

## Using PROC CDISC

Chauthi Nguyen, GlaxoSmithKline, King of Prussia, PA

### ABSTRACT

The paper will present the use of PROC CDISC to convert SAS® data sets in Clinical Trials of different domains into CDISC data sets in XML format. The new version of PROC CDISC will be used for the task. This paper is written mainly for SAS users who are new to PROC CDISC.

### INTRODUCTION

The SAS Institute updated PROC CDISC in March 2006 (version 2.15.52). This procedure can be downloaded for use in Windows, UNIX, and z/OS operating environments for SAS V8.2 and later. It currently supports 15 of the 23 CDISC Study Data Tabulation Model (SDTM) domains [1].

Why do we need to use PROC CDISC? It provides a way to export and import an XML document in CDISC ODM format based on SDTM standards. To export an XML document is to write a SAS data set to an XML file. To import an XML document is to read an XML file into a SAS data set. How do we use PROC CDISC? This paper demonstrates the use of PROC CDISC in the Windows environment to export and import an XML document. It also demonstrates a new feature in the PROC CDISC for data validation on SDTM V3.1 data sets. The discussion of how to map legacy data sets into SDTM 3.1 data sets and how to view the XML files is not within the scope of this paper.

What will work and what will not work when using PROC CDISC from a user's standpoint?

### USING PROC CDISC

#### 1. CHECKING THE CLINICAL SAS DATA SETS FOR THEIR COMPLIANCE TO THE SDTM 3.1 FORMAT

A sample of Medical History data set is used:

Obs	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM
1	abc123456	MH	abc123456.000023	1	CHAGAS DISEASE
2	abc123456	MH	abc123456.000226	1	LEFT VENTRICULAR EJECTION FRACTION DECREASED
3	abc123456	MH	abc123456.000227	1	LEFT VENTRICULAR EJECTION FRACTION DECREASED
4	abc123456	MH	abc123456.000227	2	MITRAL VALVE PROLAPSE
5	abc123456	MH	abc123456.000227	3	MITRAL REGURGITATION

  

Obs	MHDECOD	MHBODSYS	VISIT	VISITNUM	MHENRF
1	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
2	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
3	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
4	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
5	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR

Sample SAS code for data content checking

```
Libname in 'C:\nesug2006\data';

PROC CDISC MODEL = SDTM;
SDTM SDTMVersion = "3.1";
DOMAINDATA data = in.mh
           domain = MH
           category = events;
RUN;
```

There is no output SAS data set created with MODEL = SDTM. Check the SAS log for any NOTES, WARNING, or ERROR message. Entry to the parameter is not case sensitive.

Category in SDTM 3.1 for Medical History

Input SAS data set to be checked

A NOTE in the SAS log will be generated for any permitted variable that is not in the input data set:

```
NOTE: Variable MHGRPID is permitted in this domain(MH), but is not present.
NOTE: Variable MHREFID is permitted in this domain(MH), but is not present.
...
NOTE: Variable MHDY is permitted in this domain(MH), but is not present.
NOTE: DomainData MH data content is valid.
```

An ERROR message will be generated for any variable in the data set that is not defined in the SDTM V3.1. Sample log when variable `__FORMOID` exists in the MH input data set:

```
ERROR: Variable __FORMOID is not defined in this domain (MH) and must be
dropped/deleted.
ERROR: SDTM-compliant datasets may contain only variables defined in the domain
(MH).
NOTE: The SAS System stopped processing this step because of errors.
```

## 2. EXPORTING A CLINICAL STANDARD SAS DATA SET TO XML FILE:

**STEP 1:** Define required metadata for the XML file (to be included in the XML file before the Clinical data records).

```
data odm;
  ODMVersion      = "1.2";
  fileOID        = "2006-06-20 Transfer";
  FileType       = "Snapshot";
  Description    = "MH SDTM v 3.1";
run;
data study;
  StudyOID       = "ABC123456";
run;
data globalvariables;
  StudyName      = "Study ABC123456";
  StudyDescription = "ABC123456 - For testing";
  ProtocolName   = "ABC123456";
run;
data metadataversion;
  MetadataVersionOID = "SDTMv3.1";
  Name = "Metadata Version Name";
run;
```

**STEP 2:** Create required metadata variables for each Clinical data record.

```
data mh;
  length __STUDYOID __METADATAVERSIONOID __STUDYEVENTOID
         __STUDYEVENTREPEATKEY __SUBJECTKEY __FORMOID __FORMREPEATKEY
         __ITEMGROUPOID __ITEMGROUPREPEATKEY __TRANSACTIONTYPE $100.;
  set sdm.mh; /* Valid SDTM input data set */
  retain  __STUDYOID          "abc123456"
         __METADATAVERSIONOID "SDTMv3.1"
         __STUDYEVENTOID     "Medical History"
         __STUDYEVENTREPEATKEY " "
         __FORMOID           "FD.MyForm"
         __FORMREPEATKEY     "FD.MyForm"
         __ITEMGROUPOID      " "
         __ITEMGROUPREPEATKEY " "
         __TRANSACTIONTYPE   "Insert";
```

PROC CDISC in **step 3** will not work if the data set does not contain these derived metadata variables.

```

__SUBJECTKEY = trim(usubjid) || '.' || left(mhseq);
run;
proc sort data = mh;
  by __SUBJECTKEY;
run;

```

**\_\_SUBJECTKEY** is a derived variable to uniquely identify a Clinical data record in the given data set.

**STEP 3:** Have a filename statement for the output XML file and invoke PROC CDISC to export the Clinical SAS data set to the XML file.

```

filename odmxml 'C:\nesug2006\output\mh_odm.xml'; /* Assign output XML file */
PROC CDISC MODEL = ODM
  Write = odmxml;

ODM          data = odm;
STUDY        data = study;
GLOBALVARIABLES data = globalvariables;
METADATAVERSION data = metadataversion;

CLINICALDATA data = mh
  domain = "MH"
  origin = "Test Medical History data"
  purpose = "To transfer the Medical History data"
  name = "Medical History"
  comment = "Sample MH records";

RUN;
filename odmxml; /* Deassign the external file */

```

SAS data set created from **step 1**

Medical History data set created from **step 2**.

