

ADaM Validation and Integrity Checks

Wednesday 12th October 2011

Louise Cross

ICON Clinical Research, Marlow, UK

- ADaM Validation document
- Adaptability **ICON in house macro**
- Design and scope
- Interpretation
- Validation of the validation
- Future development

- ICON have a depth of experience to draw upon when developing and validating ADaM datasets
 - ADaM and SDTM have been standard within the department for many years
- Validating ADaM datasets
 - Timely specification review
 - Dual programming
 - Open CDISC Validation tool
 - Core team review
 - Check SAP content is covered and ADaM IG is adhered to
- Large conversion studies
 - Need for further streamlined tools

The definitions within the ADaM Implementation Guide 1.0 include specific guidelines and rules for defining and creating ADaM data sets. This document contains a list of requirements which may be used to validate datasets against a subset of these rules which are objective and unambiguously evaluable. The validation checks within this document are defined to be machine readable (i.e. programmable within computer software) and capable of being implemented by ADaM users.

- Taken from “ADaM Validation Checks v1.1.pdf” on www.CDISC.org

- In September 2010, CDISC released the initial version of the ADaM Validation checks document
 - January 2011 saw the release of a new version 1.1
- Version 1.1 has incorporated feedback from users on the first release regarding clarification, errors, duplicate checks
- Additional checks suggested by users and members of CDISC are being reviewed with the hope to include in the next version of the document
- User feedback is crucial to the development of such a document

- Excel document containing an extensive list of close to 150 recommended checks, with supporting explanatory documentation stored in a pdf file
- In the members section of the CDISC website, a link to the new ADaM Validation checks specification

[HOME](#) [BLOG](#) [SITE MAP](#) [FAQ](#) [CONTACT](#) [SEARCH](#) [PORTAL](#)



[ABOUT](#) [STANDARDS & INNOVATIONS](#) [RESOURCES](#) [NEWS](#) [EDUCATION & EVENTS](#) [MEMBERSHIP](#) [MEMBERS ONLY](#)

The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.



Strength through Co

- Terminology Governance and Implementation
- Therapeutic Area Standards
- **New ADaM Validation Checks****
- FDA CDISC Meetings
- PGx SDTM Domains Review Package
- Case Studies
- SDTM/ADaM Pilot Project
- Business Case for Use of Standards
- Webinars, Presentations and Slides
- Team Information
- Tools

- The ADaM Validation Checks document is logically designed with detailed information regarding each check
- The documentation contains seven columns:
 - Check Number
 - ADaM IG Section Number
 - Text from ADaM IG
 - ADaM Structure Group
 - Functional Group
 - ADaM Variable Group
 - Machine-Testable Rule

3 Standard ADaM Variables

This section defines the required characteristics of standard variables (columns) that are frequently needed in analysis datasets. The ADaM standard requires that these variable names be used when a variable that contains the content defined in Section 3 is included in an analysis dataset.

Section 3.1 describes variables in ADSL. Section 3.2 describes variables in the BDS.

In this section, ADaM variables are described in tabular format. The two rightmost columns, “Core” and “CDISC Notes” provide information about the variables to assist users in preparing their datasets. These columns are not meant to be metadata submitted in define.xml. The “Core” column describes whether a variable is required, conditionally required, or permissible. The “CDISC Notes” column provides more information about the variable. In addition, the “Type” column is being used to define whether the variable being described is character or numeric. More specific information will be provided in metadata (e.g., text, integer, float).

Values of ADaM “Core” Attribute

- **Req** = Required. The variable must be included in the dataset.
 - **Cond** = Conditionally required. The variable must be included in the dataset in certain circumstances.
 - **Perm** = Permissible. The variable may be included in the dataset, but is not required.
- Unless otherwise specified, all ADaM variables are populated as appropriate, meaning nulls are allowed.

