

A System of SAS[®] Macros for Producing Statistical Reports

Greg Grandits, M.S.

Ken Svendsen, M.S.

Division of Biostatistics, University of Minnesota

ABSTRACT

Monitoring clinical trials requires periodic generation of statistical reports for Data and Safety Monitoring Board (DSMB) reviews and other purposes. These reports display comparisons between treatment groups and consist of summary statistics and significance levels comparing the groups. The summary statistics can be simple descriptive summaries (counts, means, SD's, etc.) or summaries from more complicated analyses (e.g. hazard ratios and confidence intervals from proportional hazards regression) and often include a combination of both. Existing SAS[®] reporting procedures do not provide a straightforward method of generating such reports, in part because these reports require the collation of information from several procedures. However, with procedure output datasets, the output delivery system (ODS), and the data-step, SAS has the tools necessary to produce such reports. This paper describes a set of macros that use these tools to create reports in which statistical information from several procedures is combined and displayed on a single report page, in either simple text or html format. These macros have been used extensively by the Division of Biostatistics at the University of Minnesota to produce reports for many clinical trials and observational studies.

INTRODUCTION

In clinical trials periodic reports are generated to monitor study progress and to compare treatments for relevant variables of interest. Often these reports are presented to Data and Safety Monitoring Board (DSMB) committees. These reports usually contain summary statistics of variables (counts, means, SD's, etc.) for each group intertwined with other statistical information such as p-values, Z-statistics, hazard ratios, and confidence intervals. A method to automate the generation of these reports is important, to help ensure accuracy and reduce programming time: one that does not require transcription of numbers, typing, or editing of computer output. SAS[®] reporting procedures, in general, are inadequate to generate these types of reports, in part because these reports include information from several procedures. However, with procedure output datasets, the output delivery system (ODS), and the data-step, SAS has the necessary tools to produce such a report. This paper describes a system of macros that produce customized statistical reports that are easy to program and modify, and give complete flexibility to placement of text and summary statistics onto the report page.

The user first defines columns across the report page. Text or data values (summary statistics) are then moved to these columns and specified lines using macros MOVE and NMOVE. Summary statistics are available from calling a macro which runs a procedure (GLM, PHREG, etc.), outputs statistics to a SAS dataset, and then compresses the statistics into a one observation dataset. Statistics are placed into array type names which can be moved to the report page after a SET statement. Example: %nmove(p1-p8, col=7, line=12L8) moves 8 p-values to the 7th defined column starting on line 12.

The description and use of the macros for moving text and data values to the report page have been given in an earlier SUGI paper (1). These macros are briefly outlined here and are followed by a description of the statistical macros that make available information from the SAS procedures typically used for monitoring clinical trials. The report macros have also been enhanced to allow for generation of html tables which can be displayed by a web browser or imported into a word processing or spreadsheet document. This system of macros provide a comprehensive package for generating statistical reports for a variety of research applications.

STEPS IN WRITING REPORT

Report programs are made up of the following statements:

1. %REPORT statement that indicates a new report is starting
2. %COLSET statement that defines columns and column widths across the report page.
3. %MOVE statements which move text to the report. Features include centering, underlining, and repeating text.
4. SET statement(s) that read in statistical information from a SAS dataset. These one observation datasets are generated from one of the statistical generating macros described below
5. %NMOVE statements which place the statistical information to the report page.

SYNTAX FOR REPORT MACROS

1. %REPORT is used simply as %REPORT which indicates the start of a new report.
2. %COLSET is used as follows:
 %COLSET (column1 size column2 size ...)
 Example: %COLSET (25 2x 10 10 10)
 This statement sets up 4 columns. The first column is 25 positions wide and the last 3 columns are 10 positions wide. Two spaces are placed between the first and second columns. This is used to offset text from other text. For html output these column widths are converted to appropriate pixel widths.
3. %MOVE is used as follows:
 %MOVE ('string 1': 'string 2': ..., line=, col=, center=, under=)
 This is best illustrated by examples.
 'Men': 'Women': 'Total' text strings to be placed on report
 line = 12 21 33 moves strings to lines 12, 21, and 33
 line = 12L3 moves strings to lines 12, 13, and 14

 col = 3 4 8 moves strings to defined columns 3, 4, and 8
 col = 2-3 4-5 6-7 moves strings to columns formed by combining columns 2-3, 4-5, and 6-7
 col = 2.8 moves strings to columns 2 *through* 8

 under and center set to "y" to underline and center text

 Example: %MOVE ('Men': 'Women': 'Total', col=1, line=10L3)
4. %NMOVE is used as follows:
 %NMOVE (var1-var(n), line=, col=, fmt=);
 The line and col parameters are identical to those in %MOVE. The fmt parameter formats the values.
 Example: %NMOVE (m1-m20, col = 2 3, line=12L10, fmt=6.2)
 This statement would move the values of m1 through m20 to columns 2 and 3, and to lines 12 through 21.

