

An Automated Approach to Patient Exclusion Accountability Summaries Using SAS® Macro Language

Cam Huffman, Burroughs Wellcome Co.
Randall Austin, Burroughs Wellcome Co.
Beth Sherrill, Burroughs Wellcome Co.
Greg Silva, Trilogy Consulting Corporation
Carl Arneson, Burroughs Wellcome Co.

ABSTRACT

Often in clinical trials, some or all of a patient's data may be unreliable due to various circumstances. It may be desirable to exclude these data from certain analyses or summaries. However, it is also important to keep detailed accounts of when and why such exclusions are made. This can be a difficult task, especially when such exclusions are made in many different programs, and the information about the exclusions is continually being updated. We have implemented a system for simplifying this process using the SAS Macro Language. All exclusions are listed in a central set of files, and the information in these files is read into each program upon execution. Whenever a program is run, all relevant exclusions are incorporated into the analyses, and a detailed report of these exclusions is created. This system makes it easier for the user to keep track of and report exclusions, resulting in more reliable and efficient analyses.

INTRODUCTION

Often in clinical trials, some or all of a patient's data may be unreliable due to various circumstances. For example, the patient could have been misdiagnosed, the data collection instrument could have malfunctioned, or there could have been a concomitant medication violation. It is often of interest to exclude such data from some of the analyses or summaries. However, it is also necessary to keep track of such exclusions, since the FDA requires strict accountability for all patient data. The practice of excluding invalid patient information from some analyses can make it difficult to account for the total number of patients represented in a given analysis. This is especially true when only a portion of a patient's information is excluded.

SAS special missing values may be used to track patient exclusions in a clinical trials preferred analysis. The special missing values in SAS may be used to exclude selected numeric values from analysis. These special missing values can be formatted to correspond to the reason for the exclusion (e.g., .C for Concomitant Medications). These are then used to create an accountability listing of patients included or excluded from each variable in an analysis. Hence, the SAS special missing values act as a code for tracking patient exclusions.

In this system, exclusion information for all patients in a study is kept in a central location rather than being inserted piecemeal into individual analysis programs. If exclusions are added or changed, analysis programs only need to be rerun, not modified, to incorporate the new exclusion information. This can save time and ensure accuracy during the hectic days prior to a submission of an NDA.

The main advantage for our clinical colleagues who do most of the writing of study reports is that this system saves them time. The clinical authors at our company tell us it has literally saved them days in the report preparation cycle.

Clinical authors and reviewers do not have to spend time simultaneously referring to patient listings and accountability tables

to determine who is in or out of an analysis and for what reason. The accountability summary makes it easy to see, at a glance, how many patients are excluded and for what reasons. This is very convenient for analyses where the number of patients is large. When there are many analysis summaries in a report, the time saved reviewing accountability issues can be substantial.

OVERVIEW

The process of accounting for patient exclusions has three basic steps:

1. Use of the %XUPDATE¹ macro to create exclusion data sets
2. Use of exclusion data sets in analysis programs and the %XCALC² macro to retain special missing values through mathematical operations
3. Use of the %ACCOUNT¹ and %ACC_LIST¹ macros to count the number of patients included/excluded from each analysis variable and list them by reason for exclusion

THE %XUPDATE MACRO

The %XUPDATE macro takes raw data sets and replaces selected variable values with SAS special missing values. The exclusions to be made are entered into the system via one of two flat files, the XDATA file or the XSAS file. The XDATA file is a flat file with columns defined for the key variables:

```
DSNAME PATIENT VARIABLE_NAME EXCLUSION_CODE
```

If there are multiple observations per patient, an optional TIME variable can be included to uniquely identify each record.

The XSAS file allows for more specialized cases as it actually consists of lines of data step code which are pulled into the %XUPDATE macro. This file may be used to handle cases where a record-by-record list of exclusions would be unwieldy. An example of an XSAS file might be as follows:

```
if DSNAME="HRDATA" then do;
  if PT in(101,104,105,201)
    and GROUP <> 10 and TIME > 100 then do;
    VAR1 = .E;
    VAR2 = .E;
  end;
  if PT=285 and TIME<50 and GROUP=80
    then VAR1=.Z;
end;
```

The %XUPDATE macro cycles through all the raw data sets for a particular study (identified by common file type). It uses a data step to make the changes coded into the XSAS file (creating a work data set). The entries into the XDATA file are pulled into a SAS data set and an UPDATE statement is then used to make these changes to the work data set. PROC COMPARE checks to see if this new data set is any different from the exclusion data set already available. If they are identical (i.e. no new exclusions have been added), nothing is done. If there are differences, the

exclusion data set is replaced. In this way, the date of the exclusion data set always corresponds to the date of the last change to any exclusion file. We have the program that drives this macro scheduled to run automatically every night, so our exclusion data sets stay current.

These exclusion data sets are then used in place of raw data sets in the programs which run analyses. When summary procedures such as UNIVARIATE or MEANS are run on these data sets, the special missing values are counted as missing and do not contribute to the overall sample size.

