Paper 177 – 2010

Tips for Automating Univariate Outcomes Analysis in Hematopoietic Stem Cell Transplantation

Peigang Li, MS; CIBMTR, Medical College of Wisconsin, Milwaukee, Wisconsin
Xiaochun Zhu, MS; CIBMTR, Medical College of Wisconsin, Milwaukee, Wisconsin
Min Chen, MS; CIBMTR, Medical College of Wisconsin, Milwaukee, Wisconsin

ABSTRACT

Hematopoietic stem cell transplantation has been used to treat patients diagnosed with a variety of diseases including leukemia, severe aplastic anemia, Hodgkin's disease, non-Hodgkin's lymphoma, multiple myeloma, a number of blood disorders, and some solid tumor cancers. Typical outcomes include treatment-related mortality (TRM), relapse (REL), progression free survival (PFS), and overall survival (OS). *Univariate* outcomes analysis is characterized as either survival or cumulative incidence rate without adjusting for other covariates, i.e., *unadjusted analysis* (Klein & Moeschberger 2003; Klein et al. 2001, Part I). It is desirable to automatically generate summary tables containing probability, confidence intervals and p-values. Log-rank or pointwise p-values are used to look at equality over the strata or at a given fixed point in time such as three-year survival. The process was facilitated with a set of SAS macros.

KEYWORDS

SAS, MACRO, ODS, SQL, IML, LIFETEST, TIMELIST; Kaplan-Meier Estimator; Competing Risk; HSCT

INTRODUCTION

Univariate analysis is a basic analysis of outcomes of interest in hematopoietic stem cell transplant (HSCT) studies, including survival and competing risk analysis (Klein et al. 2001, Part I; Kim 2007; Scrucca 2007). Currently all outcomes for survival and competing risk analysis can be done using macros that export the result into the SAS output or Excel file format and cannot handle more than two groups. However, extra steps must be taken to make a desirable summary table. In addition, p-values must be calculated separately if outcomes comparison is needed and the evaluable number of observations is usually retrieved by reading the SAS output. The whole process is time-consuming and error-prone, especially when the univariate analysis has to be redone if the requirements have changed. This paper will describe SAS macros that retrieve SAS ODS® outputs, calculate p-values and create a summary table automatically.

METHOD

Standard SAS source code was converted to SAS macros to facilitate repeated generation of summary tables for outcomes analysis. The critical requirements for the macros are:

(a) able to automatically identify group levels and call the corresponding utility macros
(b) able to take a list of time points as input
(c) able to calculate survival and mortality (=1-survival) in the same macro
(d) able to generate p-values if group levels are greater than one

The core algorithm uses PROC SQL to extract the probability (Prob) and standard error (S.E.) or the confidence interval (CI) directly. It works fine under the following assumptions:

- The survival probability and the lower or upper bound of the confidence interval are in decreasing trend
- The cumulative incidence with competing risk is in increasing trend

The Kaplan-Meier estimator used in the survival function or cumulative incidence justifies these assumptions. The arcsine confidence intervals were used. The SQL statements will need to be revised to handle situations that patients obtained a therapy after relapse (e.g., DLI) where a multi-state model is used for survival analysis. Interested readers might check other references for multivariate outcomes analysis (Klein et al. 2001, Part 2; Klein 2006; Logan et al. 2006; Scrucca 2010).

The newly created macros export a series of numbered SAS data sets that can be combined and formatted to fit the summary table requirements. The user can then generate the summary table in either Excel or RTF file.

