

Macro to get data from specific cutoff period

Kiranmai Byrichetti, SCRI, Nashville, TN

Jeffrey Johnson, SCRI, Nashville, TN

ABSTRACT

During the clinical development of an investigational drug, periodic analysis of safety information is of great importance to the ongoing assessment of risk to trial subjects. It is also important to inform regulators and other interested parties (e.g., ethics committees) at regular intervals about the results of such analyses and the evolving safety profile of an investigational drug, and apprise them of actions proposed or being taken to address safety concerns. For periodic analysis, data needs to be pulled based off of specific time period and doing this on each and every output or dataset is time consuming and having a macro to do such task is definitely time saving. Also, Manual selection of data based off particular dates can be prone to errors. For this purpose, a macro has been developed and this paper discusses a macro that applies cutoff based off specific dates on all SDTM datasets available in both production and validation folders. This macro is developed in SAS 9.3 and users with basic SAS knowledge would be able to use the macro.

INTRODUCTION

As a compound moves through the various stages of clinical development, and more and more data are collected, different type of deliverables have to be generated. Sponsor more often requests data between specific dates or sometimes cumulative data but only before a particular date. This selection of data can be done in many ways and can be done at different levels (Raw data, SDTM or at ADaM). However, this type of request may be is valid only for one deliverable and data may doesn't need any manipulation for the other deliverables. Slicing and dicing of data at different levels can be a daunting task. Especially when done manually, code has to be added or removed based on the requirement and that can lead to errors. Hence, a macro has been developed to serve this purpose which can be only run, when there is a request from the sponsor and another set of data is created based off the particular dates of interest.

This paper explains a macro that actually applies data cutoff on all SDTM production and validation datasets after all the datasets are generated by the normal run. Changing at the SDTM level does have an advantage of more control on the variable date names as followed by the SDTM implementation guide. The following challenges/questions had to be addressed before initiating the macro call:

- Is the data needs to be pulled before a date or between two specific dates, for example: DSUR annual report, where reports have to be generated with annual data between two specific dates.
- Is the study following CDISC guidelines in creating SDTM datasets?
- Are there any SUPPQUALL datasets in the study?
- Will this macro changes the cumulative SDTM data with the cutoff data?
- How to call these cutoff SDTM datasets into ADaM?
- How to handle partial dates?

This paper will detail the actions taken to address the aforementioned challenges, including rules established before any SAS programing was started and steps taken to prepare for the challenges.

Our work is done using SAS 9.3 in a windows environment.

PRELIMINARY CONSIDERATIONS

Our company follows CDISC standards for all the studies and based on our company policies all deliverables follow data flow from RAW → SDTM → ADaM, we decided to apply the data cutoff at the SDTM level as SDTM follows a standard naming convention for the date variables (for example: all dates end with DTC). Whereas Raw data can be collected in many different ways and there is no way all dates ends with common name, hence this macros has not been created at the raw dataset level and same is the case with ADaM, it will have lots of different dates created for the analysis and for the usage in the listings, it will be hard to decide on which this cutoff date needs to be applied on. Hence, SDTM is used here for the filtering of data. Moreover, SDTM DTC variable may include data that is

Macro to get data from specific cutoff period, continued

represented in ISO 8601 format as a complete date/time, a partial date/time, or an incomplete date/time. This standard naming convention is a plus in developing this macro.

What all are needed before running a macro:

- A datacutoff folder needs to be created in each of the Production and validation folder at the SDTM level.
- Add a global macro in each ADaM program where actual production SDTM and validation SDTM libraries are redirected to datacutoff folder.
- Run both production and validation bat files and then batch submit cutoff macro.
- Remember to change %let study= to your current study that you are using your macro on.
- This macro can be copied over to the study specific folder for the usage.
- Change cutoff1 and cutoff2 variable to whatever the dates that you want your data cutoff on.
- Cutoff1 date should be start of cutoff and always before or equal to cutoff2.
- Example:
%let cutoff1=%str ("10APR2000"d);
%let cutoff2=%str ("10APR2015"d);
- If client wants data to be before one particular date cutoff1 and cutoff2 will be the same,
- If sponsor wants data between two cutoff dates, for example 10apr2014 and 10apr2015 then cutoff1 and cutoff2 should be updated accordingly.

```
*****creating temporary libnames to use in the macro for both production and
validation data****;

%let lib1=%str (S:\&study\Deliverable\SDTM\Production\Data);
%let lib2=%str (S:\&study\Deliverable\SDTM\Validation\Data);

*****creating libnames that saves data based off cutoff into datacutoff folder****;

%let p_sdtm_c=%str(S:\&study\Deliverable\SDTM\Production\Data\Datacutoff);
%let v_sdtm_c=%str(S:\&study\Deliverable\SDTM\Validation\Data\datacutoff);

LIBNAME source "&lib1";
LIBNAME compare "&lib2";

LIBNAME p_sdtm_c "&p_sdtm_c";
LIBNAME v_sdtm_c "&v_sdtm_c";
```

Once all the above mentioned steps have been taken care of, this macro can be run and datasets for each SDTM data for both production and validation folder will be created. Temporary libnames that were created will be used by the rest of the macro. Libnames were created in the macro based on the assumption that all studies have standard structural flow and this can be easily changed as per the company standard at the start of the macro. This macro has several autocalls and nothing needs to be changed in the code, &libname has values of both "Prod" and "Vald" which uses production and validation data respectively, &source has values of both "p_sdtm_c" and "v_sdtm_c" where cutoff data goes respectively.

The SAS System generates and maintains valuable information at run time about SAS libraries, data sets, catalogs, indexes, macros, system options, titles, and views in a collection of read-only tables called dictionary tables. Although called tables, Dictionary tables are not real tables. Information is automatically generated at runtime and the tables contents are made available once a SAS session is started. This Macro uses dictionary tables to access data based off the libnames created. When the autoexec is run for a study, All the libnames defined in the autoexec will be available in the dictionary tables and are ready for use.

```
/******Get all the datasets in the respective libraries*****/

PROC SORT DATA = sashelp.vtable OUT = data (keep = memname memlabel libname where =
(libname="&libname"));
By memname;
Run;

DATA data;
SET data;
```

Macro to get data from specific cutoff period, continued

```
%if &libname = SOURCE %then %do;

if substr(memname,1,4) = "SUPP" then sup = "&"||"dsn="||substr(memname,5) ;
%end;

%else %if &libname=COMPARE %then %do;
if substr(memname,3,4)= "SUPP" then sup1="&"||"dsn=V_"||substr(memname,7) ;

%end;
run;

/*****Get all the columns from all the datasets*****/

PROC SORT DATA = sashelp.vcolumn OUT = var(keep = memname name type LENGTH label
format libname where = (libname = "&libname"));
by memname;
run;

PROC SORT DATA = var;
by memname;
run;

DATA var_name;
SET var;

/*****using informed consent date as the date where cutoff needs to be applied in
SDTM.DM, this can be changed based on study requirements*****/
/*****Looks for variable whose variable name ends with DTC as per SDTM standards*****/

where substr(reverse(strip(name)),1,3)='CTD' and substr(name,3,3) ^='END' and
LENGTH(name) ^> 7;
if memname in ("DM" "V_DM") and name ^='RFICDTC' then delete;
run;

****STDTC dataset and use last. to get STDTC date picked if dataset has both DTC and
stdtc****;

PROC SORT;
by memname name;
run;

DATA var_name;
SET var_name;
by memname name;
if last.memname;
run;

/*****Get datasets with no date variables available to apply cutoff*****/

DATA nodate;
LENGTH col $50;
MERGE var_name(in=a keep = memname) data(in=b where =(substr (memname,1,4)^in ("SUPP"
"V_SU")));
by memname;
if b and not a;

nodatestr="&"||"dsn="||strip(memname) ;

col="% "||"cutoff("||strip(libname)||', '||strip(memname)||', , "||")";
run;
```

