

## **A Solution to Look-Ahead Observations**

Yongjiang (Jerry) Xu, CSL Behring, King of Prussia, PA  
Yanhua (Katie) Yu, PRA Health Sciences, Blue Bell, PA

### **ABSTRACT**

In clinical trials, it is common to compare certain laboratory test results across a period or various time points. For example, the confirmation of drug responses in multiple myeloma studies and some hematology indicators emerging back to safe levels from nadir (the lowest point). These data are usually mapped into finding domains in SDTM and programmed into BDS structure type ADaM datasets. However, it is difficult to look ahead at the values of future observations or look back to previous observations in SAS® because SAS reads one observation at a time into the Program Data Vector (PDV). It is a known issue in SAS data management that we cannot do comparisons across observations. This paper will present a practical solution to transpose and do data comparison and calculations across observations using DO loops and arrays.

### **KEYWORDS**

Look-ahead observations, Array, DO loop, Transpose

### **INTRODUCTION**

In order to make logical decisions about a set of variables, programmers need to look ahead (or look back) to results in subsequent (previous) visits/measurements. Some examples include the derivation of response status according to IMWG response criteria or RECIST 1.1, the time emerging from nadir of absolute neutrophil count (ANC) to its ‘safe’ level in chemotherapy treatment (a ‘safe’ ANC level means that the patient’s activities do not need to be restricted on the basis of the ANC).

By default, SAS initializes all the variables in a data set to missing each time it processes a new observation and only the current observation is available for use. Several methods of look-ahead/look-back have been published, for example, the LAG function to look back at previous observation. Arrays, on the other hand, provide a more powerful solution. An array in SAS provides a means for repetitively processing variables using a DO-loop. Arrays are a convenient way of grouping variables and do not persist beyond the data step in which they are used. Arrays allow iterative processing of alike variables saving machine and programmer time. This paper illustrates one solution using an array and a DO loop on transposed data to look at multiple observations to identify key events for a patient.

### **THE SCENARIO**

Following chemotherapy and other medical treatments, neutrophils (part of white blood cells) and platelets are often depleted leaving the person more susceptible to infection and bleeding. When a lab value gets to its lowest value, this is consider nadir for that lab test. It is very

important to know when the person’s lab counts have returned to a safe level, often allowing them to be discharged from the hospital and return to more normal activities. Clinically, having three daily consecutive neutrophil counts above 500 following nadir is considered the date of recovery.

## THE PROGRAMMING APPROACH

Assuming the Absolute Neutrophil Count (ANC) values are stored in dataset XX (below) and a new sequence variable SQNM has been created for each observation remaining after removal of observations with missing values and selecting the lowest ANC value per calendar day).

SUBJECT	SQNM	XXTESTCD	XXDTC	XXSTRESN	XXSTRESU
001	1	ANC	2019-04-01T08:30	620	/uL
001	2	ANC	2019-04-02T08:30	600	/uL
001	3	ANC	2019-04-03T08:30	610	/uL
001	4	ANC	2019-04-04T08:30	300	/uL
001	5	ANC	2019-04-05T18:30	350	/uL
001	6	ANC	2019-04-07T08:30	450	/uL
001	7	ANC	2019-04-08T08:30	510	/uL
001	8	ANC	2019-04-09T08:30	490	/uL
001	9	ANC	2019-04-10T08:30	550	/uL
001	10	ANC	2019-04-12T08:30	700	/uL
001	11	ANC	2019-04-13T08:30	1000	/uL
001	12	ANC	2019-04-15T08:30	1100	/uL
001	13	ANC	2019-04-19T08:30	1150	/uL

To determine the maximum number of rows (SQNM) for any given subject, a macro variable, &maxseq, is created using a PROC SQL statement as follows:

```
proc sql noprint;
  select max(sqnm) into: maxseq trimmed from xx;
quit;
```

### Step 1. Transposing Data

The dataset XX, which doesn’t have multiple results on a given calendar day, is transposed from vertical data structure to horizontal creating variables (\_1, \_2, etc.) from observations that can be used in an array:

```
proc transpose data=xx out=xx1;
  by subject;
  id sqnm;
  var xxstresn;
run;
```

In addition, the transpose step could also be performed using an array. It is required that the dataset XX be sorted previously by subject (to allow the use of the BY statement).

```
data xx1 (drop=lbstresn);
  set xx;
  by subject;
  array nvar{&maxseq} _1 - _&maxseq;
  retain _1 - _&maxseq;

  if first.subject then do;
    do i=1 to &maxseq;
      nvar{i}=.;
    end;
  end;
  nvar{sqnm}=lbstresn;
  if last.subject;
run;
```

## Step 2. DO Loop and Array.

Then find the lowest ANC value (nadir), determining which position in the array this value falls. In addition, search for three consecutive values in the array that are after the nadir position and above 500.

```
data xx2 ;
  set xx1;

  ****Find the nadir value****;
  nadir = min(of _1 - _&maxseq.);

  ****Set up the Array****;
  array nad{&maxseq.} _1 - _&maxseq. ;

  ****Find the first lowest point (nadir)***;
  do i = &maxseq to 1 by -1;
    if nad{i}=nadir then nd_i= i;
  end;

  **The 3 sequential results >500 after nadir**;
  do i = &maxseq-2 to 1 by -1 ;
    if nd_i< i and nad{i}>500 and nad{i+1}>500 and
      nad{i+2}>500 then sqnm=i;
  end;
  keep subject nadir nd_i sqnm ;
run;
```

### Step 3. Find the Date.

Once the desired index variable (sqnm) is identified, a merge of datasets XX2 and XX using this sqnm value will get the date the ANC values begin return to a safe level.

```
data xx3;
  length safdtc $19;
  merge xx (in=a) xx2 (in=b where =(sqnm
  ne .));
  by subject sqnm;

**** Assign ANC recovery date based on SQNM ****;
  if b then ancrdtc=lbdtc;
run;
```

### CONCLUSION

There are many needs to look ahead at observations within datasets in clinical trials. A solution to look ahead to multiple observations in SAS datasets using DO loop and array is illustrated by one clinical example, which is to find the date the ANC values begin to return to a safe level from nadir.

### REFERENCE

1. Kumar, S., Paiva, B., Anderson, K. C., etc. (2016). International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *The Lancet Oncology*, 17(8), e328-e346. doi:10.1016/S1470-2045(16)30206-6
2. E.A. Eisenhauer, P. Therasse, J. Bogaerts, etc. (2009). New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1), 45 (2), *European Journal of Cancer*, pp. 228–247.
3. Seymour L., Bogaerts J., Perrone A., etc. (2017). “iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics”. *The Lancet Oncology*, 18 (3) e143-e152. doi: 10.1016/S1470-2045(17)30074-8.
4. Dunn, Toby and Chang Y. Chung, (2005), “Retaining, Lagging, Leading, and Interleaving Data,” *Proceedings of the PharmaSUG Conference 2005*.

### ACKNOWLEDGEMENTS.

I would like to acknowledge Daphne Ewing for encouragement and review of this paper.

## **AUTHOR CONTACT**

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

Yongjiang (Jerry) Xu, Ph.D.  
CSL Behring LLC.  
1020 First Ave,  
King of Prussia, PA 19406  
Phone: 6102907549  
Email: jerry.xu@cslbehring.com

Yanhua (Katie) Yu  
PRA Health Sciences  
721 Arbor Way,  
Blue Bell, PA 19422  
Phone: 215-591-6918  
Email: yukatie@prahs.com

## **TRADEMARK INFORMATION**

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.