The macros were cross-validated in two outcomes studies:

- Study 1: only Prob and CI were required
- Study 2: p-values were required in addition to Prob and CI

USEFUL TIPS

(a) Writing SAS macros
Primary macros: passing parameters

Survival analysis:

```
%macro survstatgrp(
   data,             /* Input SAS data set */
   group,            /* could be one or more levels */
   timept,           /* list of fixed time points */
   event,            /* indicator variable */
   censor,           /* censoring status (0 or 1) */
   intv,             /* study time or interval in months */
   life              /* flag, 1=survival, 0=mortality */
   pval              /* P-value requested: 1=Yes, 0=No */
   i                 /* Counter  */
   )
```

Cumulative incidence:

```
%macro cifstatsgrp(
   data,             /* Input SAS data set */
   group,            /* could be one or more levels */
   timept,           /* list of fixed time points */
   event,            /* indicator variable */
   competerisk,      /* competing risk for the event */
   censor,           /* censoring status (0 or 1) */
   intv,             /* study time or interval in months */
   pval              /* P-value requested: 1=Yes, 0=No */
   i                 /* Counter  */
   )
```

Utility macros:

```
%incid():cumulative incidence rate
%arcsinci(): arsine confidence interval
%pointwise2gp(): pointwise p-value for 2 groups; simplified quadratic form
                 method
%pointwise3gp(): pointwise p-value for 3 groups; simplified quadratic form
                 method
%pointwise4gpandmore(): pointwise p-value for 4 or more groups; general
                 quadratic form method (Klein et al. 2001, Part I)
```

(b) ODS Tables: getting the levels in a group

```
*RETRIEVE N groups;
  ods trace on;
  ods output NLevels=nlev;
proc freq data=temp NLEVELS; tables &group; run;

DATA _NULL_; set nlev;
  if _n_ =1 then
  call symput("_NGroups", NLevels);
run;
```

(c) SAS arrays: retrieving the time list

```
%let npt = %eval(1 + %numparams(&timept));
DATA out1;
     ARRAY sarr(&npt) sarr1-sarr&npt;
     sarr {1} =0;

   timelim =99;
   do i=1 to &npt-1;
```

```
            if sarr {i} < time <= sarr {i+1} then timelim =i ;

         end;
      run;
```

(d) SQL Query Window: extracting Prob and SE or CI
Step 1:

```
      *DIRECT OUTOUT FROM SAS SQL QUERY WINDOW WITH MINOR MODIFICATIONS;
      PROC SQL;
      create table irtci as
      Select
      IRT.timelim,
      IRT.group,
      MAX(IRT.CI1) as CI11 label="MAX(CI1)"
      from IRT
      group by IRT.timelim,
      IRT.group
      having
      IRT.timelim
      LT &npt AND (Max(IRT.CI1) GT 0)
      order by
      IRT.timelim ASC,
      IRT.group ASC
      ;
       quit;
      run;
```

Step 2: omitted.

(e) PROC IML: pointwise p-values
Convert existing PROC IML code into a module (equivalent of R function) and write a wrapper to invoke the module.

```
      * CALCULATE POINTWISE P-VALUES FOR FOUR GROUPS;
      START pointwise3(dat);
          ngp=NROW(dat);
          (...omitted)
          return (pval);
      FINISH;
```

EXAMPLES and RESULTS

All macros must be included in the path and a time list format library is output.

```
      %include 'cif2.sas';
      %include 'arcsinci.sas';
      %include 'pointwisemcr.sas';
      %include 'cifstatgrpmcr.sas';
      %include 'survstatgrpmcr.sas';

      proc format;
      value $timeptf (DEFAULT=100)
        '3.29' = '  @ 100 days'
          '12' = '  @ 1 year'
          '36' = '  @ 3 years'
          '60' = '  @ 5 years'
         '120' = '  @ 10 years'
         '144' = '  @ 12 years';
      run;
```

The primary macros can be invoked with parameters by value directly or by reference with global variables.

