

ADaM Validation and Integrity Checks

Louise Cross, ICON Clinical Research, Marlow, UK

ABSTRACT

The release of the CDISC ADaM Validation Checks Version 1.0, dated September 20, 2010 and updated version 1.1, dated 21JAN2011, is a positive step in highlighting sense checks to perform on ADaM data checks to ensure compliance with the model. At ICON, this document has resulted in the creation of a key validation tool, especially useful when working on large conversion projects. We have created a standard SAS macro to use across any ADaM project, which performs these checks and outputs any issues to be addressed. When working on a large conversion project covering over 30 studies, this became a valued tool to ensure consistency and accuracy.

INTRODUCTION

ICON are now seeing an increased demand for conversions of legacy studies to CDISC, while many new studies are setup which adhere to the CDISC standards. The knowledge and expertise gained from implementing ADaM and SDTM since the early days of its introduction, means we have a great knowledge base to utilize on such projects. When working on a single study it can be straightforward to ensure ADaM standards and compliance are tracked, however when working on a 30 legacy study project converting to ADaM with over 300 ADaM datasets to create, tools must be developed to streamline the validation process.

Enhanced documentation and streamlined validation techniques are vital in projects such as this, to maintain a high standard of ADaM deliverables. The ADaM model and implementation guide provide a sound basis for the desired structure of all ADaM datasets, however the validation checks document released by CDISC has proved a key document to check ADaM rules are adhered to and hence, ensures datasets are ADaM compliant. Within ICON, this has led to the creation of a macro which performs all of the checks detailed in the document, and outputs a clear and concise document highlighting any potential non-compliance. This paper will outline the basics of the macro, processes undertaken to carry out these checks, common findings and interpretations of the macro output.

RATIONALE FOR THE DEVELOPMENT OF THIS VALIDATION CHECKS MACRO

While working on a large conversion study within the ICON Biostatistics department, it became increasingly apparent that there was the need for more stringent and streamlined checks to take place when checking the conversion of legacy studies to ADaM datasets. It was clear there was no pre-existing electronic tool to check that studies producing CDISC standard ADaM datasets are compliant with the rules outlined in the implementation guide.

The OpenCDISC validate tool is a useful program to aid integrity checks, however this focuses more on the define vs. dataset checks, so does not check the content of the datasets. It was found that this was not a suitable level of validation to ensure ADaM compliance, therefore the macro was developed. It concentrates more on the structure of ADaM datasets and the adherence of the data to the rules specified in the ADaM implementation guide. In addition, the OpenCDISC validate was developed based on V1.0 of the validation checks document, hence, any datasets completed in 2011, were not utilising the latest information and checks provided by CDISC, released in v1.1 of the validation checks document. Therefore, a macro was developed to apply across all studies that were part of the conversion project and furthermore, it can be applied to any future ADaM study undertaken.

As work progressed, consistency, integrity and compliance of the ADaM model was paramount and as the project was so large, a time efficient validation process was developed and implemented. The validation checks document released by CDISC provided an excellent foundation to build a program to ensure the conversions were fully compliant and hence this was why the macro was created.

PRACTICALITIES OF THE VALIDATION CHECKS MACRO

Initially, the macro was written based on version 1.0 of the validation checks macro released in December 2010. However, upon the release of the ADaM check specification version 1.1, in January 2011 the macro was updated to

PhUSE 2011

reflect any additional checks. The macro utilises the SASHELP tables, VCOLUMN and VTABLE in conjunction with the ADaM datasets to perform the checks. We took advantage of the SASHELP tables, as this reduced the need for such frequent interrogation of the datasets, which sometimes could be quite large, resulting in longer run times. The macro contains series of SAS code to run each individual check as detailed in the validation checks document. It is to be run interactively in SAS, and the only input requirement of the macro is that the libnames must be specified to point to the location of both SDTM and ADaM datasets. An output report is produced showing the status of each check, displaying the variables and datasets checked as part of the validation, and details of why a check has passed or failed. The macro creates an excel spreadsheet which mirrors the format of the CDISC validation checks spreadsheet, with additional columns to easily identify any issues. The individual checks within the macro are numbered according to the CDISC ADaM validation check specification; this will then make any modifications to the macro easier when CDISC release new versions of the check specification. CDISC have noted that any updates to the validation checks will not impact the numbering, hence we know that check 5 will always remain check 5. If a check was deleted then that check number would not be used again.

