Tests	Findings	LB.xpt	203	197	4	1	
MH	Medical History	Events	MH.xpt	1138	36	1	0	
PE	Physical Examinations	Findings	PE.xpt	462	0	6	1	
QS	Questionnaires	Findings	QS.xpt	721	0	722	1	
SC	Subject Characteristics	Findings	SC.xpt	308	206	80	0	
SUPPQUAL	SUPPQUAL	Special Purpose	SUPPQUAL.xpt	2109	0	0	0	
VS	Vital Signs	Findings	VS.xpt	3476	0	1	1	
Total				10240	1259	820	10	
Unprocessed Sources								
Name	Label	Class	Reason	Errors	Warnings	Infos		
Total				0	0	0		
Grand Total				10240	1259	820	10	
▶ ▶ Dataset Summary Issue Summary Details Rules 🔍 ◀ ▶								



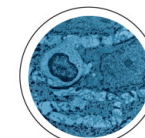
Issue Summary

OpenCDISC Validator Report		
Configuration: P:\business\opencdisc-validator\config\config-sdtm-3.1.1.xml		
Define.xml: Not provided		
Generated: 30-Jun-2010 15:33:41		
Issue Summary		
Rule ID	Message	Found
Errors		
SD0002	Null value in variable marked as Required	28
SD0003	Invalid ISO 8601 value	205
SD0005	Non-unique value for SEQ	1026
Total		1259
Warnings		
SD0017	Invalid value for --TEST variable	718
SD0026	Missing units on value	79
SD0040	Inconsistent value for Name of Measurement, Test, or Examination	2
SD0063	SDTM/dataset variable label mismatch	21
Total		820
Infos		
SKIP_SD0006	Validation rule SD0006 could not be executed because required data from EG could not be found.	1
SKIP_SD0065	Validation rule SD0065 could not be executed because required data from SV could not be found.	5
SKIP_SD0066	Validation rule SD0066 could not be executed because required data from TA could not be found.	1
SKIP_SD0070	Validation rule SD0070 could not be executed because required data from EX could not be found.	1
SKIP_SD0071	Validation rule SD0071 could not be executed because required data from TA could not be found.	1
SKIP_SD1015	Validation rule SD1015 could not be executed because required data from TA could not be found.	1
Total		10
▶ ▶ Dataset Summary Issue Summary Details Rules 🔍 ◀ ▶		



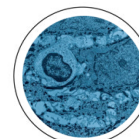
Details

Name	Record	Variable	Value	Rule ID	Message	Category	Type	Severity
VS		VARIABLE, LABEL	VSORRES, Results or	SD0063	SDTM/dataset variable label mismatch	Metadata	Warning	Low
VS	1			SKIP_SD0065	Validation rule SD0065 could not be executed because required data from	System	Information	Low
DM	1			SKIP_SD0006	Validation rule SD0006 could not be executed because required data from	System	Information	Low
DM	1			SKIP_SD0066	Validation rule SD0066 could not be executed because required data from	System	Information	Low
DM	1			SKIP_SD0070	Validation rule SD0070 could not be executed because required data from	System	Information	Low
DM	1			SKIP_SD0071	Validation rule SD0071 could not be executed because required data from	System	Information	Low
MH		VARIABLE, LABEL	MHOCCUR, Medical	SD0063	SDTM/dataset variable label mismatch	Metadata	Warning	Low
MH	36	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	42	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	53	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	57	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	61	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	87	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	93	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	107	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	114	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	122	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	124	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	127	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High
MH	401	MHTERM	null	SD0002	Null value in variable marked as	Presence	Error	High



