The Statistician (1988) 37, pp. 59–64

The crossing hazard function problem

NATHAN MANTEL[1] & DONALD M. STABLEIN[2]

[1]*Department of Mathematics, Statistics & Computer Science, American University, 4900 Auburn Avenue, Bethesda, MD 20814 and* [2]*The EMMES Corporation, 11325 Seven Locks Road, Potomac, MD 20854, U.S.A.*

Abstract. Application for the Mantel-Haenzel statistic to the analysis of survival distributions with crossing hazard functions is discussed. Both the situation of a prespecified crossing time and a suspected crossing sometime during the study's course are examined. An example from a cancer clinical trial is included.

1 Introduction

Real-life situations in which there is a reversal at some point in time in the direction of the difference between two survival or time-to-response distributions are attracting current statistical attention. Fleming *et al.* (1980) discuss a Kolmogorov-Smirnov statistic which is sensitive to crossing hazard situations. Stablein *et al.* (1981) focused attention on a situation in which initial analysis had shown limited difference in two survival curves over the full duration of a clinical study. But re-examination of the data showed that the evident superiority for one treatment early on in the study was negated by its apparent inferiority after about a year. Mantel (1980) alluded to the Red Dye 40 problem in which an acceleration model was used to explain why the early evidence for a neoplastic effect of Red Dye 40 was not reflected by the results of full-duration statistical analysis—the tumors had been accelerated in appearance by the Red Dye 40 with a consequent deficiency in later appearing tumors. Breslow *et al.* (1984), Stablein & Koutrouvelis (1985) have continued the examination of distributional testing under acceleration or crossing hazard alternatives.

Mantel (1966) discussed the situation of two exact survival patterns (i.e. without sampling variability) which crossed each other at one or more points in time. It was pointed out that one cannot properly address the problem of comparing two sample survival curves when it is unclear how to say which is the superior of two exactly-known but crossing survival curves. Specifically, the curves relative superiority is a function of one's weighting or value system.

In their work comparing hazard functions, Stablein *et al.* (1981) followed the lead of Cox (1972) by which no functional form for the survival distribution is postulated. But their approach did have to postulate a functional form by which the relative hazard function varied with time; a quadratic relationship in time was used. The analysis of their example demonstrated that patients with non-resectable gastric carcinomas who had received both radiation and chemotherapy had initially poorer survival, but after about 1 year, surviving combined modality patients showed lower risks of death than did survivors who had received chemotherapy alone. While it was the crossing hazard phenomenon which had brought attention to the situation the fact that the relative hazard function was not constant was the focus of the solution. The time-dependence of the relative hazard function does not necessarily connote that the hazard functions cross during the time period of interest. Thus, if the relative hazard

function had remained, though time-dependent, on one side of unity there would probably have been little concern that the time dependence had been ignored.

The present work is concerned with the crossing hazard function aspect of the problem rather than with the aspect of non-constancy of the relative hazard function. Note that in order for survival curves to cross an earlier crossing of the hazard is required. While not referring to the multiple strata or covariates, any methodology to be suggested should be equally suitable to the homogeneous case and to the case of multiple strata.

2 The crossing point pre-specified, exactly or approximately

Consider the following situation. Suppose that initially two survival curves separate, with one showing the lesser survival probability and greater cumulative probability of death. The two curves may later come back together and, should they cross, the curve previously showing the higher survival probability would subsequently show the lower survival probability. (If data as of the crossing point were analysed by some logrank approach, the fact that the survival curves crossed would not preclude the possibility of a statistically significant result in favor of the therapy that had seemed superior at the start). A reversal in the relative position of the hazard function (i.e. force-of-mortality or instantaneous death rate) is required prior to the survival curves coming together again.

For one survival curve to show initially poorer survival would have required that its hazard function was originally worse of the two—but for the survival curves to come back together would have additionally required that from some later point, the second curve must have a cumulatively higher hazard function. Suppose we have some advance basis for knowing that at a time point, s, there is a reversal in the merits of two hazard functions. Admittedly, this is a difficult situation to envisage, but, for completeness, it warrants consideration. Also, it permits contrast with the handling of other situations.

