A SAS MACRO FOR PERFORMING REPEATED SIGNIFICANCE TESTING IN CENSORED SURVIVAL ANALYSES

Gordon W. Davis, Glaxo Inc.
James F. Ward, Glaxo Inc.

Abstract

A common practice in lengthy clinical trials with serial patient entry is the performance of repeated significance testing to determine if significant results exist so the trial can be terminated early. A repeated significance testing framework can be implemented using the asymptotic joint distribution of a class of sequentially computed logrank statistics derived by Tsiatis (1982). A documented SAS MACRO for performing these calculations is provided with an example analysis.

I. Introduction

Clinical trials have become a tool of very widespread use in testing the efficacy and safety of proposed new drugs for treatment of human diseases. Lengthy trials are often used to compare the time to therapeutic failure among groups of patients being treated for chronic diseases. Typically, new patients enter the trial over many months or years. They may be discontinued for such reasons as study protocol violation or death unrelated to disease, which prevents observation of therapeutic failure (censoring). The results of these trials are usually closely followed in order to detect the development of any significant differences in the failure time distributions among the treatment groups. Early detection of differences can not only save time and money, but it can minimize patient exposure to inferior therapy. Assessment may occur on a regular basis, such as annually.

In order to perform repeated significance testing of this type of clinical trial data, Tsiatis (1981 and 1982) derived the asymptotic joint distribution of a class of sequentially computed statistics. A member of this class is of particular interest because it corresponds to a sequential version of the logrank test in common use for survival data analysis. Additionally, it takes a particularly simple form within the class aiding computation.

A SAS MACRO is provided that calculates, at each assessment time, the Tsiatis statistic and variance corresponding to the logrank test. It permits repeated significance testing at planned assessment times after separate computation of test boundaries. An example illustrating the use of the macro and the power of the methodology is presented.

II. Repeated Significance Testing Using Tsiatis Class of Statistics

In his papers, Tsiatis provides a derivation of the asymptotic joint distribution of a class of statistics that are special cases of a general class of nonparametric statistics characterized by Tarone and Ware (1977). These sequentially computed statistics include the logrank, Gehan's modified Wilcoxon, Prentices's generalization of the Wilcoxon, and the tests based on score statistics proposed by Harrington and Fleming. This class of statistics is used to test the null hypothesis that the failure time distribution is not related to a general covariate Z.

The Tsiatis papers should be consulted for a detailed study of the derivation of the joint distributions of the general statistic $S_n(t)$. Briefly, Tsiatis approximates $S_n(t)$ over time by a sum of identically and independently distributed random vectors and then applies the Central Limit Theorem to obtain the asymptotic distribution. We will present the formulas he derives following a description of his notation. For the data observed at time t, Tsiatis defines the following variables: Z is the covariate or treatment indicator, X(t) is the study time to failure or censoring observed by real time t, and $\Delta(t)$ is an indicator variable for failure ($\Delta(t) = 1$ if the subject has failed by real time t, 0 otherwise). At any time t, the data of the n patients can be represented as the i = 1, ..., n independent and identically distributed random vectors $\{X_i(t), \Delta_i(t), Z_i\}$. The Tsiatis class of statistics have the form:

$$S_n(t) = \sum_{i=1}^{n} \hat{Q}(t, X_i(t))\Delta_i(t)$$

$$\left\{ Z_i - \sum_{j \in R(t, X_i(t))} Z_j / n(t, X_i(t)) \right\}$$

$R(t, X_i(t))$ denotes the "risk set" which consists of those patients with study times to failure or censoring which are greater than or equal to that of the ith patient; $n(t, X_i(t))$ denotes the number of patients in the "risk set." The random function $\hat{Q}(t,x)$ corresponds to the weighting functions used to produce the nonparametric statistics described by Tarone and Ware.

