A Nonparametric Approach for the Statistical Analysis of Time to First Event in Censored Paired Data

Roderick Lashley, Dennis King, Pam Marks
STATKING Consulting, Inc.

Abstract

Analyzing time to first event data is a common problem in drug development. For a randomized block design or crossover design yielding time to first event data on multiple treatments, standard nonparametric tests, such as the Log-Rank or Gehan Test prove to be inadequate since they do not take into account within subject correlation. This paper will examine a procedure commonly referred to as the Prentice-Wilcoxon Test that can be used to test for differences between the time to first event profile of two treatments tested in a randomized block design. We will use an example from the pharmaceutical industry to illustrate the implementation of this procedure with the SAS® System.

Introduction

In time to first event studies, clinical trials will sometimes consist of a randomized complete block (RCB) design with two treatments tested. The two treatments may be active drug and placebo, or two different active drugs. Two types of experimental designs are possible. In the first, each subject is randomly assigned one of the two treatments, the subject is monitored, and the time until the event takes place is recorded. If the event does not take place before a specified amount of time has elapsed, this is noted and treated as a censored data point. A washout period may then be used, during which the effect of the first treatment will disappear ("washout") from the subject. The subject is then given the other treatment, and is again monitored. As in the previous leg of the study, the time of the event, or the end of the study period is recorded.

In the second type of RCB design, two treatments are simultaneously used on, say two skin areas, of each subject. Time to event is measured on each of the skin sites. The time to event data may be censored. This design, as does the first, yields two correlated time to first event data values from each subject.

Prentice-Wilcoxon Test

We are interested in testing for treatment differences using time to first event data. Common nonparametric procedures are inadequate in this analysis, due to their inability to take into account within subject correlation. However, the Prentice-Wilcoxon test (Kalbfleisch and Prentice 1980) can be used in this situation. Given that we have *n* subjects with 2 observations from each subject, Le (1997) describes a test consisting of the following steps:

1) Pool the data to form a right-censored sample of size 2n.
2) Arrange the observations from smallest to largest. Let n_j be the number of events still waiting to happen at time t_j. Define

$$s_i = \prod_{j=1}^{i} \frac{n_j}{n_j + 1}$$

3) If the subject event is the i^{th} observed event, assign the score $c_i = 1 - 2s_i$.
   If the subject event is censored between the i^{th} and $(i+1)^{th}$ event, assign the score $c_i = 1 - s_i$.
4) Let the scores for the l^{th} pair be $(c_{1l} - c_{2l})$, and define $\Delta_l = c_{1l} - c_{2l}$ and

$$Z = \frac{\sum_{l=1}^{n} \Delta_l}{\left\{ \sum_{l=1}^{n} \Delta_l^2 \right\}^{1/2}}$$

Under the null hypothesis of no difference between the two members of pairs (i.e. no treatment difference), *z* has a standard normal distribution. This test has also been shown (Woolson & O'Gorman 1992) to provide adequate power regardless of the underlying distribution of the data.

SAS MACRO

We have written a program using SAS Macro Language to compute the scores for the

individual observations, the test statistic, and the associated p-value for the Prentice-Wilcoxon Test. The program takes as input a SAS dataset name (dsn), the name of the variable containing the time to event information (tim2evnt), the name of the variable containing censoring information (the value of the variable should be 1 if the data is censored and 0 otherwise) (censor), and a string containing the value of one of the two treatments (treat1). The macro computes the scores for each observation, the delta and delta-squared values, the test statistic, and the p-value for the Prentice-Wilcoxon Test. The test statistic and p-value are stored in macro variables, while the other data is captured in the output dataset. The code for the macro along with the macro call for the following example is shown in the Appendix.

Example

Consider the following example. Pattern baldness is a problem that affects many males. To test a treatment for the re-growth of hair, treatment can be applied to a selected area on the scalp. The number of days for the first new hair follicle to appear can be measured. If hair does not grow before the end of the study period, the number of elapsed days is recorded and the value is considered censored. Simultaneously the alternative treatment is applied to another area of the scalp, and the time to event data is also recorded on this site. Sample data for an experiment of this type follows.

Subject	Days to first growth Treatment A	Days to first growth Treatment B
1	12	17
2	9	7
3	19+	8
4	19+	10
5	8	7
6	9	8
7	9	10
8	19+	13
9	19+	8
10	19+	19+

A '+' indicates an observation that was censored.