By value:

```
      %survstatgrp(data=survout, group=yeargp, timept=%str(3.29 12 36 60 120 144),
      event=dead, censor=0, intv=intxsurv, life=1, pval=1, i=8);
```

By reference:

```
%let preq=1; %let group=%str(yeargp); %let timept=%str(12 36 60);

%survstatgrp(data= survout, group=&group, timept=&timept, event=dead, censor=0,
intv=intxsurv, life=1, pval=&preq, i=8);
```

The final outcomes results can be exported into .XLS or .RTF file (Haworth 2001).

```
DATA final; %output(1,8); format variable timeptf. ; run;
filename myfile &myfile;
ods html body=myfile;
PROC PRINT data=final noobs; run;
quit;
ods html close;

ods escapechar='^';
ods rtf file='Table2.rtf' style=custom bodytitle;
title 'Table 2: Univariate analysis of GVHD,Mortality,TRM,REL,PFS,OS';
PROC REPORT data= final nowd split='*';
column variable _1 _2 _3 cProb;
define variable /"Outcomes" order=data style(header)=[just=left];
define _1  /"1989-1994 ^n Probability ^n (95% CI)";


define _2  /"1995-2000 ^n Probability ^n (95% CI)";
define _3  /"2001-2005 ^n Probability ^n (95% CI)";
define cProb  /"P-values (Log-Rank and Pointwise)";
run;
ods rtf close;
```

Note that the group variable must start at a number ≥ 1. The macro *%incid()* threw an error at PROC IML line "ngrp=MAX(x[,4])" if starting at 0.

The result for Study 1 is shown in Table 1 and that for Study 2 is shown in Table 2 in the **Appendix**.

CONCLUSION

It's been quite a convenience using the new macros. As more groups at different time points are compared, the p-values must be adjusted for multiple comparisons (Klein et al. 2001, Part I). Future improvements might include a list of numeric covariates (prognostic variables) to be tested for association with the failure time.

REFERENCES

1. Klein JP, Moeschberger ML. *Survival Analysis*. 2nd edn. Springer: New York, 2003.
2. Klein JP, Rizzo JD, Zhang MJ, Keiding N. *Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part I: unadjusted analysis*. Bone Marrow Transplant 2001; 28: 909–915.
3. Klein JP, Rizzo JD, Zhang MJ, Keiding N. *Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part 2: Regression modeling*. Bone Marrow Transplant 2001; 28: 909–915.
4. Klein, JP. *Modeling Competing Risks in Cancer Studies*. Statistics in Medicine 25(6), 1015-1034, 2006
5. Logan BR, Zhang MJ, Klein JP. *Regression models for hazard rates versus cumulative incidence probabilities in hematopoietic cell transplantation data*. Biol Blood Marrow Transplant 2006; 12: 107–112.
6. Kim HT. *Cumulative incidence in competing risks data and competing risks regression analysis*. Clin Cancer Res 2007; 13: 559–565.
7. Scrucca L, Santucci A, Aversa F. *Competing risk analysis using R: an easy guide for clinicians*. Bone Marrow Transplant 2007; 40: 381–387.
8. Scrucca L, Santucci A, Aversa F. *Regression modeling of competing risk using R: an in depth guide for clinicians*.. Bone Marrow Transplant. 2010 Jan 11.
9. Lauren Haworth 2001. *Output Delivery System: The Basics*. Cary, NC: SAS Institute Inc.
10. SAS Institute Inc. *Base SAS® 9.2 Procedures Guide*. Cary, NC: SAS Institute Inc.
11. SAS Institute Inc. *SAS® 9.2 Macro Language: Reference*. Cary, NC: SAS Institute Inc.
12. SAS Institute Inc. *SAS® 9.2 Output Delivery System: User's Guide*. Cary, NC: SAS Institute Inc.
13. SAS Institute Inc. *SAS® 9.2 SQL Query Window User's Guide*. Cary, NC: SAS Institute Inc.
14. SAS Institute Inc. *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.
15. SAS Institute Inc. *SAS/IML® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

ACKNOWLEDGEMENTS

We thank the Center for International Blood and Marrow Transplant Research (CIBMTR) for providing us with the test data sets and SAS sample code.