The logrank computing procedure as given in Mantel (1966) neatly resolves the situation. That computing procedure requires that at each response time, or for each short observation period, a 2×2 contingency table be constructed. Then, following the methodology of Mantel & Haenszel (1959), one obtains for each contingency table the observed, and expected, number of failures in, say, the study group, together with the associated hypergeometric variance. That is for the ith table with N_{ji} subject and O_{ji} failures on treatment j ($j=1,2$),

$$E_i=\frac{N_{1i}\,(O_{1i}+O_{2i})}{N_{1i}+N_{2i}}$$

and

$$V_i=\frac{N_{1i}\,N_{2i}\,(O_{1i}+O_{2i})\,(N_{1i}+N_{2i}-O_{1i}-O_{2i})}{(N_{1i}+N_{2i})^2\,(N_{1i}+N_{2i}-1)}$$

But now, instead of considering the cumulated difference between observed and expected over all tables, one considers separately the difference of the cumulated differences prior to and subsequent to the anticipated crossing point.

Thus, we define the statistic

$$W_2=[\Sigma(O_{1i}-E_i)\;Z\,(t_i,S)]^2/V$$

where

$$Z(t_i,s)=\begin{cases} 1 & t_i<s \\ -1 & t_i>s \end{cases}$$

and

$V=\Sigma V_i$. W_s is distributed as a X_1^2.

3 Letting the data suggest the crossing point

A more likely situation than knowing where the true crossing point occurred would be one in which we suspected the hazard functions to cross, but had no prior basis for designating the point in time. Minimally, we can see where the data suggest the crossing point to be.

For this purpose, suppose we try every possible point in time as a candidate crossing point. Actually we need consider as potential candidates only a single time point between successive sample response times. For each candidate crossing time we would proceed as in the preceding section, cumulating the difference between observed and expected prior to the candidate crossing point and adding the negative of the subsequent cumulated difference. The crossing point suggested by the data would then be the time point or rather time interval between two successive failures, for which the calculated chi-square is maximal (Note a relationship to the suggestion in Mantel (1966) for using the maximal chi-square over time as a test statistic as developed by Muenz *et al.* (1977).)

One can calculate

$$W=\sup_s(W_s)$$

the maximal chi-square statistic to test for the equality of the two survival distributions. The null distribution of W can be determined by simulation. Table 1 contains for the given sample sizes and censoring levels simulated (Monte Carlo) critical values for the statistic obtained from 2000 random samples of distinct survival times. Random samples under the null distribution of distributional equality were generated as in Meunz *et al.* (1977) using the GGUBFS (IMSL (1980)) random number generator. When censored, a uniform censoring variate was generated for each observation and applied to the same sequence of failure times, for the 20% censoring level the same sequence of censoring variates were used as for the 10% level.

Table 1. Maximum chi-square critical values for test of distributional equality

Sample size per treatment	% censoring	Critical values =0·10	=0·05	=0·01
20	0	6·69	8·33	12·04
	10	6·52	8·10	12·42
	20	6·82	8·23	11·80
30	0	6·59	8·17	11·19
	10	6·71	8·20	11·13
	20	6·49	7·91	10·68
40	0	6·57	8·19	11·49
	10	6·55	7·94	11·82
	20	6·58	8·07	11·22
50	0	6·56	8·12	11·11
	10	6·36	7·91	11·32
	20	6·68	7·93	12·13

If there is a separation, but no crossing of the hazard functions, the maximum of the chi-square value so described should tend to be after the last, or possibly before the first, observed failure time. But due to chance variation, it would not be unlikely to have the maximal chi-square correspond to one of the very early or one of the very late intervals between successive failures. Examples would be where either the first or the last response time corresponded to the group which otherwise displayed the lower degree of risk. In all this, we must recognise that there can be no real crossing point unless the relative hazard function differs from unity. If there is no treatment difference, there can be no crossing point.