General Variable Naming Conventions

1. In a pair of corresponding variables (e.g., TRTP and TRTPN, AVAL and AVALC), the primary or most commonly used variable does not have the suffix or extension (e.g., N for Numeric or C for Character).
2. The names of date imputation flag variables end in DTF, and the names of time imputation flag variables end in TMF.
3. The names of all other character flag (or indicator) variables end in FL, and the names of the corresponding numeric flag (or indicator) variables end in FN. If the flag is used, the character version (*FL) is required but the numeric version (*FN) can also be included.
4. Any ADaM variable whose name is the same as an SDTM variable must be a copy of the SDTM variable, and its label, meaning, and values must not be modified. ADaM adheres to a principle of harmonization known as “same name, same meaning, same values.”

- Positive step in highlighting sense checks to perform on ADaM datasets to ensure compliance with the IG
- ICON have developed a key validation tool from this specification
- SAS macro has been created and validated to use on any ADaM study
 - The macro performs all recommended checks
 - Outputs any issues to be addressed
- Conversion studies
 - Valued tool to ensure consistency and accuracy

Design and Scope of the Macro

Clinical Research

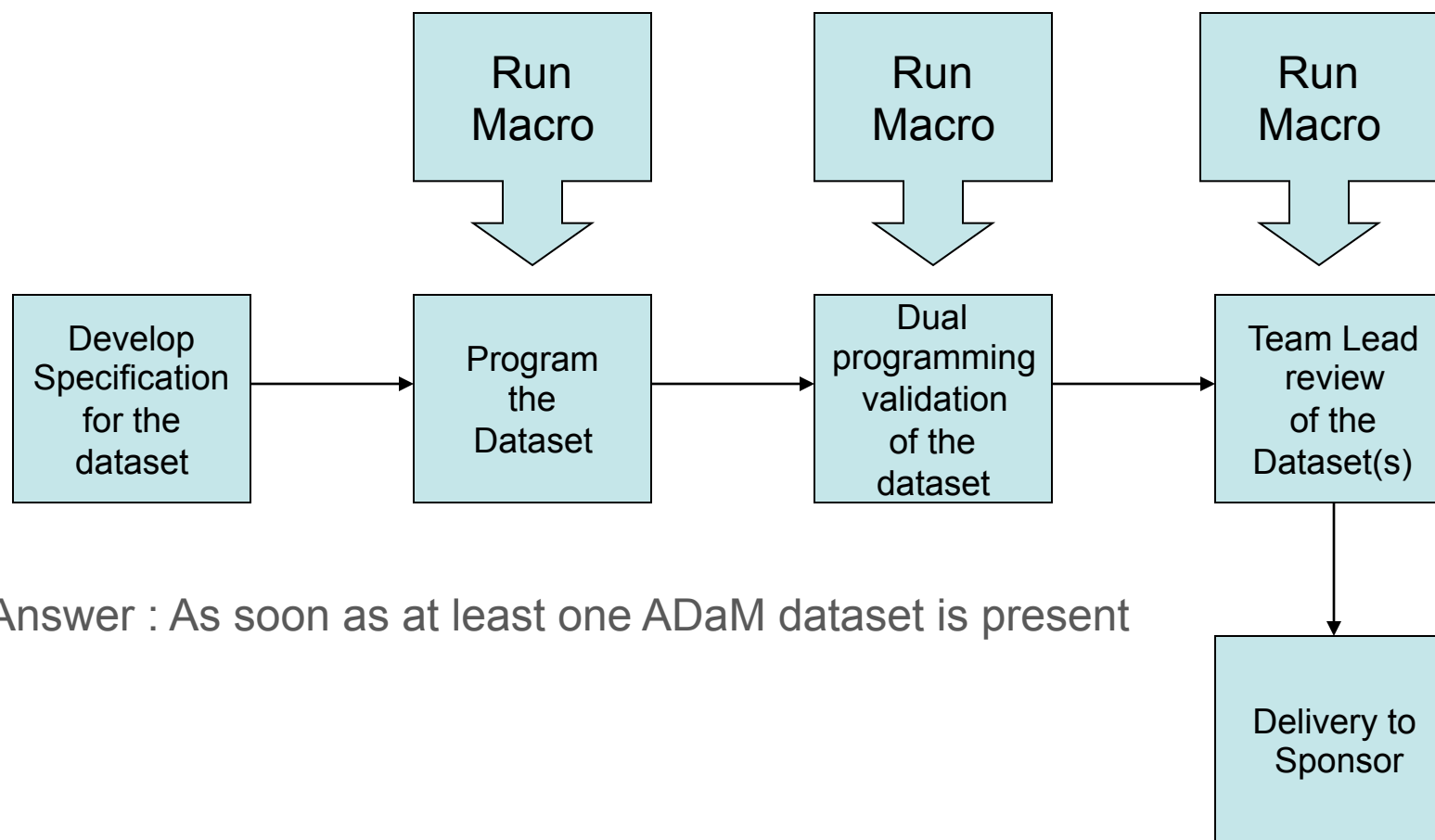
- Lack of existing electronic tool to check that studies producing CDISC standard datasets are CDISC compliant
 - Manual checking of datasets using the specification was extremely time consuming
- End goal - Create a flexible and maintainable validation tool to apply across all ADaM studies
- Uses of the macro output are three fold
 - Development
 - Training
 - Validation
 - Internal and External proof of quality