STATISTIC GENERATING MACROS

Below is a listing of several of the statistical generating macros, the SAS procedure that is called by the macro, the statistics that are available, and a brief description of the macro.

MACRO	PROCEDURE	STATISTICS	DESCRIPTION
BREAKDN	SUMMARY	N, mean, SD, etc.	Summary statistics by level of class variables
FREQDIS	SUMMARY	Counts, percents, cumulative percents	Distribution of variable by level of another variable
GLMP	GLM	ANOVA p-values	Statistics from analysis of variance
REGP	REG	p-values, betas, t-stat, etc.	Statistics from linear regression
PHREGP	PHREG	Betas, SEs, HRs, CIs, p-values, etc.	Statistics from Cox regression
LOGISTP	LOGIST	Betas, SE, ORs, CIs, p-values, etc.	Statistics from logistic regression
CHISQP	FREQ	CMH p-values	Stratified contingency table analyses
Several others are also available, and others can be added as needed.			

For illustration, two of the macros, %BREAKDN and %PHREGP, are described in more detail. Other macros have similar syntax.

%BREAKDN (DATA=, CLASS=, VAR=, OUT=, SFIRST=);

This macro reads the SAS dataset specified in DATA using PROC SUMMARY and computes summary statistics for each variable specified in VAR by each level of the variable(s) specified in CLASS. A one observation dataset containing these statistics is written to the dataset specified in OUT.

The statistics calculated are N, MEAN, MEDIAN, SDEV, SE, SUM, MIN, and MAX. They are contained in the variables N1-N?, M1-M?, MED1-MED?, S1-S?, SE1-SE?, SUM1-SUM?, MIN1-MIN?, and MAX1-MAX?, where ? depends on the number of variables in VAR and the number of levels in the variables in CLASS.

The parameters specified are illustrated by examples.

Parameter/Value	Description
class = sex 2T	statistics are stored for both levels of the variable SEX and the total
class = sex 2 group 6	statistics are stored for each level of SEX crossed with GROUP.
var = age dbp sbp chol	This is the list of variables for which to compute statistics .

DATA is the SAS dataset to be read; OUT is the SAS dataset statistics are written to and contains one observation, and SFIRST indicates the order in which the statistics are stored.

EXAMPLE:

```
%BREAKDN (class = sex 2T, var = age dbp sbp chol, out = table1, sfirst=VAR)
```

The n's for the 4 variables where sex = 1 are stored in n1-n4.

The n's for the 4 variables where sex = 2 are stored in n5-n8.

The n's for the 4 variables for men and women combined are stored in n9-n12.

The variables are stored similarly for the other statistics.

%PHREGP (PARAMETERS)

PHREGP runs PROC PHREG to perform proportional hazards regression and saves results for factors of interest into SAS datasets.

Parameter	Description
data =	SAS dataset to be read
dlist =	Dependent variable list. An analysis is done for each variable listed. The variables are event indicators coded as 1 if event, 0 if censored.
ilist =	Independent variable list. Used for each dependent variable given in dlist.
tlist =	Failure or censoring time list corresponding to events in dlist.
factor=	Independent variable (s) for which statistics are output to a SAS dataset.
units =	Value regression coefficients are multiplied by before relative risks are computed.
strata =	Optional list of stratifying variables.
out =	SAS dataset(s) to which statistics are written.

The statistics and the names of the variables that contain them are as follows:

e1-e?	regression coefficients for factor
se1-se?	standard errors of coefficients
z1-z?	z-statistics for factor
p1-p?	p-values for factor
rr1-rr?	relative risks (hazard ratios) for factor
u1-u?	upper 95% CI for RR
l1-l?	lower 95% CI for RR

? is the number of variables in dlist, which is the number of analyses run.