The statistic $S_n(t)$ converges in probability to a bivariate normal distribution with mean 0 and covariance matrix of the form

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

The covariance σ_{12} can be estimated by

$$n^{-1}\sum_{i=1}^{n}\Delta_i(t_1)\hat{Q}_1(t_1,X_i(t_1))\hat{Q}_2(t_2,X_i(t_1))$$

$$\left\{\sum_{j\in R(t_1,X_i(t_2))}\left\{Z_j - \hat{\mu}(t_1,X_i(t_1))\right\}^2/n(t_1,X_i(t_1))\right\}$$

where $\hat{\mu}(t,X_i(t)) = \sum_{j\in R(t,X_i(t))} Z_j/n(t,X_i(t))$

For the logrank statistic, $\hat{Q}(t,x)$ equals 1, which simplifies the calculations. In this case, $S_n(t)$ has asymptotically independent increments since $\sigma_{12} = \sigma_{11}$. As a result, the normalized test statistics T_i of the form

$$T_i = n^{-\frac{1}{2}} S_n(t_i)/(\hat{\sigma}_{ii})^{\frac{1}{2}};\ i=1,\ldots,\ K$$

are jointly distributed as a multivariate normal with mean 0 and correlations estimated by

$$\hat{\sigma}_{ij}/(\hat{\sigma}_{ii}\hat{\sigma}_{ij});\ i=1,\ \ldots,\ K.$$

Sequential testing requires computing boundaries $D_1, \ldots D_k$ at each of the k assessment times. Partial significance levels $\alpha_1, \ldots, \alpha_k$ must be specified in advance such that they sum to an overall level of α. The choice is arbitrary but usually the α_i increase overtime. See Sluc and Wei (1982) for a discussion of this issue and for details on computing the boundaries.

III. SAS MACRO Description

A MACRO for calculating the Tsiatis logrank statistic is documented in Appendix A. When the macro is called, it must be sent parameters containing the name of a SAS file, the name of the variable containing the time point data, the desired level of the time point variable, the names of the variables containing the failure event indicator, and time to event values, as well as a one word label describing the failure event. The event indicator equals one if the event has occurred, and zero otherwise.

Briefly, the algorithm of the MACRO first involves selecting the data pertaining to the specified time point of interest, and then sorting the data in order of decreasing time to failure event. This facilitates using nested loops to compute several required summation terms which involve all of the subjects who have greater than, or equal to, failure event time compared to the subject being processed. Such a group of subjects is termed the risk set. The sorted data is accessed by PROC MATRIX code in which an outer loop processes each subject's observation and computes overall summation terms based on interim summations computed in the nested loops.

Use of the MACRO is illustrated with the following example.

IV. Example

This example analyzes modified data generated by a real ten-year clinical trial of a new hormonal treatment for breast cancer. The response is time to death (failure) as measured from time of entry into the trial, with death time censored if it was greater than five years. Patients entered and were withdrawn for various reasons over the ten-year period. Appendix B summarizes events within treatment and control groups with all deaths displayed, even if after five years. A listing of the data collected for one of the patients is presented in Appendix C, along with a description of the variables.

The output generated from the MACRO is shown in Appendix D. The results (with the normalized statistic squared to compare with a one-degree-of-freedom chi square distribution) are shown in Table 1. This table also shows the usual logrank statistic proposed by Mantel (1966) and computed using the supplemental SAS SURVTEST procedure. From Table 1, it may be seen that the Tsiatis logrank statistic is very similar to that obtained using the Mantel logrank statistic. The Table 1 p-values do not control for repeated testing.

As discussed in Section II, a bivariate normal distribution can be used to compute the repeated test boundaries which do control for repeated testing. To perform these calculations, a BASIC program was written which accessed the IMSL subroutine MDBNOR. (MDBNOR computes the cumulative bivariate normal function.) The method of bisection was employed to find the test boundaries corresponding to the specified partial significance levels and calculated correlations. The repeated significance testing was performed for the data of the years 1978

through 1982. The correlations derived from the MACRO output and used in the analysis are presented in Table 2. The resulting test boundaries for two sets of α_i are shown in Table 3.

