

## Selecting Analysis Dates - A Macro Approach Using Raw Data for Expediting Results

Abdul Ghouse, Seattle Genetics, Inc., Bothell, WA

### ABSTRACT:

Clinical trials data contain a variety of values including dates which support several key analyses such as Time to Event (TTE). It's often seen that the algorithms defined to conduct TTE analyses require the selection of a particular date that can be found across large number of datasets. Using a macro to create a single dataset with all the CRF collected dates allows selecting the event or censoring per patient basis.

This approach simplifies the process improving efficiency and quality. In addition, using data from the raw SAS library allows for faster output creation and review of interim efficacy endpoints for decision making in a phase 1 study.

### INTRODUCTION:

Dates of events in clinical trials are used for various analyses both for safety and efficacy. For safety we can use the date events start and stop to assess duration of an event. This is fairly straightforward as all the dates needed are in a single AE dataset. When assessing time to event for death or progression of disease, the dates needed can come from any number of datasets. It is critically important to select the latest date to ensure the analysis is correct.

This paper discusses a macro approach for collating all the CRF collected dates in to a single dataset called "alldates" dataset that can enable us to perform a number of efficacy analyses based on timing of events including overall survival, progression free survival or event free survival analysis (relapse free survival),. This dataset can also be used for safety analyses such as duration of serious adverse events report with the AE coded terms, and time to first onset of an event of interest.

SAS code:

```
%macro all;
/*Gathers all the raw dataset and variable names*/
proc sql;
create table columns as
select strip ('SAS raw library reference. ' || strip (memname) as dsname, name as
varnames,
label, format from dictionary.columns
where libname='SAS raw librarary reference' and format in ('DATE9.', 'DATETIME22.3')

/*creating macro parameters for the dataset and variable names*/
select distinct dsname, varnames into :dsr, :vrs separated by ' '
from columns;
quit;

/* SYMBOLGEN: Macro variables resolve to string of dataset and variable names */
%put &dsr;
%put &vrs;

/*reference section 1*/
data final (keep= subject &vrs response lbtest lborres lborresu aeterm aeser);
set &dsr;
run;
```

A single dataset with all relevant dates to support analysis requiring start and/or stop dates of events is created as below. In addition to the date fields, the critical variables required to do analysis like response can be kept in the dataset.

Subject name or identifier	eCRF page name	Investigator assessment of OVERALL RESPONSE at this timepoint	BMDT	EXSTDT	LBDT	VSDT
899997	Response Assessment	Progressive Disease	23APR2015			
899997	Response Assessment	Not Evaluable	23APR2015			
899997	Response Assessment	Not Evaluable	23APR2015			
899997	Response Assessment					
9999991	Response Assessment	Complete Remission (CR)				
9999991	Response Assessment	Not Evaluable				
9999991	Response Assessment	Not Evaluable	23APR2015			
9999992	Response Assessment	Progressive Disease	23APR2015			
9999992	Response Assessment	Partial remission (PR)	23APR2015			
999993	Response Assessment	Not Evaluable				
999993	Response Assessment	Progressive Disease	23APR2015			
999993	Response Assessment	Progressive Disease	23APR2015			
899999	Study Drug Administration			23MAR2015		
899997	Study Drug Administration			23MAR2015		
899997	Study Drug Administration			23MAR2015		
9999991	Study Drug Administration			23MAR2015		
9999991	Study Drug Administration			23MAR2015		
9999992	Study Drug Administration			23MAR2015		
999993	Study Drug Administration			23MAR2015		

```

/*create variable lcondt with all the numeric dates*/
data Alldates;
set final;
array nums (*) _numeric_;
call symput('x',hbound(nums));
do i=1 to dim (nums);
if nums(i) ne . then lcondt=datepart (nums (i));
end;
format dts date9.;
run;

/*censoring patients dataset with minimum and maximum dates of all the date
variables*/
%macro t;
proc sql;
create table allcnsr as
select subject, %do i=1 %to &x; min(%scan(&a,&x) )as mn%scan(&a,&x) format=date9.,
max(%scan(&a,&x)) as mx%scan(&a,&x) format=date9., %end; &i
from alldates
group by subject;
quit;
%mend ;

/*macro call*/
%t

/*creating events dataset for progressive disease and death*/

data evntpd ;
set alldates;
if rsall='Progressive Disease' then do;
pdflg='Y';
output;
end;

```

```

if deathdt ne . then do;
dthflg='Y';
output ;
run;

proc sort data=evntpd out=evntpd2(keep=subject,pdflg dthflg) nodupkey;
by subject;
run;

/*join the unique event subjects and indicator variable with the censored dataset*/

Proc sql;
Create table anlysrdy as
Select *
From allcnsr left join evntpd2
On allcnsr.subject=evntpd2.subject
Order by subject;
Quit;

%mend all;

%all;

```

It is also important to have an indicator variable for whether or not the event occurred for some TTE analyses. When the event does not occur the patient is considered “censored” at the last follow-up time. The censoring of patients could be the last response assessment date or the assessment done during long term follow-up or considering the response dates before a subsequent treatment date. Details of how to perform the censoring will be found in the statistical analysis plan.

The use of the “alldates” dataset is extensible in an efficient manner. .For example, it can be used to provide listings of dose modifications or dose reduction due to an adverse event related to the study drug, and identify missing lab id’s resulting in missing reference ranges and toxicity grades.

```

/*creating safety listings*/

/*reference section 1*/
data final (keep= subject &vrs lb: ae:);
set &dsr;
run;

```

## CONCLUSION

Creation of a single dataset with all relevant dates can increase efficiency of SAS programming for a trial. For example, we can expedite the creation of output and review from the time of data extract for safety end points and efficacy endpoints in a Phase 1 trial where review of individual patient safety is critical for trial progression. We can also improve efficiency for later phase trials by creating macros to read from the “alldates” data set and create a variety of outputs.

## REFERENCES

(<http://support.sas.com/documentation/cdl/en/sqlproc/63043/PDF/default/sqlproc.pdf>)

(<http://support.sas.com/documentation/cdl/en/sqlproc/62086/HTML/default/viewer.htm#a002595788.htm>)

## ACKNOWLEDGMENTS

I would like to thank Aileen Murphy, Rajeev Karanam and Jay Gadhiya for reviewing this paper.

## CONTACT INFORMATION

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

Name: Abdul Ghouse  
Enterprise: Seattle Genetics, Inc.  
Address: 21823 30<sup>th</sup> Drive Southeast  
City, State ZIP: Bothell, WA 98021  
E-mail: [aghouse@seagen.com](mailto:aghouse@seagen.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.