Design and Scope of the Macro

Clinical Research

- The macro has been developed generically so that it can run successfully on any ADaM study
- Series of programs/macros to run each individual check
- The individual checks within the macro are numbered according to the ADaM check specification numbering
- Utilises SASHELP data, e.g. VCOLUMN and VTABLE
 - Reduced running time if not directly accessing datasets each time

When to use it?



Answer : As soon as at least one ADaM dataset is present

- A report is produced to show:
 - ‘Check’ details, as per the CDISC specification
 - The status of each check, *Passed* or *Failed*
 - Display the failure reasons where applicable
- The output format was designed to mirror the format of the excel version of the specification
 - To ensure consistency between the two documents for traceability purposes
- The macro will help ensure consistency across studies/projects where CDISC standard ADaM datasets are created

Specification vs Macro output

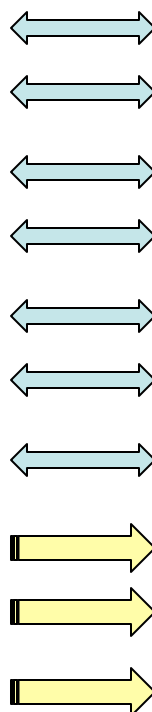
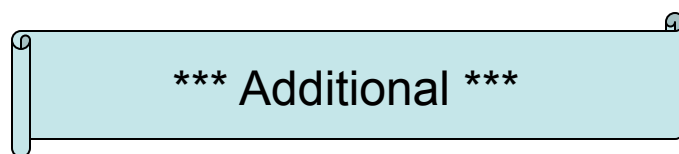
Clinical Research

CDISC ADaM Validation Checks Specification

- Check Number
- ADaM IG Section Number
- Text from ADaM IG
- ADaM Structure Group
- Functional Group
- ADaM Variable Group
- Machine-Testable Rule

ICON Validation Checks Macro Output

- Check Number
- ADaM IG Section Number
- Text from ADaM IG
- ADaM Structure Group
- Functional Group
- ADaM Variable Group
- Machine-Testable Rule
- Pass/Fail
- Reason for Failure
- Additional Information



Output – List file output

- The report was initially only outputted as dated list file

SAS System Viewer - [ADaM_v1_1_Validation_Checks_21SEP11.lst]

File Edit View Window Help

13:50 Wednesday, September 21, 2011 1

ADaM IG 1.0 Check Number	Section Number	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria	Pass/ Fail	Reason	Additional Information
1	S2.3.1	There is only one ADSL per study	ADSL	Present/ Populated	General	ADSL dataset does not exist	Passed		
								13:50 Wednesday, September 21, 2011	2
ADaM IG 1.0 Check Number	Section Number	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria	Pass/ Fail	Reason	Additional Information
2	S3	Any ADaM variable whose name is the same as an SDTM variable must be a copy of the SDTM variable, and its label and values must not be modified	ALL:SDTM	Metadata	Data Point Traceability Variables	A variable is present in ADaM with the same name as a variable present in SDTM but the variables do not have identical labels	Failed	BRTHDTC: SDTM - Date/Time of Birth ADaM - Date/Time of Birth - Char	Variables checked: AEACN, AEACNOH, AEBODSYS, AEDECOD, AEENDTC, AEENRTPT, AEENTPT, AEMODIFY, AEOUT, AEREFID, AEREL, AESEQ, AESER,

