

Poster: PO-20**Proc CDISC: Implementation and Assessments**

Sheetal Nisal, Independent Consultant
Shilpa Eduganti, Independent Consultant

Abstract:

With the rapid developments in the industry standards like CDISC and with demanding FDA requirements, CROs and sponsor companies are trying to implement various data models developed by clinical data interchange standards consortium (CDISC). Proc CDISC is one of the important procedures which provide a way to import and export an XML document in CDISC ODM (operational data model) and validates the SDTM (Study Data Tabulation model) requirements. It provides functionality and features that is based on specific CDISC model. This paper introduces and describes application of proc CDISC with the context of SDTM for different domains, and illustrates the differences in output produced by proc CDISC when applied to different type of metadata.

Introduction:

Clinical Data Interchange Standards Consortium or CDISC has come up with different standards to facilitate efficient exchange of information between industry and regulatory bodies. Food and Drug Administration or FDA has recommended use of CDISC submission data standards which are based on study data tabulation model (SDTM). FDA developed a clinical data repository, called JANUS, which provides a central access to standardized data and creates an integrated platform for tools used in analysis and review. Because of benefits to the industry in communicating and sharing standardized data and FDA requirements, pharmaceutical industry are mainly rapidly adopting the SDTM for submission purpose.

SDTM is a very robust data model. However, considering its purpose to support the industry needs, it is very generic in nature and allows the sponsor companies to interpret the SDTM requirements in their own manner. This results into lot of interpretations of SDTM guidelines. In order to ensure that industry interprets SDTM appropriately and yet use it to represent their submission data in most efficient manner, industry felt a need to seek guidance about compliance of SDTM.

SAS® Institute responded to this industry need by developing a procedure called 'proc CDISC'. This procedure currently supports import and export of data as per operational data model (ODM- version 1.2) and helps ensuring compliance of data with SDTM (version 3.1). This paper mainly focuses on CDISC SDTM compliance and provides user insight and resources on how to optimize use of proc CDISC for SDS compliance. With this SDTM model proc CDISC performs data content validation on SAS data set to validate the compliance requirements of data. Through the ODM option, the CDISC procedure in SAS supports the import, export and processing of XML documents and

content that are in CDISC defined format. SAS institute update PROC CDISC in March 2006 version (2.15.52).

SDTM Option in Proc CDISC:

For successful application as well as implementation of proc CDISC it is important to understand SDTM model completely. CDISC SDTM defines standard structure for study data tabulation data sets that are submitted as a part of product application to FDA. CDISC SDTM includes several defined domains that grouped within broad categories. CDISC-SDTM guidelines categorize the study end point data into standardized domains. These domains represent certain types of data such as demog, or exposure, or concomitant medication etc. SDTM classifies domains into certain classes. These classes are defined based on the purpose of data. Domains are mainly divided in three different classes, which is Interventions, Event and Findings. Apart from these, special purpose domains are specified in the SDTM. All domains fall in either of the classes. Although SDTM version 3.1.2 comes up with quite a few additional domains related to Pharmacokinetic data, and microbiological data, CDISC procedure confines validation of certain domains from SDTM version-3.1.

Domain - The CDISC procedure currently supports 15 of the 23 SDTM domains. Each CDISC domain specifies unique two character domains code. Each domain is a collection of observations that are common to specific subject. These include:

Class: Special purpose domains
Demographic characteristics (domain- DM)
Comments (domain- CO)

Class: Interventions
Concomitant Medications (domain- CM)
Exposure (domain- EX)
Substance Use (domain- SU)

Class Events:
Adverse Events (domain- AE)
Disposition (domain- DS)
Medical History (domain- MH)

Class Findings:
ECG Test Results (domain- EG)
Inclusion/Exclusion Exception (domain- IE)
Laboratory Test Results (domain- LB)
Physical Examinations (domain- PE)
Questionnaires (domain- QS)
Subject Characteristics (domain- SC)
Vital Signs (domain- VS)

Based on the support documentation of proc CDISC, the SDTM option in this procedure has capabilities to perform the following checks:

- Verifies that all required variable are present in the data set.
- Reports as an error if any variable in the data is not defined in the domain.
- Reports a warning if any expected domain variable is not present in dataset.
- Verifies that all domain variables are of the expected data type & proper length.
- Note any permitted variable which are not in the data set.
- Notes all required variable fields if they contain missing value.
- Detects the conformance of all ISO-8601 specification assigned values; including date, time, date time, duration, and interval types
- Notes correctness of yes/no and yes/no/null responses.
- Detects any domain variables that are assigned a controlled terminology specification by the domain and do not have a format assigned to them.