Macro to get data from specific cutoff period, continued

```
/******Get datasets with date variables available to apply cutoff*****/  
  
DATA var_name;  
SET var_name ;  
run;  
  
/******SET both datasets together*****/  
  
DATA var_name1;  
SET var_name nodate;  
  
var=_n_;  
%if &libname=SOURCE %then %do;  
  
dsn_=memname;  
  
%end;  
  
%else %if &libname=COMPARE %then %do;  
  
dsn_=substr(memname,3);  
  
%end;  
  
col="%||"cutoff("||strip(libname)||','||strip(memname)||","||strip(name)||","||strip(  
dsn_)||","||strip(put(var,best.))||")";"  
run;
```

In the above code, macro variables are being created in the data step to store snippets of code that contains name of another macro variable which is one of the input parameter of the macro. All the datasets and columns from the required datasets are pulled. SDTM demographics (DM) dataset stores various different dates such as RFICDTC, RFSTDTC, RFENDTC, Any of these dates can be pulled off of the SDTM based on the sponsor requirement. Here, RFICDTC is used to apply the cutoff. Hence, all other dates are deleted when creating a macro variable that is again used as one of the input parameters.

Some datasets may not contain dates to apply cutoff on, like Trial domains TA, TE. This has been taken into consideration and macro checks for the datasets with no date variables and save the data set as is into the new data cutoff folder.

```
PROC SQL noprint;  
  select col into: datastr separated by ' ' from data1 where var=1;  
  select col into: dsnstr separated by ' ' from var_name1 ;  
  
%if &libname=SOURCE %then %do;  
  
  select supp into: memn separated by ' or ' from data where substr(memname,1,4)=  
"SUPP" ;  
  
  %end;  
  
%else %if &libname=COMPARE %then %do;  
  
  select suppl into: memn separated by ' or ' from data where substr(memname,3,4)=  
"SUPP" ;  
  
  %end;  
  select nodatestr into: nodatestr separated by ' or ' from nodate ;  
  
quit;  
  
OPTION NOFMterr;  
  
%macro cutoff (base,dsn,name,dsn_,var);
```

Macro to get data from specific cutoff period, continued

```
%if &nodatestr %then %do;

/*****getting datasets with no date variables to apply cutoff into the new library
without any change*****/

DATA &source.&dsn;
SET &base.&dsn;
run;

%if &libname=SOURCE %then %do;

%if &memn %then %do;
DATA &source..supp&dsn;
SET &base..supp&dsn;
run;

%end;

%end;

%else %if &libname=COMPARE %then %do;

%if &memn %then %do;

DATA &source..supp&dsn;
SET &base..V_supp&dsn_;
run;

%end;
%end;
%end;

%else %do;

/*****applying cutoff and saving the data into the new library *****/

DATA &dsn(drop=&name._n);
SET &base.&dsn;

if LENGTH(&name)>=10 then &name._n=input(substr(&name.,1,10),yymmdd10.);
format &name._n date9.;

if . < (input(put(&cutoff1, date9.), date9.)) le &name._n le (input(put(&cutoff2,
date9.), date9.)) > . or (&name._n=.) ;

run;

/*****Sort datasets by usubjid and sequence number to have the sorting right
As all datasets with dom.seq number is assigned based off a specific sorting order
*****/

PROC SORT DATA=&dsn;

%if %upcase(&dsn.) = DM or %upcase(&dsn.) = SV or %upcase(&dsn.) = V_DM or
%upcase(&dsn.) = V_SV %then %do;
by usubjid ;
%end;

%else %do;
by usubjid &dsn_.seq;
%end;
```

Macro to get data from specific cutoff period, continued

```
run;

DATA &source..&dsn;
SET &dsn;
run;

%if &memn %then %do;
DATA s_&dsn;
  SET
  %if &libname=SOURCE %then %do;
    &base..supp&dsn.;
  %end;

%else %if &libname=COMPARE %then %do;
&base..V_supp&dsn_.;
%end;

%if %upcase(&dsn.) = DM or %upcase(&dsn.) = SV or %upcase(&dsn.) = V_DM or
%upcase(&dsn.) = V_SV %then %do;
  &dsn_.seq=.;
%end;

%else %do;
if idvarval ^= ' ' then &dsn_.seq=input(idvarval, best.);
%end;

run;

/*****Get only those idvarval records corresponding to the datacutoff parent
dataset seq records*****/

PROC SORT;
by usubjid &dsn_.seq;
run;

DATA %if &libname=SOURCE %then %do; &source..supp&dsn. %end;
%else %if &libname=COMPARE %then %do;
&source..V_supp&dsn_.
%end;
(drop=&dsn_.seq);
MERGE &dsn (in=a keep=

%if %upcase(&dsn.) = DM or %upcase(&dsn.) = SV or %upcase(&dsn.) = V_DM or
%upcase(&dsn.) = V_SV %then %do;
Usubjid

%end;

%else %do;

usubjid &dsn_.seq

%end; )
s_&dsn(in=b);

%if %upcase(&dsn.) = DM or %upcase(&dsn.) = SV or %upcase(&dsn.) = V_DM or
%upcase(&dsn.) = V_SV %then %do;
by usubjid
%end;
```