Output – Excel format

- Feedback : no more List files

Microsoft Excel - ADaM_v1_1_Validation_Checks_21SEP11.xls

File Edit View Insert Format Tools Data Window Help Adobe PDF

Type a question or search

E9 Metadata

Check Number	ADaM IG 1.0 Section Number	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria	Pass/Fail	Reason	Additional Information
1	S2.3.1	There is only one ADSL per study	ADSL	Present/Populated	General	ADSL dataset does not exist	Passed		
2	S3	Any ADaM variable whose name is the same as an SDTM variable must be a copy of the SDTM variable, and its label and values must not be modified	ALL:SDTM	Metadata	Data Point Traceability Variables	A variable is present in ADaM with the same name as a variable present in SDTM but the variables do not have identical labels	Failed	BRTHDTC: SDTM - Date/Time of Birth ADaM - Date/Time of Birth - Char	Variables checked: AEACN, AEACNTH, AEBODSYS, AEDECOD, AEENDTC, AEENRTPT, AEENTPT, AEMODIFY, AEDOUT, AEREFID, AEREL, AESEQ, AESER, AESEV, AESPID, AESTDTC, AETERM, AGE, AGEU, ARM, BRTHDTC, CMCAT, CMCLAS, CMCLASCO, CMDECOD, CMDOSE, CMDOSFRM, CMDOSFRQ, CMDOSTXT, CMDOSU, CMENDTC, CMENRTPT, CMINTPT, CMINDC, CMROUTE, CMSEQ, CMSPID, CMSTDTC, CMTRT, COUNTRY, DACAT, DAREASND, DASEQ, DASTAT, DASTRESC, DASTRESN, DASTRESU, DMDTC, DMDY, DSCAT, DSDECOD, DSSEQ, DSSTDTC, DSSTDY, DSTERM, EGCAT, EGSEQ, EGSTAT, ETHNIC, EXDOSE, EXDOSTXT, EXDOSU, EXENDTC, EXSEQ, EXSTDTC, EXTRT, FABLFL, FACAT, FAEVAL, FAOBJ, FASCAT, FASEQ, FASPID, IEACAT, IESEQ, IESPID, INVIO,

- Interpreting the output of the macro – failures!
 - Most checks show up obvious rules that are not followed in the I.G whether due to programming errors or data issues.
 - Some checks show up data inconsistencies could be allowed to fail, in the case of legacy studies
 - The use of the excel spreadsheet meant the output could be filtered and additional columns for tracking added

Clinical Research

- There are some issues that have arisen regularly on the conversion study
- In non-ADaM datasets it can be ok that a check would not pass due to the specific requirements of the data and study e.g. an SDTM variable copied to an AD and renamed
 - BUT not with ADaM
- Each check MUST be resolved either by changing the specification/dataset to match the IG rules OR by reporting the “allowed” failure in the Reviewer’s Guide
- No check is allowed to fail without investigating possible solutions

A variable is present in ADaM with the same name as a variable present in SDTM but the variables do not have identical lengths

A variable is present in ADaM with the same name as a variable present in SDTM but the variables do not have identical labels

A variable is present in ADaM with the same name as a variable present in SDTM but the variables do not have identical formats

A variable is present with the same name as a variable present in ADSL but the variables do not have identical formats

SAFFL is present and has a value that is not Y or N

Within a given value of PARAMCD, BASETYPE is populated for at least one record and is not populated for at least one record

Within a given PARAMCD for a subject, more than one record has ABLFL="Y" and BASETYPE is not present

*SDY is less than or equal to *EDY if both are non-missing

Within a given value of PARAMCD, there is more than one value of AVALC for a given value of AVAL