PROCESS OF VALIDATION CHECKS MACRO

The macro assumes the SDTM dataset to be contained in one SAS library, and the ADaM datasets to be contained in separate SAS library. Once these are defined in libname statement, the macro can be run interactively across the SDTM and ADaM libraries. The macro runs through a series of checks, and within each check there are numerous loops to run the code over the ADaM database.

Example:

Check 2

This is a snapshot of the ADaM validation checks spreadsheet taken from the CDISC website.

	A	B	C	D	E	F	G
1	Check Numb	ADaM IG 1.0 Section Numb	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria
3	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

The macro uses the SASHELP tables to check the consistency of all variables that are present in SDTM and ADaM, and checks the attributes that are associated with this.

The macro creates two outputs. An excel spreadsheet and a list file.

Originally, the first version of the macro created a SAS list file from SAS, containing all the checks carried out and detailing the pass/fail criteria observed. The output is as below

ADaM IG 1.0	ADaM	ADaM	ADaM	Machine-Testable	Pass/Fail	Reason	Additional Information		
Check Section	Text from ADaM IG	Structure Group	Functional Group	Failure Criteria	Fail				
Number Number									
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 LBREASND: SDTM - Reason Test Not Done ADaM - Reason Text Not Done	Variables checked: AEACN, AEACNOTH, AEBODSYS, AEDECOD, AEENDTC, AEENRTPT, AEENTPT, AEMODIFY, AEOUT, AEREFID, AEREL, AESEQ, AESER, AESEV, AESPID, AESTDTC, AETERM, AGE, AGEU, ARM, BRTHDTC, CMCAT,

However users found this was not very easy to read and use, despite containing a lot of information. So in addition to this, an excel spreadsheet was created, which would output this data to a spreadsheet and which proved a more versatile tool. This could be annotated with investigatory comments and could be filtered and hence

PhUSE 2011

can form into a very useful tracking document for resolving issues found. This format has proved more user-friendly; it mirrors the list file and is also the same format as the validation checks spreadsheet that CDISC provides.

Check	ADaM	Text from ADaM IG	ADaM	Functional	ADaM	Machine-Testable Failure	Pass	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 LBREASND: SDTM - Reason Test Not Done ADaM - Reason Text Not Done	Variables checked: AEACN, AEACNOH, AEBODSYS, AEDECOD, AEENDTC, AENERT, AENRPT, AEMODIFY, AEOU, AEREFID, AEREL, AESEG, AESAESEV, AESPID, AESTDTC, AETERM, AGE, AGEU, ARM, BRTHDTC, CMCAT, CMCLAS,

The macro adds the last three columns, which contain the outcomes of the macro. The Passed column contains whether a check has passed or failed, the reason column contains specifics on a failure (if present) and the additional information column contains details of the data that has been checked. This proves to be a useful tool when checking back against the ADaM specifications.

INTERPRETING THE VALIDATION CHECKS MACRO

Most check failures show up obvious rules that are not followed from the implementation guide. Much of this is due to misinterpretation of the implementation guide and the macro raises these issues so that the programmer can update the ADaM datasets to ensure a consistent and compliant approach across all datasets and studies. The validation check output spreadsheet can be used to track these issues so when an issue is fixed this can be signed off in the spreadsheet. Much in the same way a comments spreadsheet can be received from a senior or sponsor review of outputs, the spreadsheet created by the macro can develop into a useful tracking tool.

There may be instances where a check that has failed, can not be fixed. This is particularly true in the case of legacy studies, which have CRF designs that did not consider CDISC standards, as they were created prior to CDISC being developed. Once the ADaM datasets are finalized, any outstanding issues highlighted by the macro, would be discussed with the sponsor and potentially form part of the Reviewer's Guide if no resolution can be reached. Some checks show up data inconsistencies which can be allowed to fail, especially for the case of legacy studies where the data cannot be queried.