### 3. IMPORTING AN XML FILE TO A CLINICAL STANDARD SAS DATA SET:

Sample SAS code:

```

libname out 'C:\nesug2006\output'; /* directory of the output SAS data set */
filename xmlin 'C:\nesug2006\output\mh_odm.xml'; /* XML file to be imported */

PROC CDISC          model = ODM
                   read = xmlin
                   formatActive = Yes
                   formatNoReplace = No
                   ;
ODM          ODMVersion = "1.2"
           ODMMinimumKeyset = YES /* If = YES, only _SubjectKey was kept */
           ;
CLINICALDATA out = out.mh
           SasDatasetName = "MH"
           ;

RUN;
filename xmlin; /* Deassign the external file */

```

Imported SAS data set OUT.MH

**Output** from PROC CONTENTS (using VARNUM option) and PROC PRINT of the imported SAS data set:

The CONTENTS Procedure					
Variables in Creation Order					
#	Variable	Type	Len	Informat	Label
1	__SubjectKey	Char	100		
2	STUDYID	Char	10		Study Identifier
3	DOMAIN	Char	2		Domain Abbreviation
4	USUBJID	Char	20		Unique Subject Identifier
5	MHSEQ	Num	8	12.2	Sequence Number
6	MHTERM	Char	200		Reported Term for the Medical History
7	MHDECOD	Char	20		Dictionary-Derived Term

8	MHBODSYS	Char	24		System Organ Class
9	VISIT	Char	20		Visit Name
10	VISITNUM	Num	8	12.2	Visit Number
11	MHENRF	Char	20		End Relative to Reference Period

Obs	__SubjectKey	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM
1	abc123456.000023.1	abc123456	MH	abc123456.000023	1	CHAGAS DISEASE
2	abc123456.000226.1	abc123456	MH	abc123456.000226	1	LEFT VENTRICULAR EJECTION FRACTION DECREASED
3	abc123456.000227.1	abc123456	MH	abc123456.000227	1	LEFT VENTRICULAR EJECTION FRACTION DECREASED
4	abc123456.000227.2	abc123456	MH	abc123456.000227	2	MITRAL VALVE PROLAPSE
5	abc123456.000227.3	abc123456	MH	abc123456.000227	3	MITRAL REGURGITATION

  

Obs	MHDECOD	MHBODSYS	VISIT	VISITNUM	MHENRF
1	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
2	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
3	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
4	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR
5	Dummy Preferred Term	Dummy System Organ Class	Screening	10	PRIOR

## DOWNLOADING PROC CDISC

### FROM THE SAS WEB SITE:

The current version is 2.15.52. Press CTRL + click to follow link, then follow instructions in Readme.txt:  
[http://www.sas.com/apps/demosdownloads/proccdisc\\_PROD\\_9\\_sysdep.jsp?packageID=000369](http://www.sas.com/apps/demosdownloads/proccdisc_PROD_9_sysdep.jsp?packageID=000369)

### USER'S NOTES

1. The derivation of the metadata variable `__SUBJECTKEY` is very important when exporting the Clinical SAS data set to the XML file. In the above example, each subject may have more than one Medical History event, so if the variable `__SUBJECTKEY` used only the variable `USUBJID`, we would have only 3 records in the imported SAS data set `OUT.MH`. Only the last Medical History event would be kept for each subject. With the combination of `USUBJID` and `MHSEQ`, the imported SAS data set `OUT.MH` has 5 records.
2. We can have more than one clinical data statement in PROC CDISC to export/import multiple SAS data sets.
3. An error message will be generated if the input SAS data set to PROC CDISC is missing the required metadata variables (see STEP 2).
4. To verify that your XML file is exported as expected, the generated XML file can be imported back into the SAS data set to be compared with the original SAS data set.
5. To check for the version of PROC CDISC, run the following code: PROC CDISC **version**; run;

## CONCLUSION

The beauty of PROC CDISC is to be able to validate the Clinical SAS data sets for compliance with the SDTM V3.1 standards. It also provides a flexible way to convert your Clinical SAS data sets to XML file and vice versa. We are looking forward to having PROC CDISC support additional domains in the near future.

## REFERENCES

[1] "The CDISC Procedure for SAS Software, Release 8.2 and Later", and "CDISC Procedure for the CDISC SDTM 3.1 Format", SAS Institute Inc.  
<http://support.sas.com/rnd/base/topics/sxle82/cdiscsdm.html>

[2] CDISC web site, <http://www.cdisc.org/>

**ACKNOWLEDGMENTS**

The author has special thanks to Dr. Shi-Tao Yeh for his suggestion and encouragement on the paper's topic .

**RECOMMENDED READING**

For in-depth usage and reference material, press CTRL+click to follow link:

Proc CDISC Field Response Release,

<http://support.sas.com/rnd/base/topics/sxle913/fieldresponse.htm>

<http://support.sas.com/rnd/base/index-xml-resources.html>

**CONTACT INFORMATION**

Chauthi Nguyen

GlaxoSmithKline Pharmaceutical, Inc.

2301 Renaissance Boulevard

King of Prussia, PA 19406

Work phone: (610) 787 - 3853

E-mail: [Chauthi.Nguyen@gsk.com](mailto:Chauthi.Nguyen@gsk.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.

Other brand and product names are trademarks of their respective companies.