Rules

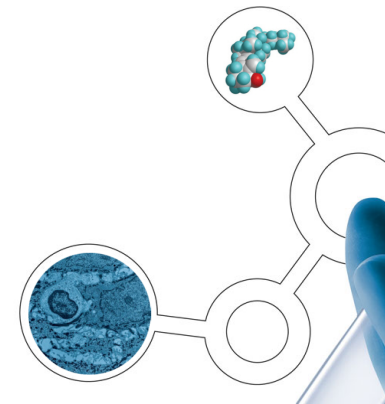
Rule ID	Message	Description	Category	Type	Severity
CT0004	Value for AGEU not found in AGEU controlled terminology	Variable values should be populated with terms found in 'Age Unit' (C66781) CDISC controlled terminology	Terminology	Error	High
CT0006	Value for COUNTRY not found in COUNTRY controlled terminology	Variable values should be populated with terms found in 'Country' (C66786) CDISC controlled terminology	Terminology	Warning	Medium
CT0020	Value for IECAT not found in IECAT controlled terminology	Variable values should be populated with terms found in 'Category for Inclusion/Exclusion' (C66797) CDISC	Terminology	Error	High
CT0034	Value for SEX not found in SEX controlled terminology codelist	Variable values should be populated with terms found in 'Sex' (C66731) CDISC controlled terminology codelist	Terminology	Error	High
CT0038	Value for --STRF not found in STENRF controlled terminology	Variable values should be populated with terms found in 'Relation to Reference Period' (C66728) CDISC	Terminology	Error	High
CT0039	Value for --ENRF not found in STENRF controlled terminology	Variable values should be populated with terms found in 'Relation to Reference Period' (C66728) CDISC	Terminology	Error	High
CT0044	Value for AETOXGR not found in TOXGR controlled terminology	Variable values should be populated with terms found in 'Common Terminology Criteria for Adverse Events'	Terminology	Warning	Medium
CT0059	Value for --BLFL not found in NY controlled terminology codelist	Variable values should be populated with terms found in 'No Yes Response' (C66742) CDISC controlled	Terminology	Warning	Medium
CT0060	Value for --DRVFL not found in NY controlled terminology	Variable values should be populated with terms found in 'No Yes Response' (C66742) CDISC controlled	Terminology	Error	High
CT0061	Value for --FAST not found in NY controlled terminology codelist	Variable values should be populated with terms found in 'No Yes Response' (C66742) CDISC controlled	Terminology	Error	High
CT0062	Value for --OCCUR not found in NY controlled terminology	Variable values should be populated with terms found in 'No Yes Response' (C66742) CDISC controlled	Terminology	Error	High
CT0064	Value for AESER not found in NY controlled terminology codelist	Variable values should be populated with terms found in 'No Yes Response' (C66742) CDISC controlled	Terminology	Error	High
CT0065	Value for AESCONG not found in NY controlled terminology	Variable values should be populated with terms found in 'No Yes Response' (C66742) CDISC controlled	Terminology	Error	High
	Value for AFSDISAB not found in	Variable values should be populated with terms found in			





CDISC Ambiguities

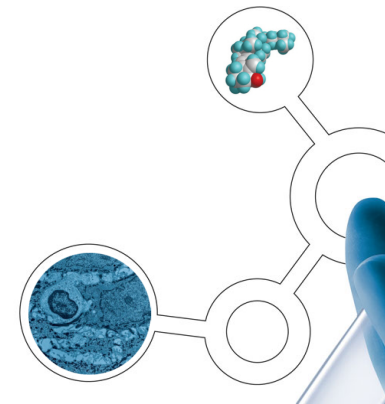
- SDTM 3.1.2 covered
 - Accepted by FDA
- ODM ambiguities
 - Empty XML tags
- **Label** vary between classes
- **Role** vary between classes
- **Controlled Terminology** vary between classes
- Variable names reused for different purposes
- Unique variables for all uses and definitions
- **Keys** and little else should have the same name