Results

The results of analyzing the hair growth data with the tim2evnt macro (see Appendix) are as follows:

Analysis of Time to First Event
Using the Paired Prentice-Wilcoxon Test

Subject	Score A	Score B	Delta	Delta**2
1	-0.413	-0.131	-0.283	0.080
2	-0.684	-0.905	0.221	0.049
3	0.435	-0.805	1.239	1.536
4	0.435	-0.555	0.989	0.979
5	-0.805	-0.905	0.100	0.010
6	-0.684	-0.805	0.120	0.014
7	0.435	-0.554	0.989	0.979
8	0.435	-0.272	0.707	0.499
9	0.435	-0.805	1.239	1.536
10	0.435	0.435	0.000	0.000

$\sum_{l=1}^{n} \Delta_l = 5.322; \quad \sum \Delta_l^2 = 5.681;$

Z Test Statistic=2.23
P-Value=0.0256

The test shows a statistically significant difference between treatments A and B (p=.0256). Note that the times to first hair growth for treatment B are, in general, shorter than those of treatment A.

Conclusion

Clinical trials utilizing randomized block designs are of great value to the drug development process. This design reduces the number of subjects by allowing each subject to serve as their own control. However, the analysis of the resulting time to first event data is more difficult than in parallel studies. We have shown an application of the Prentice-Wilcoxon test using SAS® software that demonstrates a proper statistical analysis of this type of clinical trial.

References

Kalbfleisch, J. D. and Prentice, R.L. (1980), **The Statistical Analysis of Failure Time Data**, New York, NY: John Wiley & Sons.

Le, C.T. (1997), **Applied Survival Analysis**, New York, NY: John Wiley & Sons.

Woolson, R.F., and O'Gorman, T.W. (1992), A Comparison of Several Tests For Censored Paired Data, ***Statistics in Medicine**,* Volume 11, 193-208.

Author Contact

Roderick Lashley
Dennis King
Pam Marks
STATKING Consulting, Inc.
759 Wessel Drive - Unit 6
Fairfield, OH 45014
Phone: 513-858-2989
Fax: 513-858-3022
email: rod@statkingconsulting.com;
dennis@statkingconsulting.com;
pam@statkingconsulting.com

Appendix

```
%macro tim2evnt(dsn,tim2evnt,censor,treat1);
 ************************************;
** Macro to compute Prentice-Wilcoxon;
** test for pair-matched data.;
** STATKING Consulting, Inc.;
** October 1999;
** dsn= data set name ;
** tim2evnt = variable containing;
** time to event information;
** censor = variable containing;
** censor information ;
** treat1 = value of first treatment;
************************************;
proc sort data=&dsn; by &tim2evnt;
run;
   ** compute s_sub_i and score for each point;
  data &dsn; set &dsn nobs=nobs;
    retain s_sub_i 1;

** get time to event for last obs;
    lasttime=lag(&tim2evnt);


          ** compute s_sub_i;
    if ((lasttime ^= &tim2evnt) and (&censor ^= 1)) then
                                        ** nsubj / (nsubj+1);

s_sub_i=s_sub_i*((nobs-_n_+1)/(nobs-_n_+2));
          ** compute score;
    if (&censor=1) then score=1-s_sub_i;
      else score=1-(2*s_sub_i);
    drop lasttime;
  run;

  proc sort data=&dsn; by subject treatmnt;
  run;

 data &dsn; set &dsn end=eof; by subject treatmnt;
       ** y1 is score and censor info for first treatmnt;
       ** y2 is score and censor info for second treatmnt;
  length y1 $ 8  y2 $ 8;
  retain y1 y2 surv1 surv2 score1 score2;
  retain sumd 0 sumd2 0;
     ** get scores for treatment 1 and censoring info;
  if treatmnt= "&treat1" then do;
    score1=score;
    if censor=1 then y1=put(&tim2evnt,5.2)||'+';
    else y1=put(&tim2evnt,5.2);
    surv1=s_sub_i;
  end;
  else do;
     ** get scores for treatment 2 and censoring info;
    score2=score;
    if censor=1 then y2=put(&tim2evnt,5.2)||'+';
    else y2=put(&tim2evnt,5.2);
      surv2=s_sub_i;
  end;
          ** compute delta_sub_l, sum of delta_sub_l
          and; ** sum    of delta_sub_l squared;
  if last.subject then do;
    delta=score1-score2;
    delta2=delta**2;
    if (delta ^= .) then do;
      sumd=sumd+delta;
      sumd2=sumd2+delta2;
    end;
       ** output scores and summations;
    output;
  end;
       ** end of data, compute z statistic and probability;
  if eof=1 then do;
    z=sumd/(sqrt(sumd2));
    probz=2*(1-probnorm(abs(z)));
    call symput('z',put(z,4.2));
    call symput('probz',put(probz,6.4));
  end;
  drop &tim2evnt score treatmnt &censor z probz;
 run;
%mend tim2evnt;

%tim2evnt(dsn=data, tim2evnt=hours, censor=censor,
treat1=PLACEBO);
```