Macro to get data from specific cutoff period, continued

```
%else %do;
by usubjid &dsn_.seq
%end;;
if a and b;
run;

%end;
%end;

%mend cutoff;
OPTIONS MPRINT MLOGIC SYMBOLGEN;

&&dsnstr;

%mend datacut;

%datacut(libname=SOURCE,source=p_sdtm_c);

%datacut(libname=COMPARE,source=v_sdtm_c);
```

Here, Source and compare are already defined in the libname statements at the start of the macro and these names should not change as the macro uses these names for various conditional statements.

HOW THE MACRO WORKS?

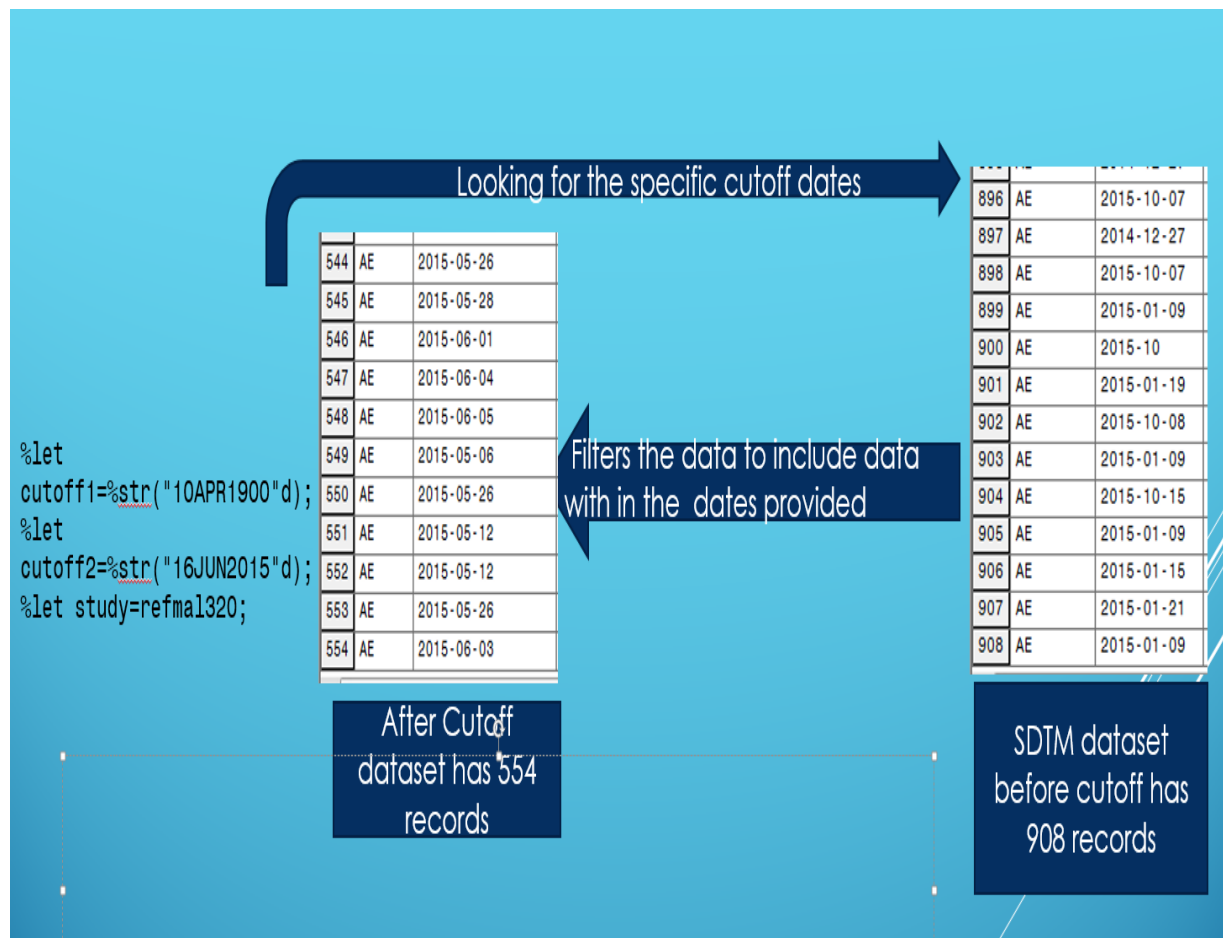


Figure 1. How the macro works?

Macro to get data from specific cutoff period, continued

- Basically, Macro gets all datasets from the library given in the autoexec and picks variables ending with DTC.
- Once all the datasets with date variables are identified, a numeric version of date is created and is used to compare with that of the cutoff1 and cutoff2 dates and filters the data to include records that falls in between input dates given into the macro.
- If a study has SUPPQUALL variables, SUPPQUALL datasets are identified and are merged back with that of the parent dataset to include only data for the patients from the parent dataset.
- Finally, Dataset that are created with the cutoff filter are sorted by USUBJID and sequence number and are saved into the new folder created.
- For the datasets which doesn't have sequence number such as Demographics, only USUBJID is used in the sorting order.

```
=====*/
%let study=refmal320;
*****creating temporary libnames to use in the macro for both production and validation data****;
%let lib1=%str(S:\&study\Deliverable\SDTM\Production\Data);
%let lib2=%str(S:\&study\Deliverable\SDTM\Validation\Data);