Validating a CDISC SDTM:-

SAS data set validation mainly divided in three parts.

- Proc CDISC statement specifies MODEL= SDTM
- SDTM statement to specify the SDTM version number
- DOMAINDATA statement to specify the SAS data set to be validated, two charter domain code and domain model type (category).

Illustrations of proc CDISC:

Following illustration explains how proc CDISC validates the compliance of data with SDTM model. Consider the concomitant medication data called 'cmdat'. A brief snap shot of this data is provided below. Please note that the following snap shots of data doesn't represent the complete data for CM domain and the variables used below are for illustration purpose only:

STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMDOSE	CMDOSTXT	CMROUTE
001	CM	001-001	1	TRT1	.		
001	CM	001-001	2	TRT2	1	1	PO
001	CM	001-001	3	TRT3	1	1	PO
001	CM	001-001	4	TRT4	1	1	PO
001	CM	001-001	5	TRT5	1	1	PO
001	CM	001-001	6	TRT6	1	1	PO

Following code can be used to validate the concomitant data.

```
proc cdisc model = sdtm;
  sdtm sdtmversion = "3.1";
  domaindata data = cm domain = cm category = interventions;
run;
```

As you may note, above code specifies the category as 'intervention' as concomitant medication is an intervention type domain. Running this code provided following output:

NOTE: Variable CMSPID is permitted in this domain(CM), but is not present.

NOTE: Variable CMSCAT is permitted in this domain(CM), but is not present.
 NOTE: Variable CMOCCUR is permitted in this domain(CM), but is not present.
 NOTE: Variable CMCLAS is permitted in this domain(CM), but is not present.
 NOTE: Variable CMCLASCD is permitted in this domain(CM), but is not present.
 NOTE: Variable CMDOSTOT is permitted in this domain(CM), but is not present.
 NOTE: Variable CMDOSRGM is permitted in this domain(CM), but is not present.

This indicates that above permissible variables (CMSPID, CMSCAT, CMOCCUR, CMCLAS, CMCLASCD, CMDOSTOT, and CMDOSRGM) are missing from the data.

Now let's consider that above data snap shot does not have the variable 'cmseq', which is a required variable.

STUDYID	DOMAIN	USUBJID	CMTRT	CMDOSE	CMDOSTXT	CMROUTE
001	CM	001-001	TRT1	.		
001	CM	001-001	TRT2	1	1	PO
001	CM	001-001	TRT3	1	1	PO
001	CM	001-001	TRT4	1	1	PO
001	CM	001-001	TRT5	1	1	PO
001	CM	001-001	TRT6	1	1	PO

If the proc CDISC code is executed on the above data it gives following output:

```
ERROR: Required parameters not contained on DOMAINDATA(Domain=CM) statement.
       Required parameter CMSEQ not present
```

In this example, proc cdisc gives error since the input data does not contain the required variable 'cmseq', which reads the sequence number of concomitant medication.

Now consider the case, when the required variable has missing record. In the following data, the variable CMSEQ has a missing value in the first record.

STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMDOSE	CMDOSTXT	CMROUTE
001	CM	001-001	.	TRT1	.		
001	CM	001-001	2	TRT2	1	1	PO
001	CM	001-001	3	TRT3	1	1	PO
001	CM	001-001	4	TRT4	1	1	PO
001	CM	001-001	5	TRT5	1	1	PO
001	CM	001-001	6	TRT6	1	1	PO

In such case if the required variable has a missing record, which does not adhere to compliance requirements of SDTM, proc cdisc gives following error:

```
ERROR: Required variable CMSEQ has a MISSING value in observation 1.
```

As shown above, if variable CMSEQ has a missing record, proc cdisc gives error. This is because a required variable like CMSEQ can not have missing value as per the SDTM compliance requirements.

Consider the following meta-data of Exposure dataset.

