GROUP SEQUENTIAL METHODS FOR SURVIVAL ANALYSIS WITH STAGGERED ENTRY

Anastasios A. Tsiatis

Harvard University School of Public Health and Sidney Farber Cancer Institute Boston, Massachusetts

1. Introduction

In many clinical trials, especially in the study of chronic disease, we are often interested in the comparison of time to failure among different treatment groups. Many procedures have been developed in the past decade to analyze such failure time data, the most popular being the logrank test (Mantel, 1966; Peto and Peto, 1972) and modifications of the Wilcoxon test (Gilbert, 1962; Gehan, 1965; Breslow, 1970; Peto and Peto, 1972; Prentice, 1978).

Typically in such a trial, patients enter the study serially, are then assigned according to some random mechanism to different treatment arms and are followed until they either fail or the study is terminated. Ordinarily, these studies are designed so that after sufficient amount of patient accural and follow-up time a single terminal analysis will be made to test whether the failure time distribution is the same among the different treatment groups. In practice, however, as well as for ethical considerations, the data are monitored periodically and if sufficient differences are found between the treatment groups, a decision might be made to stop the study early. It is, therefore, very important to study the sequential properties of the tests used in survival analysis in order that correct and efficient methods be employed in monitoring the data.

Breslow (1969) and Breslow and Haug (1972) provide sequential methods of comparing exponential distributions. However, most of the recent work examines the more commonly used nonparametric statistics. Jones and Whitehead (1979) have looked at the sequential logrank and modified Wilcoxon tests. Nagelkerke and Hart (1980) also indicate how to extend the sequential probability ratio test using partial likelihoods. In a more rigorous fashion Tsiatis (1981) has derived the group sequential distribution of the logrank score and Slud and Wei (1982) that of the modified Wilcoxon score.

Most of the commonly used nonparametric statistics are special cases of a general class characterized by Tarone and Ware (1977) and Prentice and Marek (1979). In this paper, the asymptotic joint distribution of the sequentially computed test statistics, within this general class of nonparametric tests, will be derived.

These results will prove useful in constructing group sequential procedures.

2. Notation and Formulae

Let the positive random variables X,Y denote failure time and time of entry into study, respectively. Also let the positive random variable W denote time to censoring such as loss to follow-up. We wish to test the null hypothesis, H_0, that the hazard rate for failure is not related to a covariate Z. That is

$$H_0 : \lambda(x|z) = \lambda(x) ,$$

for all $x \geq 0$, where $\lambda(x|z)$ denotes the conditional hazard rate at time x given that the covariate Z is equal to z.

The time of entry into study, Y , and time to censoring, W , will have distributions $H(y|z) = P(Y \leq y|Z = z)$ and $G(w|z) = P(W \leq w|Z = z)$, respectively, which may depend on the covariate Z . We will denote the survival distribution $1 - G(w|z)$ as $\bar{G}(w|z)$. Assume that, given the covariate Z, the random variables

X,Y,W are independent.

If the data were examined at time t, the following variables could be observed; time to failure or censoring $X(t) = \max\{\min(X, t-Y, W), 0\}$, and an indicator variable for failure $\Delta(t) = 1$ if $X < \min(t-Y, W)$, $\Delta(t) = 0$ otherwise. At time t, the data can be represented as n identically and independently distributed random vectors $\{X_i(t), \Delta_i(t), Z_i\}$ for $i = 1,\ldots,n$.

The class of tests for testing H_0 will be similar to those of Tarone and Ware (1977) and Prentice and Marek (1979). Using the notation of this paper, we define a class of tests Θ, characterized by statistics of the form

$$S_n(t) = \sum_{i=1}^{n} \hat{Q}(t,X_i(t))\ \Delta_i(t)\ \{Z_i - \sum_{j \in R(t,X_i(t))} Z_j/n(t,X_i(t))\} \quad ,$$

where R(t,x) denotes the risk set at time x if the data were observed at real time t, $x \leq t$. That is, R(t,x) denotes the set of indices $\{j = 1,\ldots,n\}$ such that $\{X_j(t) \geq x\}$. Letting I(A) denote the indicator function of the event A, then $n(t,x) = \sum_{j=1}^{n} I(X_j(t) \geq x)$. For fixed t, the random function $\hat{Q}(t,x)$, $x \leq t$, is assumed to converge in probability in sup norm to a function Q(t,x).