Observe that in each case the test statistic exceeds the calculated boundary in 1981, permitting an early decision to be made. The tests, which did not control for repeated "looks" (Table 1), might have erroneously suggested a decision as early as 1979.

V. Conclusions

Repeated significance testing methodology will continue to develop because of its importance to appropriate early assessment of clinical trials. The SAS MACRO discussed provides a tool for investigation of the Tsiatis logrank statistic applied in repeated testing. The code could be modified to calculate other members of the Tsiatis class.

VI. Acknowledgements

The authors would like to thank Dr. Imogene McCanless and Mr. Ted Pratt for their assistance in the development of this presentation.

Contact Author:

Gordon W. Davis
Glaxo Inc.
P.O. Box 13960
Research Triangle Park, NC 27709

REFERENCES

Mantel, N. (1966), "Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration," Cancer Chemotherapy Reports, 50, 163-170.

Slud, E.V., and Wei, L.J. (1982), "Two-Sample Repeated Significance Tests Based on the Modified Wilcoxon Statistic," Journal of the American Statistical Association, 77, 862-868.

Tarone, R.E., and Ware, J. (1977), "On Distribution Free Tests for Equality of Survival Distributions," Biometrika, 64, 156-160.

Tsiatis, A.A. (1981), "The Asymptotic Joint Distribution of the Efficient Scores Test for the Proportional Hazards Model Calculated Over Time," Biometrika, 68, 311-315.

Tsiatis, A.A. (1982), "Repeated Significance Testing for a General Class of Statistics Used in Censored Survival Analysis," Journal of the American Statistical Association, 77, 634-637.

TABLE 1

COMPARISON OF TSIATIS AND PROC SURVTEST
LOGRANK STATISTICS FOR FIVE-YEAR SURVIVAL
(P-VALUES UNADJUSTED FOR REPEATED TESTING)

YEAR OF ANALYSIS	TSIATIS LOGRANK* STATISTIC	P-VALUE	PROC SURVTEST LOGRANK STATISTIC	P-VALUE
1974	1.421	0.2332	1.42	0.2341
1975	4.647	0.0311	4.64	0.0313
1976	0.202	0.6535	0.20	0.6542
1977	1.338	0.2474	1.33	0.2485
1978	3.630	0.0567	3.57	0.0588
1979	4.850	0.0276	4.78	0.0288
1980	4.655	0.0309	4.56	0.0328
1981	5.528	0.0187	5.37	0.0205
1982	4.355	0.0369	4.22	0.0399
1983	3.782	0.0518	3.65	0.0559

*Square of statistic shown to compare with 1 df chi square.

TABLE 2

CORRELATIONS USED IN BOUNDARY CALCULATIONS
FOR REPEATED TESTING OF FIVE-YEAR SURVIVAL

YEAR 1, YEAR 2	CORRELATION (YEAR 1, YEAR 2)
1978, 1979	0.8822
1979, 1980	0.9075
1980, 1981	0.9169
1981, 1982	0.9199
1982, 1983	0.9102

TABLE 3

REPEATED SIGNIFICANCE TESTS FOR
FIVE-YEAR SURVIVAL

	YEAR OF ANALYSIS				
	1979	1980	1981	1982	1983
α_i	0.01	0.01	0.01	0.01	0.01
Boundary	2.576	2.393	2.314	---	---
Statistic	2.202	2.158	2.351	---	---
α_i	0.002	0.006	0.013	0.014	0.015
Boundary	3.090	2.670	2.345	---	---
Statistic	2.202	2.158	2.351	---	---

NOTE: Sum of α_i is 0.05 (overall significance level).