Sometimes the logic being checked is two way so there are two checks for the same thing

e.g.

A check for a start date being less than or equal to an end date

AND

A check for an end date being more than or equal to a start date

e.g. EGS DY EGED Y
5 4

This has clearly failed the check and should be 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.

ASSUMPTIONS AND LIMITATIONS OF THE MACRO

The macro can only be used for studies that create SAS datasets to CDISC standards. The macro must be run within the study development environment; the autoexec for the study will assign the libnames and global macro variable settings required for the macro to run. The macro is programmed generically and utilises the SASHELP tables. Checks are based on the rules in the ADaM Implementation Guide v1.0 and the current ADaM check specification v1.1, as released by CDISC. The macro will be updated when CDISC release new checks/updates.

One main limitation currently, is that pooled data can cause an issue with one of the checks in the macro. The check below cross matches subjects in the SDTM DM dataset with data in the ADaM ADSL dataset.

1	Check Numb	ADaM IG 1.0 Section Numb	Text from ADaM IG	ADaM Structure Group	Functional Group	ADaM Variable Group	Machine-Testable Failure Criteria
52	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

PhUSE 2011

If the ADSL dataset is a pooled dataset across two or more studies, check 53 would fail. This particular issue arose on one of our studies, where ADSL contained subjects from two stages of a study, main and extension, however the SDTM data only referenced extension subjects. Hence, subjects who did not proceed to the extension would result in a failure of this check, even though upon further investigation this was because the subjects. This is certainly something to work on for future development, however currently it provides a useful alert to ensure the data is as expected.

VALIDATION CHECKS VS OPEN CDISC

The OpenCDISC Validate tool is the only other tool available for checking ADaM datasets. This checks the ADaM data from a define perspective, hence is not checking the structure of the data. If a define.xml is not created for a study, then the OpenCDISC Validate tool cannot be used to check ADaM datasets. Once created, the define.xml is run through the OpenCDISC validate tool. This produces a report and lists any errors found in reconciling the dataset specifications with the ADaM datasets. Interpreting the OpenCDISC error report can sometimes be time consuming as the output can be vague. The OpenCDISC is excellent for checking consistency between the define.xml, the datasets and the specification, however it does not interrogate the data enough to provide a sufficient ADaM compliance check, which is why the validation checks macro is such an essential tool for this. We are also in a great position to swiftly create a validated updated version of the macro if and when CDISC release a new version of the ADaM check specifications.

FUTURE DEVELOPMENTS

Work is ongoing to develop this macro based on user feedback. A key future development for this is to build in an audit trail for this. Currently the macro outputs new versions of documents with the run date attached in the file name. However, in a development of the macro currently underway, the first time the macro is run, the excel sheet will be created and comments can be added by the reviewer. When the next run of the macro is performed, the macro will import the existing spreadsheet into SAS, so any comments regarding non resolvable issues can still be logged, and appends any additional checks to this spreadsheet. This is essential for cases where issues cannot be resolved due to legacy study data issues, and hence the same issues will re appear in every run of the macro. The document can then be used to ensure the reviewers guide is accurate in detailing what issues remain in the data and cannot be resolved.

As previously mentioned, pooled data can result in a fail for check 53. This is something to consider in future version of the macro, to enable this to work for pooled studies.

CONCLUSION

In summary, the release of the validation checks document has proved an excellent tool to build upon. It has enabled us to create the validation checks macro. Implementing this macro across ADaM datasets results in studies being delivered quicker and more consistently, thus reducing timelines which is obviously paramount with any study. Ongoing development of the macro will continue to occur, as the more studies this macro is applied to the more unique cases may occur, which will result in enhancement to this validation tool.

REFERENCES

Please see the CDISC website to download the latest validation checks documents.
www.cdisc.org

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Louise Cross
ICON Plc,
2 Globeside,
Globeside Park,
Marlow
Phone: +44 (0)1628 493516
Email: louise.cross@iconplc.com

PhUSE 2011

Web: www.iconplc.com

Brand and product names are trademarks of their respective companies.