The random function $\hat{Q}(t,x)$ corresponds to the weighting functions W_i described by Tarone and Ware (1977) and Prentice and Marek (1979). In particular, the most widely used nonparametric tests can be represented as follows:

Example 1. For the logrank test, $\hat{Q}(t,x) = Q(t,x) = 1$ for all $t > 0$, $0 \leq x \leq t$.

Example 2. The weighting function for the modified-Wilcoxon test is given by $\hat{Q}(t,x) = n(t,x)/n$, which converges in probability to the function $Q(t,x) = P(X(t) \geq x)$. Due to the independence of X,Y,W given Z, this can be expressed, under H_0, as

$$
\begin{aligned}
P(X(t) \geq x) &= E\{P(X(t) \geq x) | Z\} \\
&= E\{P(X \geq x,\ t - Y \geq x,\ W \geq x) | Z\} \\
&= \exp\{-\Lambda(x)\}\ E\{H(t-x) | Z\}\ E\{\bar{G}(x) | Z\} \quad .
\end{aligned} \tag{1}
$$

Example 3. The weighting function for Prentice's (1978) generalization of the Wilcoxon test can be expressed as

$$\hat{Q}(t,x) = KM(x) = \prod_{i=1}^{n} [n(t,X_i(t))/\{n(t,X_i(t)) + 1\}]^{N_i(t,x)} \quad ,$$

where

$$N_i(t,x) = I(X_i(t) \leq x,\ \Delta_i(t) = 1) \quad . \tag{2}$$

We note that, under H_0, $Q(t,x)$ is approximately the same as the Kaplan and Meier (1958) estimate of the survival distribution. Consequently, $Q(t,x) = \exp\{-\Lambda(x)\}$.

Example 4. The tests based on the score statistics, called G^ρ tests, proposed by Harrington and Fleming (1981) have weighting functions $\hat{Q}(t,x) = \{KM(x)\}^\rho$, and hence $Q(t,x) = [\exp\{-\Lambda(x)\}]^\rho$.

3. Asymptotic Joint Distribution of the Statistic

The key to deriving the joint distribution of the statistic $S_n(t)$ over time is to approximate it by a sum of i.i.d. random variables. All subsequent results are assumed under the null hypothesis.

We first note that the general statistic can be written as

$$S_n(t) = \sum_{i=1}^{n} \int_0^t dN_i(t,x)\ \hat{Q}(t,x)\ \{Z_i - \sum_{j \in R(t,x)} Z_j / n(t,x)\} \quad , \tag{3}$$

where $N_i(t,x)$ is given by (2).

For fixed t, $N_i(t,x)$ is a counting process as a function of x with intensity process given by $\lambda(x)\ I(X_i(t) \geq x)$. Therefore, using the results of Aalen (1977, 1978), the process

$$M_i(t,x) = N_i(t,x) - \int_0^x \lambda(u)\ I(X_i(t) \geq u)du$$

is a martingale. The statistic (3) can then be shown to be equal to

$$S_n(t) = \sum_{i=1}^{n} \int_0^t \hat{Q}(t,x)\ dM_i(t,x)\ \{Z_i - \sum_{j \in R(t,x)} Z_j/n(t,x)\} \quad .$$