Findings

	A	B	C	D	E	F	G	H	I
1	ADaM Validation Check								
2									
3	Check Number	ADaM IG 1.0 Section Number	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria	Pass Fail	Reason
4									
5									
6									
57	53	S3.1	Invalid STUDYID/USUBJID combination not found in the SDTM Demographics domain.	ALL:SDTM	Value Consistency	Subject Demographics	The values of USUBJID are not present in SDTM.DM	Failed	ADEG

- **Flags up potential data issues**
 - Why does a subject have ADEG data but no DM data?
- **Pooled data**
 - Macro is restrictive on this structure of study set up
 - Raised question of whether pooled data should be pooled at SDTM or ADaM stage

Findings (2)

Chec Numb	ADaM IG 1.0 Sectio Num	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria	Pass Fail	Reason	Additional Information
149	S3.2.4	if both AVAL and AVALC are populated then there must be a one-to-one mapping	BDS	Value Consistency	Analysis Parameter Variables	Within a given value of PARAMCD, there is more than one value of AVALC for a given value of AVAL	Failed	Variables: ADDA.AVAL vs. ADDA.AVALC (for PARAMCD: RETAMT), ADHBA1C.AVAL vs. ADHBA1C.AVALC (for PARAMCD: CHOLS, CPEPTIDS, CRPS, GLUCS, HBA1CS, HDLS, HOMABS, HOMAS, INSULINS, LDLCS, NITROS, TRIGS), ADLB.AVAL vs. ADLB.AVALC (for	Variables checked: ADDA.AVAL vs. ADDA.AVALC, ADEG.AVAL vs. ADEG.AVALC, ADFA.AVAL vs. ADFA.AVALC, ADHBA1C.AVAL vs. ADHBA1C.AVALC, ADLB.AVAL vs. ADLB.AVALC, ADVS.AVAL vs. ADVS.AVALC
150	S3.2.4	if both AVAL and AVALC are populated then there must be a one-to-one mapping	BDS	Value Consistency	Analysis Parameter Variables	Within a given value of PARAMCD, there is more than one value of AVAL for a given value of AVALC	Failed	Variables: ADHBA1C.AVALC vs. ADHBA1C.AVAL (for PARAMCD: CHOLS, CPEPTIDS, GLUCS, HBA1CS, HDLS, INSULINS, LDLCS, NITROS, TRIGS)	Variables checked: ADDA.AVALC vs. ADDA.AVAL, ADEG.AVALC vs. ADEG.AVAL, ADFA.AVALC vs. ADFA.AVAL, ADHBA1C.AVALC vs.

- This is a common failure that has arisen during development of the BDS datasets
 - Depends greatly on the population of - -STRESN and - - STRESC variables
 - Update the numeric AVAL for consistency?
 - AVAL=input(strip(AVALC),best.);

Findings (3)

Check Number	ADaM IG 1.0 Section Num	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria	Pass Fail	Reason
155	S3.2.4	Whenever there is more than one definition of baseline, the BASETYPE column is required. BASETYPE identifies the definition of baseline that corresponds to the value of BASE in each row. There is only one BASE column, and only one column for each qualifying function	BDS	Present/ Populated	Flag Variables	Within a given PARAMCD for a subject, more than one record has ABFL="Y" and BASETYPE is not present	Failed	More than one record has ABFL=Y and BASETYPE is not present

- Flags up issues in the dataset surrounding baseline derivations
 - If there is more than one ABLFL flag per subject and parameter, then there should be a BASETYPE flag
 - E.g. Randomization vs Screening

- All such cases must be reported in the Reviewer's Guide!

Check Numb	ADaM IG 1.0 Section Number	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria
98	S3.2.3	*SDY less than or equal to *EDY if both are non-missing	BDS	Value Consistency	Timing Variables	A variable with a suffix of SDY has a value greater than a value of a variable with the same root and a suffix of EDY, and both variables are
99	S3.2.3	*STDY <= *ENDY if both are non-missing	BDS	Value Consistency	Timing Variables	A variable with a suffix of STDY has a value greater than a value of a variable with the same root and a suffix of ENDY, and both variables are populated

- e.g. EGSDY
5
- EGEDY
4

- Such an example has occurred on legacy study conversions,
- It is a failure and should have been resolved
- In this case the data was final and no correction could be made so it was reported in the Reviewer's Guide as a failure due to incorrect data.