APPENDIX A

DOCUMENTED LISTING OF MACRO CODE

```
%MACRO SURVYR(SASDS,TIME,T,Z,FSTIND,FXT,FLABEL);
  DATA INDAT;
    SET &SASDS;
    IF &Z=. OR &FSTIND=. OR &FXT=. THEN DELETE;
    IF &TIME NE &T THEN DELETE;
    TITLE SURVIVAL ANALYSIS WITH &FLABEL AS FAILURE EVENT;
    TITLE2 AND &Z AS THE COVARIATE;

  PROC SORT DATA=INDAT;
    BY DESCENDING &FXT;

  PROC RANK DATA=INDAT OUT=RNKDAT TIES=LOW DESCENDING;
    VAR &FXT;
    RANKS TCNT;

  PROC MATRIX ;
    FETCH ZV DATA=RNKDAT(KEEP=&Z);
    FETCH STINDV DATA=RNKDAT(KEEP=&FSTIND);
    FETCH TCNTV DATA=RNKDAT(KEEP=TCNT);
    FETCH TXT DATA=RNKDAT(KEEP=&FXT);

    COLN='TIME':|:'STAT':|:'COV':|:'CHISQ':|:'PVAL';
    NOBS=NROW(STINDV);
    COV=0;
    ST=0;
    Q=1;
    DO I=1 TO NOBS BY 1;
      IF STINDV(I,) NE 0 THEN DO;
        ZSUM=0;
        J=1;
        DO WHILE(TXT(J,) GE TXT(I,));
          ZSUM=ZSUM+ZV(J,);
          IF J EQ NOBS THEN GO TO OUTL1;
          J=J+1;
          END;
        OUTL1:
        SUMTRM=ZSUM#/TCNTV(I,);
        IST=Q#STINDV(I,)#(ZV(I,)-SUMTRM);
        COVTRM=0;
        J=1;
        DO WHILE(TXT(J,) GE TXT(I,));
          COVTRM=COVTRM+(((ZV(J,)-SUMTRM)##2)#/TCNTV(I,));
          IF J EQ NOBS THEN GO TO OUTL2;
          J=J+1;
          END;
        OUTL2:
        ICOV=STINDV(I,)#Q#COVTRM;
        ST=ST+IST;
        COV=COV+ICOV;
      END;
    END;
    CHISQ=(ST##2)#/COV;
    PVAL=1-PROBCHI(CHISQ,1);
    OUTVEC=&T:|:ST:|:COV:|:CHISQ:|:PVAL;
    OUTPUT OUTVEC OUT=MTRX&T COLNAME=COLN;
  DATA ACCUM;
    SET ACCUM MTRX&T;
%MEND SURVYR;

%MACRO SURVRUN;
  DATA ACCUM; DELETE;
  %DO I=2 %TO 11 %BY 1;
    %SURVYR(BYYEAR,YR,&I,TREAT,DSTIND,DXT,DEATH);
    %END;
  PROC PRINT DATA=ACCUM LABEL SPLIT='#';
    ID TIME; VAR STAT--PVAL;
    LABEL STAT='LOGRANK#STATISTIC
    COV='COVARIANCE#TERM CHISQ='CHI SQUARE#STATISTIC
    PVAL='P-VALUE TIME='TIME#POINT';
%MEND SURVRUN;

%SURVRUN;
```

```
/*  Macro Variables:                                          */
/*  SASDS - name of SAS file containing data                  */
/*  TIME - variable holding time point of interest            */
/*  T - time point of interest                                */
/*  Z - covariate                                             */
/*  FSTIND - indicator of presence of failure event           */
/*  FXT - time to failure event                               */
/*  FLABEL - label describing failure event                   */

/* In this data step, observations pertaining to              */
/* time of interest are output.                               */

/* The data are sorted in order of decreasing time            */
/* to event to facilitate later computation of                */
/* summation terms. TCNT holds the count of patients          */
/* with time to event that is greater or equal to             */
/* that of the current observation.  This is the              */
/* "risk set" as defined by TSIATIS.                          */

/* The specified variables are placed into vectors            */
/* for analysis.                                              */

/* Initializing  summation terms. The function Q=1            */
/* defines the logrank statistic.  Every subject in           */
/* the file is accessed by the loop, however if a             */
/* subject has an indicator value of 0 they would             */
/* not effect the value of the test statistic or              */
/* covariance term and are therefore skipped to save          */
/* processing time.                                           */

/* This nested loop calculates the nested summation           */
/* term in the formula of the Tsiatis class of stat-          */
/* istics. Note that both this nested loop and the            */
/* next are exited when their counters reach the              */
/* count of patients.  This is to avoid an array              */
/* subscript overrun that would occur if the last             */
/* patients processed were tied on time to failure.           */
/* Contribution to test statistic by the ith patient          */

/* This loop computes the nested summation term in            */
/* the covariance term formula presented in Tsiatis.          */

/* Contribution to covariance term by the ith                 */
/* subject.                                                   */

/* The statistics pertaining to time &T are                   */
/* output and concatenated to a file holding                  */
/* all of the results produced.                               */

/* This loop is used to perform a desired analysis            */
/* by calling the macro and supplying it with param-          */
/* eters that contain the variables to be analyzed.           */
/* The loop counter identifies the time points of             */
/* the study and is sent as the second parameter.             */
/* The other parameters are described above.                  */

/* Printing the file of final results.                        */
```