Because of the complex relationship between t and x, the general martingale approach of Aalen is not directly applicable in characterizing the asymptotic joint distribution of the process $S_n(t)$. However, by noting that

$$\sum_{j \in R(t,x)} Z_j/n(t,x)$$

converges in probability to

$$(4) \qquad \mu(t,x) = [E\{ZH(t-x|Z)\ \bar{G}(x|Z)\}] / [E\{H(t-x|Z)\ \bar{G}(x|Z)\}] \quad ,$$

and $\hat{Q}(t,x)$ converges in probability to $Q(t,x)$, then the statistic $n^{-\frac{1}{2}}S_n(t)$ can be approximated by $n^{-\frac{1}{2}}\bar{S}_n(t)$, where

$$(5) \qquad \bar{S}_n(t) = \sum_{i=1}^{n} \int_0^t Q(t,x)\ dM_i(t,x)\ \{Z_i - \mu(t,x)\} \quad .$$

The difference between $n^{-\frac{1}{2}}S_n(t)$ and $n^{-\frac{1}{2}}\bar{S}_n(t)$ can be shown to be asymptotically negligible by using results similar to Tsiatis (1981b, Lemma 3.1).

The approximate statistic $\bar{S}_n(t)$ can be written as

$$\bar{S}_n(t) = \sum_{i=1}^{n} [\Delta_i(t)\, Q(t,X_i(t))\, \{Z_i - \mu(t,X_i(t))\} - \int_0^{X_i(t)} Q(t,x)\, \{Z_i - \mu(t,x)\}\, \lambda(x)dx] \quad . \tag{6}$$

Although (6) is complex it is nonetheless a sum of identically and independently distributed random variables and the asymptotic distribution can be obtained by application of the central limit theorem. We are now in a position to prove the following fundamental theorem.

THEOREM 3.1. Defining the statistics in Θ by

$$S_n^{(1)}(t) = \sum_{i=1}^{n} \hat{Q}_1(t,X_i(t))\, \Delta_i(t)\, \{Z_i - \sum_{j\in R(t,X_i(t))} Z_j/n(t,X_i(t))\} ,$$

and

$$S_n^{(2)}(t') = \sum_{i=1}^{n} \hat{Q}_2(t',X_i(t'))\, \Delta_i(t')\{Z_i - \sum_{j\in R(t',X_i(t'))} Z_j/n(t',X_i(t'))\} ,$$

where $t' \geq t$; then the random vector $n^{-\frac{1}{2}}\{S_n^{(1)}(t),\ S_n^{(2)}(t')\}$ converges in distribution to a bivariate normal distribution with mean zero and covariance matrix

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

where

$$\sigma_{11} = \int_0^t Q_1^2(t,x)\, \psi(t,x)\, \lambda(x)dx ,$$

$$\sigma_{12} = \int_0^t Q_1(t,x)\, Q_2(t',x)\, \psi(t,x)\, \lambda(x)dx , \tag{7}$$

$$\sigma_{22} = \int_0^{t'} Q_2^2(t',x)\, \psi(t',x)\, \lambda(x)dx ,$$

and $$\psi(t,x) = E[\{Z - \mu(t,x)\}^2\, H(t-x|Z)\, \bar{G}(x|Z)]\, \exp\{-\Lambda(x)\} .$$

PROOF. Because of the asymptotic approximation of

$$n^{-\frac{1}{2}}\{S_n^{(1)}(t),\ S_n^{(2)}(t')\} \text{ for } t' \geq t \ ,$$

to

$$n^{-\frac{1}{2}}\{\bar{S}_n^{(1)}(t),\ \bar{S}_n^{(2)}(t')\} \ ,$$

it suffices to find the asymptotic joint distribution of the latter. Since this is a normalized sum of i.i.d. random vectors then a routine application of the multivariate central limit theorem together with the calculation of the moments given in the appendix of Tsiatis (1982) will yield the desired results.