Variable	Type	Len	Format	Label
STUDYID	Char	200	\$200.00	STUDY IDENTIFIER
DOMAIN	Char	10	\$10.00	DOMAIN ABBREVIATION
USUBJID	Char	25	\$25.00	UNIQUE SUBJECT IDENTIFIER
EXSEQ	Num	8	BEST8.	SEQUENCE NUMBER
EXGRPID	Char	200	\$200.00	GROUP ID
EXSPID	Char	5	\$5.00	SPONSOR-DEFINED IDENTIFIER
EXTRT	Char	50	\$50.00	NAME OF ACTUAL TREATMENT
EXCAT	Char	200	\$200.00	CATEGORY FOR TREATMENT
EXSCAT	Char	100	\$100.00	SUBCATEGORY FOR TREATMENT
EXDOSE	Num	8	BEST8.	DOSE PER ADMINISTRATION
EXDOSTXT	Char	20	\$20.00	DOSE DESCRIPTION
EXDOSU	Char	10	\$10.00	DOSE UNITS
EXDOSFRM	Char	50	\$50.00	DOSE FORM
EXDOSFRQ	Char	50	\$50.00	DOSING FREQUENCY PER INTERVAL
EXDOSTOT	Num	8	BEST8.	TOTAL DAILY DOSE USING EXDOSU
EXDOSRGM	Char	10	\$10.00	INTENDED DOSE REGIMEN
EXROUTE	Char	50	\$50.00	ROUTE OF ADMINISTRATION
EXLOT	Char	50	\$50.00	LOT NUMBER
EXLOC	Char	40	\$40.00	LOCATION OF DOSE ADMINISTRATION
EXTRTV	Char	8	\$8.00	TREATMENT VEHICLE
EXADJ	Char	200	\$200.00	REASON FOR DOSE ADJUSTMENT
TAETORD	Num	8	8	ORDER OF ELEMENT WITHIN ARM
EXSTDTC	Char	200	\$200.00	START DATE/TIME OF TREATMENT
EXSTDY	Num	8	BEST8.	STUDY DAY OF START OF TREATMENT
EXENDY	Num	8	BEST8.	STUDY DAY OF END OF TREATMENT
EXDUR	Char	25	\$25.00	DURATION OF TREATMENT
EXTPT	Char	20	\$20.00	PLANNED TIME POINT NAME
EXTPTNUM	Num	8	8	PLANNED TIME POINT NUMBER
EXELTM	Char	20	\$20.00	PLANNED ELAPSED TIME FROM REFERENCE POINT
EXTPTREF	Char	200	\$200.00	TIME POINT REFERENCE

When proc CDISC is run on this exposure data, it gave following warning.

WARNING: Variable EXENDTC is expected in this domain(EX), but is not present.

As shown above, variable EXENDTC represents treatment end date/time and it is an expected variable as per version 3.1. This variable can have missing records but needs to be populated in the data. Since the exposure data used in the above example does not have this variable populated, proc CDISC gave above warning.

Above illustrations show that proc CDISC can be a good tool to ensure that the data meets minimum necessary requirements of SDTM compliance as per the requirements of version 3.1. As a minimum requirement, any SDTM compliant data must have necessary required, and expected variables and the data records should adhere the purpose of such variables. Proc CDISC goes beyond simply the types of variables and also assesses the meta data requirements in terms of variable labels, and formats.

Conclusion:

After release of proc CDISC by SAS® Institute, CDISC SDTM team released a newer version of SDTM, which had many changes including addition of few new domains. This limited the utility of proc CDISC to validation of CDISC based data as defined in version 3.1. SDTM specific options of proc CDISC are based on certain specific compliance requirements as defined by SDTM committee of CDISC. In a broader sense these requirements hold true even for newer version of SDTM. Some key changes done in newer version of SDTM include, i) addition of more domains, and ii) changing the status of certain variables from expected to permissible. Although use of proc CDISC for validation of SDTM -3.1.2 based data can give inaccurate results, an analyst doing such validation can easily custom build the checks to ensure such compliance with newer version. Proc CDISC provided a good guideline to industry about how to go about ensuring the compliance of data with SDTM model by following SDTM principles and comparing meta data and its content requirements as per SDTM guidelines. For any analyst ensuring SDTM compliance, proc CDISC can be a good starting point.

For detailed assessment of compliance of data with SDTM-3.1.2 requirements, an analyst can explore many open source tools available in the market. After appropriate validation of such tools, SDTM compliance of a data can be achieved in a fairly accurate manner. Alternatively, if a user has a thorough understanding of SDTM and requirements of this model, and if the user can interpret and implements this model appropriately, then the user can also build custom checks to ensure compliance of data with SDTM.

Reference

- Study Data Tabulation Model (SDTM) Version-3.1, and Version-3.1.2 as outlined in Study Data Tabulation Model User Guide (SDTMIG). This is provided by SDTM committee of Clinical Data Interchange Standards Consortium (CDISC).
- Proc CDISC documentation as provided by SAS® Institute.
<http://support.sas.com/documentation/cdl/en/cdisc/60755/HTML/default/viewer.htm#a003070352.htm>
- Susan J. Kenny, Michael A. Litzsinger (2005), “Strategies for Implementing SDTM and ADaM Standards”. Proceedings of the PharmaSUG 2005 conference.

Contact Information: Your comments and questions are valued and encouraged. Author can be contacted at sheetalnisal@gmail.com , and at shilpaedupganti@gmail.com

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.