*****creating libnames that saves data based off cutoff into datacutoff folder****;
%let p_sdtm_c=%str(S:\&study\Deliverable\SDTM\Production\Data\datacutoff);
%let v_sdtm_c=%str(S:\&study\Deliverable\SDTM\Validation\Data\datacutoff);

libname source "&lib1";
libname compare "&lib2";

libname p_sdtm_c "&p_sdtm_c";
libname v_sdtm_c "&v_sdtm_c";

*****If client wants data to be before one particular date cutoff1 and cutoff2 will be the same,
if sponsor wants data between two cutoff dates, for example 10apr2014 and 10apr2015 then cutoff1 and cutoff2

%let cutoff1=%str("10APR1900"d);
%let cutoff2=%str("16JUN2015"d);

%inc "s:\&study\autoexec.sas";
```

Figure 2. Sample code used for one of the studies

In the above sample code from Figure1, Sponsor wanted data that falls on/before 16 June 2016 and cutoff2 has been set up with the same date.Cutoff1 can be assigned the date making sure, no data point have a record with the date equal to/before

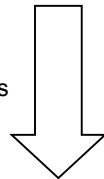
Macro to get data from specific cutoff period, continued

	USUBJID	AESTDTC
553	CX-839-001-0101014	2015-05-24
554	CX-839-001-0101007	2015-04
555	CX-839-001-0101008	2014-12-19
556	CX-839-001-0101014	2015-05-11
557	CX-839-001-0101008	2014-12-27
558	CX-839-001-0101008	2014-12-27
559	CX-839-001-0101008	2014-12-27
560	CX-839-001-0101008	2014-12-27
561	CX-839-001-0101008	2015-01-09
562	CX-839-001-0101008	2015-01-19
563	CX-839-001-0101008	2015-01-09
564	CX-839-001-0101008	2015-01-09
565	CX-839-001-0101008	2015-01-15
566	CX-839-001-0101008	2015-01-21
567	CX-839-001-0101008	2015-01-09

Figure 3. SDTM AE dataset before applying cutoff

CUTOFFMACRO

Removes records that has date falling after the cutoff date provided to the macros, in this case 13 records