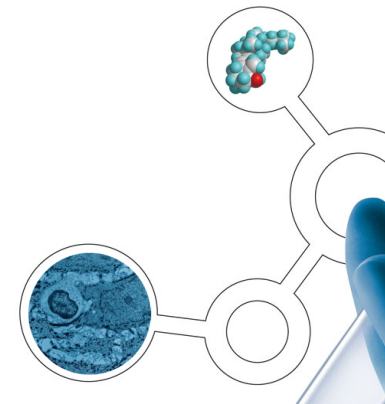
OpenCDISC Ambiguities

- Validator is fully configurable
 - ...but very near complete!
 - Custom configurations leads to version confusion
 - Unambiguous default configuration within reach
- Missing units on scores + ratios (later corrected)
 - SD0026, SD0029
- Truncation of data
 - SD0017: --TEST value > 40 characters: **ERROR**
 - SD1049: QLABEL value > 40 characters: **WARNING**
 - Both are candidates to be IDLABEL in Proc Transpose



Strategy for Handling Messages

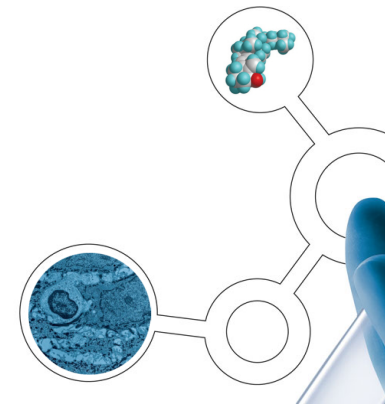
- Define level of messages acceptable
 - Errors
 - ❖ Unevenly defined: Controlled Terminology as Error/Warning
 - Warnings
 - ❖ Negative intervals, Reference to non-existing visits
 - Informational
 - ❖ Always safe to ignore
- General rules
 - Errors not acceptable
 - Warnings acceptable
 - But...





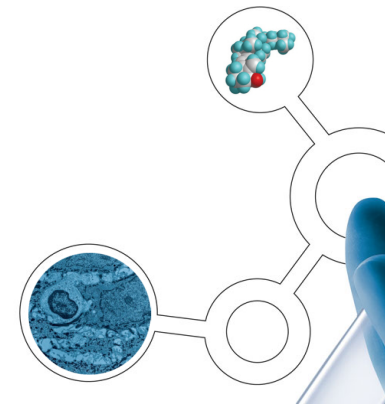
Acceptable Errors

- Missing units on scores (and ratios)
- Known data errors that cannot be corrected



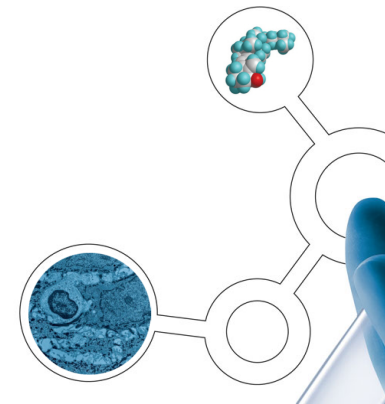
Warnings that should be Errors

- Warnings of category 'Cross-reference'
- Possible exceptions
 - SD0065: Invalid Subject Visit/Visit Number
 - SD1023: Invalid VISIT/VISITNUM
 - Dangling reference
 - ❖ Mix of planned and actual data in CDISC standards (ambiguity)
 - SD1014: Invalid TAETORD
 - SD1015: Invalid EPOCH
 - Should be down-classified as 'Consistency'
 - ❖ EPOCH is a political priority, not a data model error