4 Illustration

The data set employed by Stablein *et al.* (1981) are used for illustration purposes here. These data report on a experiment with 45 individuals receiving both chemotherapy and radiation treatment and another 45 receiving only chemotherapy for the treatment of locally unresectable gastric cancer. Of the 45 receiving the combined modality therapy, 37 were observed to die, with death times ranging from day 17 to day 1366 while the range of still-alive times for the eight survivers at the time of analysis ranged from 882 to 1472 days. The death times for the 38 failures on chemotherapy ranged from day 1 to 1271, the still-alive times for the remaining 7 range from day 381 to day 1519. Figure 1 displays the survival and log (−log (survival)) curves for the experiment; each visually indicates a lack of constancy of the relative hazard function.

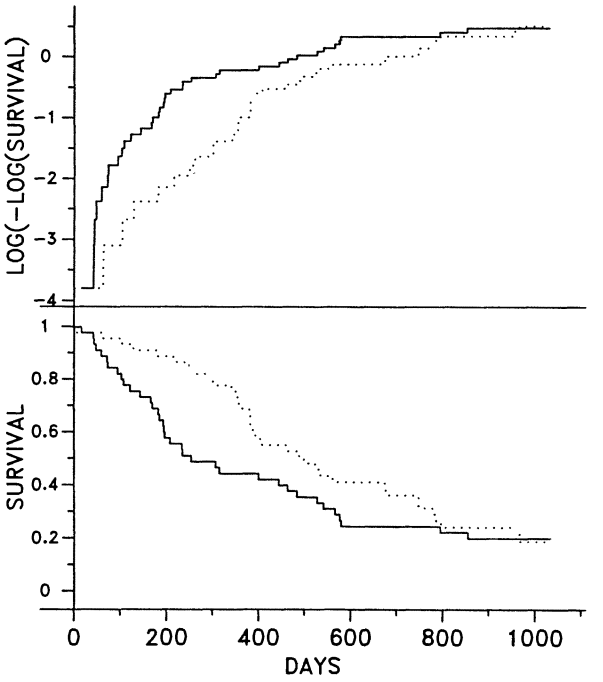

Fig. 1. Estimated survival and log (−log (survival)) curves by treatment of gastric cancer experiment ——— chemotherapy+radiation; — — — chemotherapy only.

The overall Mantel-Haenszel chi-square statistic is calculated to be

$(37-32{\cdot}13)^2/18{\cdot}03=1{\cdot}32$ with a p value of 0·254. Had one prespecified the alternative hypothesis of a crossing of the relative hazard functions at one year, the observed and expected death in the two sets created by partitioning at 365 days could be determined. The statistic W_{365} is calculated to be $[(25-16{\cdot}69)-(12-15{\cdot}44)]^2/18{\cdot}03=7{\cdot}65$ $p<0{\cdot}01$. Thus the implications of the two tests are in conflict.

More realistically one may want to protect against the possibility of crossing hazards but may not be able to prespecify the time point for the suspected crossing. For this data set there are 70 distinct non-empty partitions. The statistics W_s are plotted against time in Fig. 2. Note that for all partitions between 220 and 350 days, the statistic exceeds the 0·05 level criterian from Table 1. The maximum statistic over all partitions, W, equals 12·05 for the appparent hazard crossing between days 315 and 352. It is clear that even by searching for the maximal chi-square a difference as large as that observed is unlikely to occur if the distributions are the same. Thus it seems reasonable to claim the distributions are significantly different as the observed value is sufficiently extreme.