APPENDIX B

SUMMARY OF TRIAL EVENTS OVER TIME (YEARS)

	72	73	74	75	76	77	78	79	80	81	82	TOTAL
GROUP 1 (HORMONAL THERAPY)												
Entrants	16	24	20	19	18	10	7	7	4	1	0	126
Deaths	0	0	0	6	4	3	3	5	4	8	10	43
Withdrawals	0	2	1	3	5	9	11	13	16	16	4	80
GROUP 2 (CONTROL)												
Entrants	23	19	19	21	18	27	17	16	12	10	0	182
Deaths	0	2	3	2	8	5	6	5	6	5	8	50
Withdrawals	0	2	3	7	15	14	16	17	14	9	4	101

APPENDIX C

LISTING OF DATA FOR ONE PATIENT

PTN	YR	TREAT	DSTIND	DXT
183	1	1	0	284
183	2	1	0	649
183	3	1	0	1014
183	4	1	0	1379
183	5	1	1	1745
183	6	1	1	1901
183	7	1	1	1901
183	8	1	1	1901
183	9	1	1	1901
183	10	1	1	1901
183	11	1	1	1901

Variable Descriptions

PTN - Patient Number
YR - Year of Study. Years 1 through 11 represent 1973 - 1983.
TREAT - Type of Treatment. (1=Active, 2=Control)
DSTIND - Indicator of Failure Event. (1=Patient Died, 0=Otherwise)
DXT - Time to Failure Event or Cut-Off for Analysis in Each Year

APPENDIX D

MACRO OUTPUT FOR PRESENTATION EXAMPLE

SURVIVAL ANALYSIS WITH DEATH AS FAILURE EVENT
AND TREAT AS THE COVARIATE

TIME POINT	LOGRANK STATISTIC	COVARIANCE TERM	CHI SQUARE STATISTIC	P-VALUE
2	0.82906	0.4836	1.42130	0.233190
3	2.40632	1.2461	4.64683	0.031111
4	0.80659	3.2278	0.20156	0.653467
5	2.88367	6.2140	1.33820	0.247352
6	5.32699	7.8164	3.63041	0.056733
7	6.97892	10.0422	4.85008	0.027645
8	7.53393	12.1944	4.65461	0.030970
9	8.95428	14.5047	5.52779	0.018717
10	8.63963	17.1381	4.35540	0.036892
11	8.84475	20.6868	3.78163	0.051818