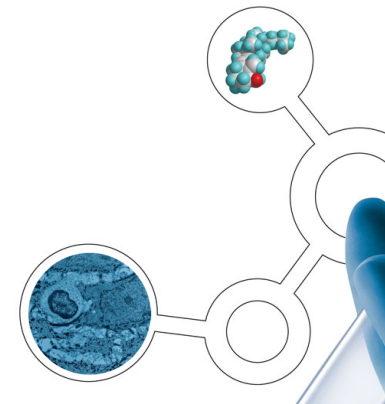
More Warnings that should be Errors

- Some 'Consistency' Warnings
 - SD0040: Inconsistent value for --TEST within --TESTCD
 - SD0046: Inconsistent value for QLABEL within QNAM
 - SD0051: Inconsistent value for VISIT within VISITNUM
 - SD0052: Inconsistent value for VISITNUM within VISIT
 - SD0090: AESDTH is not 'Y', when AEOUT='FATAL'
 - SD0091: AEOUT is not 'FATAL', when AESDTH='Y'
 - SD1062: AESER is not 'Y', when AESOD equals 'Y'
- Concerns 1:1 relationships between two variables
- Prone to create duplicates in data
- SD1060: Duplicate VISITNUM
 - Explicitly concerns duplicates



Refined Strategy for Handling Messages

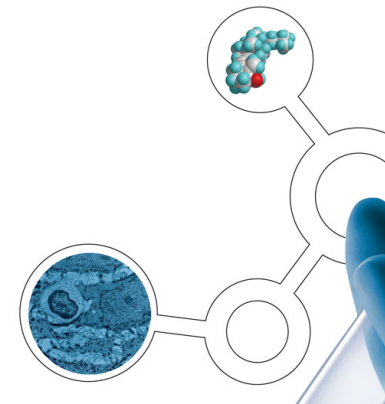
- Aim for no errors, but allow for errors regarding missing units and unavoidable data errors
- Treat warnings classified as 'Cross-reference' as if they were errors, possibly allowing for invalid TAETORD and EPOCH messages
- Treat the warnings classified as 'Consistency' **and** dealing with 1:1 relationships or duplicates as errors. Allow for no exceptions





Message Strategy Addendum

- Familiarise yourself with OpenCDISC rules
 - View the XML configuration files supplied directly
 - ❖ Display very well in a browser
 - ❖ Organised per dataset
 - ❖ Handy when constructing specific datasets
 - View any Excel report generated
 - ❖ Switch to the 'Rules' sheet
 - ❖ Apply filtering
 - ❖ Filter by 'Category' and/or 'Severity'
 - ❖ Reduces the number of rules to small sets
 - In either case
 - ❖ Read each rule in its context
 - ❖ Reflect on its implications



XML Configuration File in Browser

OpenCDISC VALIDATOR

TABLE OF CONTENTS

- Global Rules Section (GLOBAL)
- Demographics (DM)
- Comments (CO)
- Subject Elements (SE)
- Subject Visits (SV)
- Interventions (Interventions)
- Concomitant Medications (CM)
- Exposure (EX)
- Substance Use (SU)
- Events (Events)
- Adverse Events (AE)
- Disposition (DS)
- Medical History (MH)
- Protocol Deviations (DV)
- Clinical Events (CE)
- Findings (Findings)
- ECG Test Results (EG)
- Inclusion/Exclusion Criteria Not Met (IE)
- Laboratory Tests Results (LB)
- Physical Examination (PE)
- Questionnaires (QS)
- Subject Characteristics (SC)
- Vital Signs (VS)
- Drug Accountability (DA)
- Microbiology Specimen (MB)
- Microbiology Susceptibility Test (MS)
- Pharmacokinetic Concentrations (PC)
- Pharmacokinetic Parameters (PP)
- Findings About Events or Interventions (FA)
- Trial Arms (TA)
- Trial Elements (TE)
- Trial Visits (TV)
- Trial Inclusion/Exclusion Criteria (TI)
- Trial Summary (TS)

Validator Configuration

Global Rules Section

Dataset Structure

Variable	Description	Required	Data Type	Length
STUDYID	Study Identifier	✓	text	200