	USUBJID	AESTDTC
1	CX-839-001-0102006	2015-06-17
2	CX-839-001-0102006	2015-06-17
3	CX-839-001-0102006	2015-06-17
4	CX-839-001-0103004	2015-06-18
5	CX-839-001-0103004	2015-06-18
6	CX-839-001-0104008	2015-06-30
7	CX-839-001-0104008	2015-06-30
8	CX-839-001-0108022	2015-06-24
9	CX-839-001-0108026	2015-06-17
10	CX-839-001-0108026	2015-06-17
11	CX-839-001-0112003	2015-06-18
12	CX-839-001-0112003	2015-06-18
13	CX-839-001-0112020	2015-06-24

Figure 4. SDTM AE dataset with 13 records falling after the cutoff date i.e. 16 June 2015.

	USUBJID	AESTDTC
532	CX-839-001-0118003	2015-05-21
533	CX-839-001-0118003	2015-06-04
534	CX-839-001-0118003	2015-06-04
535	CX-839-001-0119001	2015-04-22
536	CX-839-001-0119001	2015-04-22
537	CX-839-001-0119001	2015-05-06
538	CX-839-001-0119001	2015-04-22
539	CX-839-001-0119001	2015-04
540	CX-839-001-0119001	2015-04-29
541	CX-839-001-0119001	2015-04-29
542	CX-839-001-0119001	2015-05-06
543	CX-839-001-0119001	2015-05-26
544	CX-839-001-0119001	2015-05-26
545	CX-839-001-0119001	2015-05-28
546	CX-839-001-0119001	2015-06-01
547	CX-839-001-0119001	2015-06-04
548	CX-839-001-0119001	2015-06-05
549	CX-839-001-0119001	2015-05-06
550	CX-839-001-0119001	2015-05-26
551	CX-839-001-0119001	2015-05-12
552	CX-839-001-0119001	2015-05-12
553	CX-839-001-0119001	2015-05-26
554	CX-839-001-0119001	2015-06-03

Figure 5. SDTM AE data after cutoff macro has been applied

Figure 2 shows the dataset with 567 total records and Figure 3 shows 13 records does have the AESTDTC that falls after the Cutoff date passed through the macro which is 16 June 2016. Figure 4 is the dataset snippet after the macro has been applied and 13 records that fall after the cutoff date have been removed and resulting dataset with 554 records (567-13=554) is generated. Likewise, all other datasets are checked by the macro at one instance and new datasets will be saved. Running of macros doesn't take much time and .log will be created after the run to check for errors, warnings and notes.

CONCLUSION

This paper is particularly useful for Statistical programmers, Statisticians and anyone who comes up with similar scenario. This macro comes handy especially when a programmer have lots of deliverables at a time or there are multiple studies that needs a data cut at the same time. The best way to use this macro is creating a copy in the study level macro folder and make appropriate changes as explained in the paper for the first time and use it whenever needed without any change except for the date. Also, when programming ADaM, Libname for SDTMs must be redirected to get filtered SDTM datasets. As this macro is developed based on the standards followed at SCRI, it may need some tweaks here and there to adapt to various studies. Users needs knowledge of SAS macros and base SAS. There are some limitations to the macro including but not limited to: It assumes that SDTM datasets has at least one SUPPQUALL dataset in the folder. This can be modified easily to not to check for SUPPQUALL but that code has not been updated/included.

Macro to get data from specific cutoff period, continued

REFERENCES

SDTM implementation guide 3.1.2

RECOMMENDED READING

- Base SAS® Procedures Guide
- SAS® For Dummies®
- SAS Macros

CONTACT INFORMATION

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

Name: Kiranmai Byrichetti
Enterprise: SCRI, Nashville, TN
Address: 3322 West End Ave Suite 900
City, State ZIP: Nashville, TN 37203
E-mail: Kiranmai.byrichetti@scri-innovations.com

Name: Jeffrey Johnson
Enterprise: SCRI, Nashville, TN
Address: 3322 West End Ave Suite 900
City, State ZIP: Nashville, TN 37203
E-mail: Jeffrey.johnson@scri-innovations.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.