WRITING HTML AND WORD TABLES

The above macros have been recently modified to allow the user to output the report as an HTML table. These files can then be viewed using a web browser or imported into Word. Most of the formatting options that are available in html such as font type, font size, font weight, and indenting and underlining text have been incorporated as well as more general html features such as background and text color. In addition an option has been added to combine data and place it in a single column with certain types of formatting, such as mean $\pm$ SD, N (%), or HR (95% CI), giving the report table a more journal finished look. Default options have been set up so that when the html table is inserted into Word the columns spread out proportionately to fill the page margins, making it unnecessary to deal with margin issues.

An example program and html output (inserted into word) is given in the appendix.

DISCUSSION

Much effort has been made by SAS and SAS users to make reporting easier. Although no individual SAS feature or procedure is sufficient to provide the ease or flexibility in producing statistical reports, with use of the system of macros described, which takes advantage of the data step, output from procedures, and the ODS, a very useful reporting system can be developed. The statistical report macros described here are simple to use and have tremendous flexibility. Programs are easy to write, understand, and modify. Typical report programs are less than one page (see example program). These macros can also be expanded to include statistics from other procedures through use of ODS.

The key to making the numeric moves in the report section is getting the statistics into one observation datasets. Then a single SET statement is all that is needed to make available the statistics, without worry of the implied loop in the data step. Information from several different sources (SAS datasets) can easily be included on the report by multiple SET/NMOVE statements. This gives the flexibility to the system. These macros have proved invaluable to the clinical trials and other projects monitored by the Division of Biostatistics at the University of Minnesota, and could be useful for any research organization or pharmaceutical company producing statistical reports for clinical trials.

REFERENCE:

1. A set of SAS macros for producing customized reports. Presented at the 19th Annual SAS Users Group International Conference 1994, Dallas, Texas.

CONTACT INFORMATION

Greg Grandits
Division of Biostatistics
2221 University Ave. SE, Suite 200
Minneapolis, MN 55414
Email: grand001@umn.edu
Phone: 612-626-9033

Macro code and documentation are available at: www.biostat.umn.edu/~greg-g

Appendix

EXAMPLE PROGRAM (GENERATES TEXT REPORT)

```
* Assume dataset temp contains all needed variables ;
* Macro variable evlist contains indicators for events;
* Macro variable tmlist contains the days until the event or censoring
* Breakdn call obtains the descriptive statistics for each group;
* Phregp call obtains statistics from Cox-Regression for group;

%let evlist = primary mi str cvddth allcvd corevas tia
              angina chf acchyp renalf ;
%let tmlist = tprimary tmi tstr tcvddth tallcvd tcorevas ttia
              tangina tchf tacchyp trenalf ;

%breakdn(data=temp, class= group 2, var=&evlist, out=out1);
%phregp(data=temp, dlist = &evlist, strata=clinic,
        tlist= &tmlist, ilist=group, factor=group, out = out2);

%report(repfile='example.txt');
%colset(25 7 7 2x 7 7 4x 7 7 7 2x 7);

%move('Number and Percent of Selected Cardiovascular Events by Treatment Group':
      'And Hazard Ratio (Experimental/Standard)from Cox Regression Analyses', col=1-9,
      line=3L2);
%move('Experimental': 'Standard', col=2-3 4-5, line=9, under=y);
%move('Cox Regression Summary', col=6-9, line=9, under=y);
%move('Endpoint', col=1, center=2, line=11, under=y);
%move('N': '%', col=2.5, line=11, u=y);
%move('HR': 'L95%': 'U95%': 'P-Val', col=6.9, line=11, u=y);
%move('Primary CVD ':
      ' MI (Fatal/NF)':
      ' Stroke (Fatal/NF)':
      ' CVD Death':
      'Any CVD Hospialization':
      ' Revascularization':
      ' TIA':
      ' Angina':
      ' Heart Failure':
      ' Hypertension':
      ' Renal Failure',
      col=1, center=n, line=14L4 19L7);

set out1;
%nmvmove(sum1-sum22 , col=2 4, line=14L4 19L7, fmt=5.0 );
%nmvmove(m1-m22 , col=3 5, fmt=5.1, scaler=100 );

set out2;
%nmvmove(r1-r11 L1-L11 u1-u11, col=6.8, fmt=5.2 );
%nmvmove(p1-p11 , col=9, fmt=6.3);
```

TEXT OUTPUT

Number and Percent of Selected Cardiovascular Events by Treatment Group
And Hazard Ratio (Experimental/Standard) from Cox Regression Analyses