THE %XCALC MACRO

An additional issue is how to keep the SAS special missing values in place if mathematical operations are to occur on the exclusion data sets prior to analysis. Usually, mathematical operations on missing values of any kind, even special missing values, result in missing numeric value. In other words, special missing values are lost in the process. For example, a patient's systolic and diastolic blood pressure may have been excluded due to concomitant medications, so we have these values updated as .C (a SAS special missing value). If we then calculate:

$$\text{Mean Arterial Pressure} = 1/3 \text{ systolic} + 2/3 \text{ diastolic}$$

in the analysis program, the result will be missing (not .C) and we will not be able to account for this patient's lack of information.

The %XCALC macro was designed to retain the special missing values in such situations. It takes common mathematical functions, compares the values of variables containing any kind of missing value and assigns a special missing value to the result based on a pre-defined hierarchy of exclusion codes. For example, for the function:

$$X = Y + Z \text{ where } Y = .C \text{ and } Z = .M$$

the %XCALC macro assigns each special missing value a numeric value corresponding to the user's hierarchy of exclusion categories (defined in a format library). This bypasses the usual SAS hierarchy of special missing values from .A to .Z, giving the user more flexibility. If .C is assigned a higher value than .M, then the result will be the value .C. Without the %XCALC macro, the result would have been only a missing numeric value with no explanation as to why this value is missing. With the %XCALC macro, all operations involving special missing values return a special missing value, thus preserving the exclusion code from one operation to the next.

THE %ACCOUNT AND %ACC_LIST MACROS

After an analysis, such as a summary table has been produced, the %ACCOUNT macro is called using the same data set that was used to produce the analysis. For each analysis variable, the %ACCOUNT macro counts the number of values that were non-missing, missing numeric value or missing due to special missing values and keeps track of which patients are in each category. The add-on macro, %ACC_LIST, runs a PROC REPORT on this accountability information. It lists the total sample size used to summarize a variable and the patients who contributed to that total by treatment group. Then it lists the patients who were not in the analysis due to missing numeric values or due to one of the special missing values formatted to identify exclusion codes.

For example, data set DATA1 contains the variables VAR1 and VAR2, which are the percent change in some measure of physical status such as blood pressure. Some of the values for VAR1 and VAR2 may not be valid due to equipment malfunction, improper procedure, assessments made at the wrong time, or for some other reason, and have been replaced with special missing values by the %XUPDATE macro. The following macro call will produce a

summary of the DATA1 data set that lists the patients who have missing values by reason for exclusion (see Appendix A). Note that the NUMMISS parameter is a list of values to be summarized.

```
%account(data=DATA1,varlist=VAR1 VAR2,
nummiss=DOTS .C .E .M .Z ALLIN,
trt=TRTMENT,subtrt=GROUP);
%acc_list(data=_OUT,trt=TRTMENT,subtrt=GROUP);
```

The %ACCOUNT macro can also summarize data based on criteria other than exclusions. In the following example, the NUMMISS parameter is defined so that the data is summarized based on the percent-change category (i.e., the number and list of patients with 20%, 30% or 40% change; see Appendix B):

```
%account(data=DATA1,varlist=VAR1 VAR2,
nummiss=-40 -30 -20 20 30 40 999 ALLIN,
trt=TRTMENT,subtrt=GROUP);
%acc_list(data=_OUT,trt=TRTMENT,subtrt=GROUP);
```

GENERALIZABILITY

The system we have described was developed for use on the CMS and VMS operating systems. It has been used on the UNIX operating system and should be easily portable to other operating systems.

In this application, raw data sets and data sets with excluded values are permanently stored even though many of the observations are identical. With large datasets or small storage space, it may be preferable to apply the exclusions as datasets are used. Faster processing for this step would allow the user to apply exclusions as needed without storing additional datasets.

CONCLUSION

This system for tracking patient exclusions from selected variables using SAS special missing values provides an easy way for users to keep up with which patients are represented in each analysis. Changes or additions to the exclusion list can be quickly incorporated by rerunning the XUPDATE driver and the analysis program.

ACKNOWLEDGMENTS

- ¹Original code by Randall Austin
- ²Original code by Greg Silva

SAS is a registered trademark of SAS Institute Inc. in the USA. ® indicates USA registration.

Appendix A:

Table A1
 Percent Changes in Variable of Interest
 Treatment = Treatment 1
 Group = Group 1
 ACCOUNTABILITY LISTING

Variable	Value	n	Patient List
VAR1	Missing Num Value	2	102 103
	Equipment Malfunction	1	109
	Non-Missing	42	101 104 105 106 107 108 110 111 112 113 114 115 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315
VAR2	Missing Num Value	2	102 103
	Equipment Malfunction	1	109
	Non-Missing	42	101 104 105 106 107 108 110 111 112 113 114 115 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315

Appendix B:

Table A2
 Percent Changes in Variable of Interest
 Treatment = Treatment 1
 Group = Group 1
 ACCOUNTABILITY LISTING

Variable	Value	n	Patient List
VAR1	30% Increase	1	305
	40% Increase	1	204
	Within Range	27	201 202 203 205 206 207 208 209 210 212 213 214 215 301 302 303 304 306 307 308 309 310 311 312 313 314 315
	20% Decrease	1	211
VAR2	40% Increase	1	211
	Within Range	26	201 202 203 205 206 208 210 212 213 214 215 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315
	20% Decrease	3	204 207 209