Validation Rules

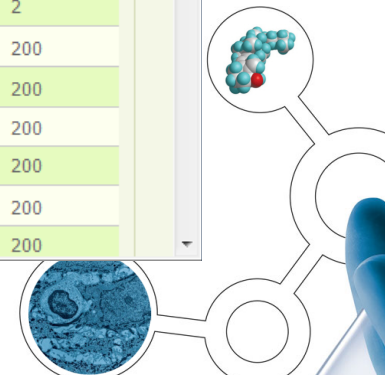
ID	Description	Validator	Message	Type	Severity	Active
SD0095	Supplemental Qualifiers special purpose dataset (SUPP--) can only be used to capture non-standard variables and their association to parent records in general-observation-class datasets (Events, Findings, Interventions) and Demographics	Condition	SUPQUAL dataset is used for non-general-observation-class Domain	❗ Error	●	✓

[back to top](#)

Demographics

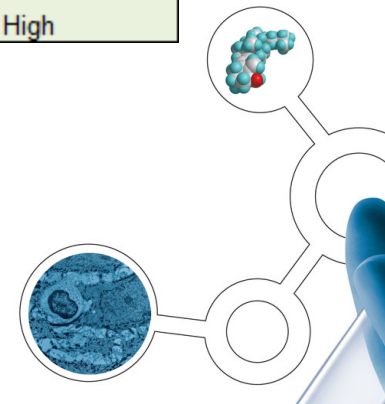
Dataset Structure

Variable	Description	Required	Data Type	Length
STUDYID	Study Identifier	✓	text	200
DOMAIN	Domain Abbreviation	✓	text	2
USUBJID	Unique Subject Identifier	✓	text	200
SUBJID	Subject Identifier for the Study	✓	text	200
RFSTDTCT	Subject Reference Start Date/Time		datetime	200
RFENDTCT	Subject Reference End Date/Time		datetime	200
SITEID	Study Site Identifier	✓	text	200
INVID	Investigator Identifier		text	200



Excel Report Rules Tab

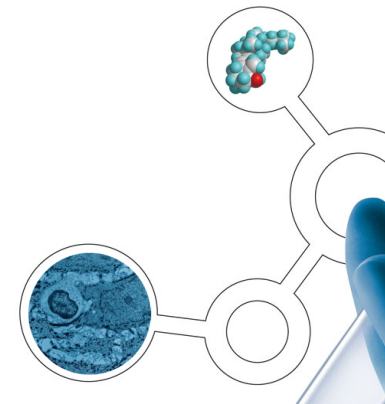
Rule	Message	Description	Category	Type	Severity
SD0003	Invalid ISO 8601 value	Dates and times of day must conform to the ISO 8601 international standard.	Format	Error	High
SD0010	VISITNUM should not be formatted to more than 3 decimal places	Identifies records where the value for Visit Number is formatted to more than three decimal places	Format	Warning	Medium
SD0017	Invalid value for --TEST variable	Identifies records that violate condition [LENGTH(Name of Measurement, Test or Examination) less than or equal to 40 characters]	Format	Warning	Low
SD0018	Invalid value for --TESTCD variable	Identifies records that violate condition [LENGTH(Short Name of Measurement, Test or Examination) less than or equal to 8 chars, cannot start with a number or contain special chars]	Format	Warning	Low
SD0019	Invalid value for TSPARM variable	Identifies records that violate the condition [LENGTH(Trial Summary Parameter) less than or equal to 40 characters]	Format	Warning	Low
SD0020	Invalid value for TSPARMCD variable	Identifies records that violate the condition [LENGTH(Trial Summary Parameter Short Name) less than or equal to 8 chars, cannot start with a number or contain special chars]	Format	Warning	Low
SD1011	Invalid ISO 8601 value	Duration, Elapsed Time, and Interval values must conform to the ISO 8601 international standard.	Format	Error	High





Conclusions

- OpenCDISC validator is great at ensuring high data quality
 - In spite of small ambiguities of both standards and tools
- A true pleasure to learn that methods have mature status
 - Still early days in the life cycle of standardising clinical data
- Vision that handling of clinical data can be fully automated
 - As standards and tools evolve
- Understanding and analysing data is a different story...





Questions