Endpoint	Experimental		Standard		Cox Regression Summary			
	N	%	N	%	HR	L95%	U95%	P-Val
Primary CVD	364	4.5	365	4.4	1.02	0.88	1.18	0.771
MI (Fatal/NF)	133	1.6	166	2.0	0.82	0.65	1.03	0.089
Stroke (Fatal/NF)	133	1.6	118	1.4	1.15	0.90	1.48	0.265
CVD Death	152	1.9	143	1.7	1.09	0.87	1.37	0.471
Any CVD Hospitalization	793	9.7	775	9.3	1.05	0.95	1.16	0.307
Revascularization	163	2.0	166	2.0	1.01	0.82	1.26	0.913
TIA	89	1.1	105	1.3	0.87	0.66	1.15	0.330
Angina	202	2.5	190	2.3	1.09	0.89	1.33	0.389
Heart Failure	126	1.5	100	1.2	1.30	1.00	1.69	0.051
Hypertension	22	0.3	18	0.2	1.26	0.67	2.34	0.474
Renal Failure	27	0.3	34	0.4	0.81	0.49	1.35	0.426

EXAMPLE PROGRAM (GENERATES HTML REPORT)

```
%wreport(htmlfile='example.html');
%colset(25 14 14 21 7);

%move('Number and Percent of Selected Cardiovascular Events by Treatment Group<br>'
      'And Hazard Ratio (Experimental/Standard) from Cox Regression Analyses',
      col=1-5, line=3, fontweight=bold, fontsize=12pt);
%move('Number of Patients<br>With Event', col=2-3, line=9, fontweight=bold);
%move('Cox Regression Analyses<br>Experimental/Standard', col=4-5, line=9,
      fontweight=bold);
%move('Experimental': 'Standard': 'HR (95% CI)': 'P-value', col=2.5, line=11,
      fontweight=bold);
%move('Endpoint', col=1, center=n, line=11, fontstyle=italic, fontweight=bold)
%move('Primary CVD ': 'Any CVD Hospitalization', col=1, center=n, line=14 19,
      fontstyle=italic, fontweight=bold);
%move('MI (Fatal/NF)':
      'Stroke (Fatal/NF)':
      'CVD Death':
      'Revascularization':
      'TIA':
      'Angina':
      'Heart Failure':
      'Hypertension':
      'Renal Failure',
      col=1, center=n, line=15L3 20L6, indent=1);

set out1;
%nmmove(sum1-sum22 m1-m22, col=2 3, scaler=1 100, combine=y, fmt=5.0 5.1, fchar=2P);

set out2;
%nmmove(r1-r11 L1-L11 u1-u11, col=4, line=14L4 19L7, combine=y, fmt=5.2 5.2 5.2,
      fchar=3);
%nmmove(p1-p11 , col=5, fmt=6.3);

stop;
run;
%makehtml;
```

HTML OUTPUT (INSERTED IN WORD)

Number and Percent of Selected Cardiovascular Events by Treatment Group And Hazard Ratio (Experimental/Standard) from Cox Regression Analyses				
	Number of Patients With Event		Cox Regression Analyses Experimental/Standard	
<i>Endpoint</i>	Experimental	Standard	HR (95% CI)	P-value
<i>Primary CVD</i>	364 (4.5%)	365 (4.4%)	1.02 (0.88 - 1.18)	0.771
MI (Fatal/NF)	133 (1.6%)	166 (2.0%)	0.82 (0.65 - 1.03)	0.089
Stroke (Fatal/NF)	133 (1.6%)	118 (1.4%)	1.15 (0.90 - 1.48)	0.265
CVD Death	152 (1.9%)	143 (1.7%)	1.09 (0.87 - 1.37)	0.471
<i>Any CVD Hospitalization</i>	793 (9.7%)	775 (9.3%)	1.05 (0.95 - 1.16)	0.307
Revascularization	163 (2.0%)	166 (2.0%)	1.01 (0.82 - 1.26)	0.913
TIA	89 (1.1%)	105 (1.3%)	0.87 (0.66 - 1.15)	0.330
Angina	202 (2.5%)	190 (2.3%)	1.09 (0.89 - 1.33)	0.389
Heart Failure	126 (1.5%)	100 (1.2%)	1.30 (1.00 - 1.69)	0.051
Hypertension	22 (0.3%)	18 (0.2%)	1.26 (0.67 - 2.34)	0.474
Renal Failure	27 (0.3%)	34 (0.4%)	0.81 (0.49 - 1.35)	0.426