It can also be shown by using the results of Tsiatis (1981a) that the covariance matrix Ω can be consistently estimated by replacing the quantities in (7) by their appropriate empirical estimates. In particular, the estimate of σ_{12} is given by

$$\hat{\sigma}_{12} = \int_0^t \hat{Q}_1(t,x)\ \hat{Q}_2(t',x)\ \hat{E}[\{Z - \hat{\mu}(t,x)\}^2\ I(X(t) \geq x)]\ d\hat{\Lambda}(x) \ ,$$

where

$$\hat{E}[\{Z - \hat{\mu}(t,x)\}^2\ I(X_i(t) \geq x)] = \sum_{j \in R(t,x)} \{Z_j - \hat{\mu}(t,x)\}^2/n \ ,$$

and $\hat{\Lambda}(x)$ is the estimate of the cumulative hazard function given by Nelson (1969), namely

$$\hat{\Lambda}(x) = \int_0^x dN_i(t,u)/n(t,u) \ .$$

COMMENT: The assumptions in this paper will apply to a completely randomized clinical trial but will not reflect a study with dynamic treatment assignment. However, with some modifications to the proofs we can allow the covariates $Z_1,\ldots,Z_n$ to be fixed values. Replacing $\mu(t,x)$ and $\psi(t,x)$ given in (4) and (7)

by the following quantities

$$\mu^n(t,x) = \sum_{i=1}^{n} Z_i H(t-x|Z_i)\ \bar{G}(x|Z_i) \Big/ \sum_{i=1}^{n} H(t-x|Z_i)\ \bar{G}(x|Z_i)$$

and

$$\psi^n(t,x) = \sum_{i=1}^{n} \{Z_i - \mu^n(t,x)\}^2\ H(t-x|Z_i)\ \bar{G}(x|Z_i)\ \exp\{-\Lambda(x)\} / n \quad ,$$

and by assuming that $\mu^n(t,x)$ and $\psi^n(t,x)$ converge to $\mu^*(t,x)$ and $\psi^*(t,x)$, as would be the case for dynamic treatment assignment, then the statistic $S_n(t)$ can be approximated by

$$\bar{S}_n(t) = \sum_{i=1}^{n} \int_0^t Q(t,x)\ dM_i(t,x)\ \{Z_i - \mu^n(t,x)\} \quad .$$

We can then prove the results of Theorem 3.1 for fixed covariates by applying a central limit theorem for independent but not identically distributed random variable to $n^{-\frac{1}{2}}\{\bar{S}_n^{(1)}(t), \bar{S}_n^{(2)}(t')\}$. In this particular case the asymptotic covariance matrix is given by σ_{ij}^* which is obtained by substituting $\psi^*(t,x)$ for $\psi(t,x)$ in formula (7). However, the consistent estimate for σ_{ij}^* is the same as $\hat{\sigma}_{ij}$ given above and hence all applications to group sequential tests given in this paper would be identical.

4. Concluding Remarks

The asymptotic convergence of the joint distribution of test statistics within the class $\mathcal{O}$, evaluated at time points $t_1 < t_2 < \cdots < t_k$, to a multivariate normal with mean zero, and covariance matrix that can be estimated, will enable us to construct group sequential tests at those time points by using methods described by Slud and Wei (1982, Section 4). In particular, we shall consider group sequential tests that will reject the null hypothesis if

$$|S_n(t_1)| \geq d_1$$

or

$$|S_n(t_1)| < d_1, |S_n(t_2)| \geq d_2$$

or

$$\vdots$$

$$|S_n(t_1)| < d_1, \ldots, |S_n(t_{k-1})| < d_{k-1}, |S_n(t_k)| \geq d_k \quad ,$$

otherwise we accept H_0. The boundary values $d_1,\ldots,d_k$ are derived by choosing $\alpha_1,\ldots,\alpha_k$ so that $\alpha_1+\cdots+\alpha_k=\alpha$ and then recursively solving the following equations:

$$P(|U_1| \geq d_1) = \alpha_1$$

where $U_1 \sim N(0,\hat{\sigma}_{11})$

$$P(|U,| < d_1, |U_2| \geq d_2) = \alpha_2$$

where $(U_1,U_2) \sim N\left(\underset{\sim}{0},\begin{pmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{12}\\ \hat{\sigma}_{12} & \hat{\sigma}_{22}\end{pmatrix}\right)$,

etc.