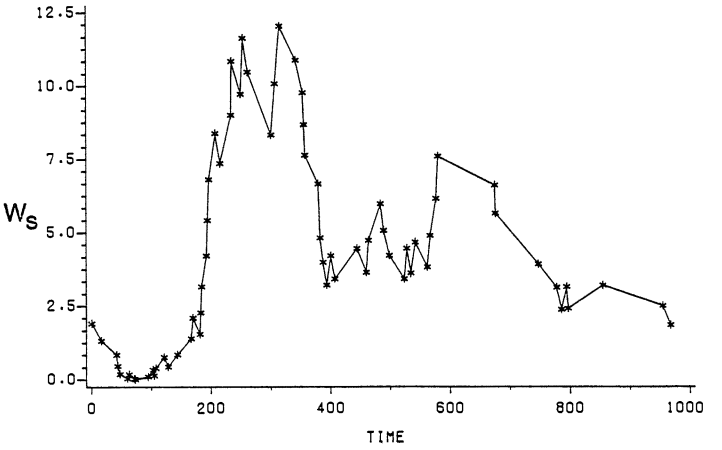

Fig. 2. Chi-square statistics W_s for gastric cancer experiment.

At 11 months, the estimated time of the hazards crossing, a nearly 40 percentage point survivorship benefit for chemotherapy is observed. In fact, the study was terminated early because of this substantial early survival advantage. By 3 years the curves had come together and further follow-up of these patients showed a persistant 10% advantage beyond 5 years for the combination arm. Increased cures, not just short term survival advantage, is the objective of treatment in this disease stage. It is hoped that radiation provides for sterilization of locally residual disease present after surgery and thus an increased cure rate. In the combination arm chemotherapy was delayed for at least 10 weeks during irradiation and it was hypothosized that this may have permitted disseminated disease spread. A current protocol addresses this issue by sandwiching radiation between chemotherapy treatments.

Discussion

Continued developments in the analyses of censored survival data when non-constant hazard ratios exist are required. Clearly the local optimality of the log rank procedure

for detection of constant hazard ratio differences is well known. The procedure discussed here is not intended as a competitor in this situation. When non-constant hazard ratios exist multiple analyses may be required, and each has attendant advantages and disadvantages. For example the modified Smirnov procedure has good power properties when hazards are equal initially and diverge at a later time point. Alternatively the proposed approach provides for an estimate of the time of the hazards crossing point. When hazard functions cross one may be interested in maximising cure rates (i.e. the tail of the curve), the mean survival, the median survival or a plethora of other approaches which could be contemplated. Selection of the objective will be a function of the medical problem, prior knowledge and the analysis technique available. The approach detailed herein should be useful when one has a single crossing of the hazard functions.

Acknowledgement

Participation by N. Mantel supported by Grant C.A. 30205 from the National Cancer Institute.

References

BRESLOW, N.E., EDLER, L. & BERGER, J. (1984) A two sample censored data rank test for acceleration, *Biometrics*, 40, pp. 1049–1062.

COX, D.R. (1972) Regression models and life tables (with discussion), *Journal of the Royal Statistical Society Series B*, 34, pp. 187–220.

FLEMING, T.R., O'FALLON, J.R., O'BRIEN, P.C. & HARRINGTON, P.P. (1980) Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right censored data, *Biometrics*, 36, pp. 607–625.

MANTEL, N. (1966) Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Reports*, 50, No. 3, pp. 163–170.

MANTEL, N. (1980) Assessing laboratory evidence for neoplastic activity, *Biometrics*, 36, pp. 381–399.

MANTEL, N. & HAENSZEL , W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of the National Cancer Institute*, 22, pp. 719–748.

MUENZ, L.R., GREEN, S.B. & BYAR, D.P. (1977) Application of the Mantel-Haenszel statistic to the comparison of survival distributions, *Biometrics*, 33, pp. 617–626.

STABLEIN, D.M., CARTER, W.H. & NOVAK, J.W. (1981) Analysis of survival data with nonproportional hazard functions, *Controlled Clinical Trials*, 2, pp. 149–159.

STABLEIN, D.M., & KOUTROUVELIS, I.A. (1985) A two sample test sensitive to crossing hazards in uncensored and singly censored data, *Biometrics*, 41, pp. 643–652.

STARK, C.R. & MANTEL, N. (1969) Maternal-age and birth-order effects in childhood leukemia. Age of child and type of leukemia, *Journal of the National Cancer Institute*, 42, pp. 857–866.