- Utilisation of the macro resulted in a more in-depth thought process regarding the interpretation of the IG and the way we develop ADaM datasets
- A few questions were raised about the IG and the validation checks specified
- Results in feedback to the relevant CDISC teams

- TRTP and TRT01P for parallel studies, what is the relevance of both variables for this study design
- *BASETYPE is populated for at least one record within a parameter then it must be populated for all records within that parameter*
 - Question use of the BASETYPE variable, all parameters must be populated, but does it make sense to populate BASETYPE for a pre-baseline record, i.e. screening?
- *PARAM and PARAMCD are present and have a one-to-one mapping*
 - Not just a one to one mapping – ensure when checking the one-to-one mapping, it is an appropriate 1-1 mapping, i.e. not “Pulse Rate (beats/min)” -> “missing” for example!

- LVOTFL

Check Number	ADaM IG 1.0 Section Numb	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria
175	S3.2.6	A maximum of one record within a parameter timepoint can have LVOTFL=Y	BDS	Value Consistency	Flag Variables	There is more than one record with LVOTFL equal to Y for a given value of PARAMCD

- The “Text from the IG” column indicates looking at a parameter time point.
- Machine readable text is clearer in stating there cannot be more than one record within a PARAMCD with LVOTFL=‘Y’ but the text from the IG is misleading in mentioning time point

- What can be done to minimise errors before this stage?
- Programming specification checks and a “Shell” program
 - Checks on the specifications to enable a more efficient process
 - If this is ok the programming errors are minimised
 - Use of a shells program which utilises the specifications to build a ‘dummy’ dataset to populate with data

- For this tool to be used across all ICON ADaM studies, it was essential the macro was fully validated to ensure confidence in the results of this tool
- Up front validation plan was designed, and suitable test data selected
- The validation consisted of three parts
 - Validation tests
 - Portability tests
 - Code review

- **Validation Tests**
 - Each check was validated using test data
 - To determine whether or not the check correctly identifies data that fails the check
- **Portability Tests**
 - Each check was validated using data from a different sponsor than that initial data used for the development of the macro
 - To ensure the macro is portable to data from other sponsors and indications.
- **Code Review**
 - In addition to testing each of the individual checks in the macro, the SAS code and log was reviewed to check for programming consistency with guidelines
 - To ensure there are no warnings, errors
 - To ensure sufficient commenting for future edits / updates by another programmer to the initial developing programmer

- At the end of each validation stage, a validation report was compiled detailing the following:
 - Each run of the process
 - The checks that passed and failed validation at each run
 - Once completed, sign off indicating the macro was fit for purpose

PASSED

- The macro has been approved by developer, validator and senior management and can now be utilised across all ADaM studies as standard within ICON

But what about OpenCDISC?

- OpenCDISC is another tool to use for checking ADaM datasets
 - Version 1.2 of the OpenCDISC validation tool, focused on adding support for ADaM validation checks
 - It utilizes the first version 1.0 of the validation checks for ADaM
- Advantages of the ADaM macro vs OpenCDISC ADaM checks
 - Can be used on an ongoing basis, prior to finalisation of all datasets
 - No need for a define.xml
 - Runs directly on SAS data files, no need to convert to xpt format
 - Comprehensible output document pointing directly to the issue
 - Macro uses the latest version of validation checks v1.1
- OpenCDISC continues to be utilised for Controlled Terminology and formatting checks (i.e. spec vs data checks)

- ADaM Validation checks specification release is a positive step in creating a summary of key compliance issues to check in a dataset
- Designed to be machine-readable, and it certainly is!
- ICON worked with this to create a fully up to date macro based on the recommended checks advised by CDISC
- Uses most recent set of checks released by CDISC
- Excellent validation tool to work with to ensure our continued ADaM compliance