This method will guarantee an overall level of significance equal to some prespecified α. In order to apply these methods in practice, however, numerical methods for calculating multivariate normal integrals have to be used. Currently such methods are very inefficient. However, in many cases, as we shall indicate later, the score statistic has uncorrelated increments and for such instances recursive integration formulas given by Armitage, McPherson and Rowe (1969) can be used to solve the above equations.

Using the results of Theorem 3.1, we have the flexibility of constructing the group sequential test by using different test statistics within the class Θ at any of the time points $t_1,\ldots,t_k$. Since different tests are more sensitive to different types of alternatives this might allow us to construct more robust group sequential procedures.

We also note that if the weighting function $\hat{Q}(t,x)$ converges to a function $Q(x)$, independent of t, then the score function $S_n(t)$ has asymptotically independent increments. This follows the fact that σ_{12} is equal to σ_{11}.

Therefore, the statistic for the sequentially computed logrank test, Prentice's generalization of the Wilcoxon test, and the G^ρ tests, described by Examples 1, 3, and 4, respectively, have independent increments. However, the modified-Wilcoxon test of Example 2 would not have this property unless all patients entered at once into the study. This contradicts Jones and Whitehead (1979) but supports the results of Slud and Wei (1982).

The results of Theorem 3.1 can be used to find the asymptotic joint distribution of any finite number of test statistics within the class θ. Therefore, the asymptotic joint distribution of the logrank test and modified-Wilcoxon test derived by Tarone (1981) would be a simple consequence of Theorem 3.1.

ACKNOWLEDGEMENT

This investigation was supported by grant NO. CA-23415, awarded by the National Cancer Institute, DHHS.

REFERENCES

Aalen, O.O. (1977). Weak convergence of stochastic integrals related to counting processes. Z. Wahrscheinlichkeitstheorie verw. Geb. 38, 261-277.

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. Annals of Statistics 6, 701-726.

Armitage, P., McPherson, C.K. and Rowe, B.C. (1969). Repeated significance tests on accumulating data. Journal of the Royal Statistical Society A 132, 235-244.

Breslow, N. (1969). On large sample sequential analysis with applications to survivorship data. Journal of Applied Probability 6, 261-274.

Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. Biometrika 57, 579-594.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Annals of Statistics 2, 437-453.

Breslow, N. and Haug, C. (1972). Sequential comparison of survival curves. Journal of the American Statistical Association 67, 691-697.

Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. Biometrika 53, 203-223.

Gilbert, J.P. (1962). Random censorship. Unpublished PhD thesis, Department of Statistics, University of Chicago.

Harrington, D.P. and Fleming, T.R. (1981). A class of rank test procedures for censored survival data. Technical Report Series No. 12, Section of Medical Research Statistics, Mayo Clinic. To appear in Biometrika, 1983.

Jones, D. and Whitehead, J. (1979). Sequential forms of the logrank and modified-Wilcoxon tests for censored data. Biometrika 66, 105-113.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457-481.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports 50, 163-170.

Nagelkerke, N.J.D. and Hart, A.A.M. (1980). The sequential comparison of survival curves. Biometrika 67, 247-249.

Nelson, W. (1969). Hazard plotting for incomplete failure data. Journal of Quality Technology 1, 27-52.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society, A 135, 185-206.

Prentice, R.L. (1978). Linear rank tests with right censored data. Biometrika 65, 167-179.

Prentice, R.L. and Marek, P. (1979). A qualitative discrepancy between censored data rank tests. Biometrics 35, 861-867.

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. (To appear).

Tarone, R.E. (1981). On the distribution of the maximum of the logrank statistic and modified-Wilcoxon statistic. Biometrics 37, 79-86.

Tarone, R.E. and Ware, J. (1977). On distribution free tests for equality of survival distributions. Biometrika 64, 156-160.

Tsiatis, A.A. (1981a). A large sample study of Cox's regression model. Annals of Statistics 9, 93-108.

Tsiatis, A.A. (1981b). 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 score statistics used in censored survival analysis. Journal of the American Statistical Association. (To appear).