CONTACT INFORMATION

Peigang Li, Biostatistician
Center for International Blood and Marrow Transplant Research (CIBMTR)
Froedtert and the Medical College of Wisconsin Clinical Cancer Center
9200 W. Wisconsin Avenue, Suite C5500
Milwaukee, WI 53226 USA
Telephone: 414-805-0700
Fax: 414-805-0714
E-mail: **peigang@mcw.edu**
Web: **http://www.cibmtr.org**

APPENDIX

Table 1: Univariate analysis of GVHD, Mortality, TRM, REL, PFS, OS

Outcomes	1989-1994 Probability (95% CI)	1995-2000 Probability (95% CI)	2001-2005 Probability (95% CI)
Acute GVHD (II-IV)			
N Eval	345	285	579
@ 100 days	29 (25-34)	33 (28-39)	35 (31-40)
Chronic GVHD			
N Eval	342	284	565
@ 1 year	31 (27-36)	28 (23-34)	43 (38-47)
@ 3 years	32 (28-37)	30 (25-36)	46 (41-50)
@ 5 years	32 (28-37)	30 (25-36)	47 (42-51)
Mortality			
N Eval	346	285	574
@ 100 days	33 (28-38)	29 (24-35)	19 (16-22)
Treatmented related mortality			
N Eval	346	285	580
@ 1 year	36 (31-41)	43 (37-48)	24 (21-28)
@ 3 years	40 (35-45)	46 (40-52)	29 (25-33)
@ 5 years	40 (35-45)	48 (42-54)	29 (26-34)
Relapse/Progression			
N Eval	346	285	580
@ 1 year	27 (22-32)	16 (12-20)	33 (29-36)
@ 3 years	36 (31-41)	23 (19-29)	47 (43-52)
@ 5 years	39 (34-44)	28 (22-33)	55 (50-61)
Progressio free survival			
N Eval	346	285	580
@ 100 days	56 (51-61)	68 (62-73)	68 (64-71)
@ 1 year	37 (32-42)	42 (36-47)	43 (39-47)
@ 3 years	24 (20-29)	31 (25-36)	24 (20-28)
@ 5 years	21 (17-26)	25 (20-30)	15 (11-20)
@ 10 years	18 (13-22)	17 (11-22)	**9 (5-15)**
Overall survival			
N Eval	346	285	580
@ 100 days	67 (62-72)	71 (65-76)	81 (78-84)
@ 1 year	50 (45-55)	50 (44-56)	60 (56-64)
@ 3 years	35 (30-41)	39 (33-45)	39 (35-44)
@ 5 years	30 (25-35)	32 (27-38)	29 (24-34)
@ 10 years	23 (19-28)	23 (17-30)	**23 (17-30)**
@ 12 years	21 (16-26)	**9 (1-27)**	

Table 2: Univariate analysis of TRM, REL, PFS, OS

Outcomes	PPCL	MM	P-value
Treatment related mortality			
N Eval	107	2132	
@ 1 year	2 (0-5)	5 (4-6)	0.035*
@ 3 years	5 (1-10)	7 (6-8)	0.317*
@ 5 years	5 (1-10)	8 (7-10)	0.128*
Relapse/Progression			
N Eval	107	2132	
@ 1 year	30 (21-40)	28 (26-29)	0.575*
@ 3 years	61 (49-72)	59 (57-62)	0.782*
@ 5 years	73 (62-84)	71 (69-74)	0.701*
Progression free survival			
N Eval	107	2132	0.703**
@ 1 year	68 (58-77)	68 (66-70)	0.971*
@ 3 years	34 (24-46)	34 (32-36)	0.898*
@ 5 years	22 (13-33)	20 (18-23)	0.788*
Overall survival			
N Eval	107	2196	0.769**
@ 1 year	86 (79-92)	87 (86-89)	0.739*
@ 3 years	62 (51-73)	65 (63-67)	0.607*
@ 5 years	48 (35-60)	47 (45-50)	0.954*

*Pointwise **Log-Rank