ESTIMATION OF THE RATIO OF HAZARD FUNCTIONS

John Crowley

University of Washington and Fred Hutchinson

Cancer Research Center

P.Y. Liu

University of California at Los Angeles

Joseph G. Voelkel

Allied Chemical Corporation

1. Introduction

In a clinical trial in which comparison of survival across treatment groups is of interest, it is useful to have a descriptive measure of the difference in survival between groups. If the hazard functions in two groups are roughly proportional, then the ratio of hazard functions has the interpretation of relative risk and has intuitive appeal as a descriptive statistic.

In this paper we investigate the large sample properties of several estimators of relative risk. The experimental setting envisioned is the clinical trial or other similar situations in which survival is being measured, and in which there are possibly different potential follow-up times for each patient.

The notation and model are given in Section 2. Section 3 presents the maximum likelihood estimator based on an exponential model, and the maximum partial likelihood estimator of Cox (1972;1975). Various other approaches are indicated in Section 4, and one, the Mantel-Haenszel (1959) estimator, is investigated in detail. Some numerical comparisons are made in Section 5.

2. Notation

For the i^{th} sample, $i=0,1$, we assume that the true survival times are observations on a non-negative random variable X_i^0, with absolutely continuous cumulative distribution function F_i. The hazard functions in each sample are defined to be $\lambda_i(t) = \frac{dF_i(t)}{1-F_i(t)}$; we assume that $\lambda_1(t) = \theta\,\lambda_0(t)$ (which implies that $1-F_1 = (1-F_0)^{\theta}$). Our objective is to investigate several estimators of θ, the relative risk.

In general, the possible follow-up time is limited by a fixed end point to the study, while patients enter a trial at possibly different times. The potential follow-up may thus vary from patient to patient, and may be modeled as non-negative random variables T_i (possibly depending on the sample) with cumulative distribution functions C_i. The T_i are assumed to be independent of the X_i^0 and what is observed is a survival time $X_i = \min\{X_i^0, T_i\}$, as well as whether the observation is uncensored ($X_i = X_i^0$) or censored ($X_i = T_i$). (This random censorship model is, in fact, more restrictive than necessary, and is used here primarily for ease of exposition.)

The distribution function of the survival times X_i is given by $H_i = 1 - (1-C_i)(1-F_i)$, and the subdistribution function of the observed deaths from each sample will be denoted by G_i, where $dG_i(t) = (1-H_i(t))\,\lambda_i(t)dt$. The number of patients in each sample will be denoted by n_i, with $n = n_0 + n_1$ and $\alpha_i = \lim_{n\to\infty} \frac{n_i}{n}$. Subscript n will be used to denote empirical versions of cumulative distribution functions, and we will also define $G_n = \frac{n_0}{n} G_{0n} + \frac{n_1}{n} G_{1n}$, $H_n = \frac{n_0}{n} H_{0n} + \frac{n_1}{n} H_{1n}$, $G = \alpha_0 G_0 + \alpha_1 G_1$, and $H = \alpha_0 H_0 + \alpha_1 H_1$.

For the combined sample of n patients let the ordered survival times, without regard to censoring, be denoted by

$$\underset{\sim}{X} = (X_1, \ldots, X_n) \quad .$$

Define a corresponding vector of sample indicators

$$\underset{\sim}{Z} = (Z_1, \ldots, Z_n) \quad ,$$

where
$$\begin{array}{ll} Z_j = 0 & X_j \text{ from sample } 0 \\ Z_j = 1 & X_j \text{ from sample } 1 \quad , \end{array}$$

and a vector of censoring indicators

$$\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_n) \quad ,$$

where
$$\begin{array}{ll} \delta_j = 0 & X_j \text{ censored} \\ \delta_j = 1 & X_j \text{ uncensored} \quad . \end{array}$$

Further, define the number at risk in the respective samples at X_j by

$$n_{0j} = \sum_{k=j}^{n} Z_k$$

and
$$n_{1j} = \sum_{k=j}^{n} (1 - Z_k) \quad .$$

3. Likelihood Approaches

3.1 Maximum Likelihood

With an exponential model, we have

$$F_0(t) = 1 - e^{-\lambda_0 t} \quad \text{and} \quad F_1(t) = 1 - e^{-\lambda_1 t} \quad .$$

The likelihood is in this case well known (Bartholomew (1957)) to be maximized by the ratio of occurrences to exposure time:

(1) $$\hat{\lambda}_0 = \sum_{j=1}^{n} \delta_j(1-Z_j) / \sum_{j=1}^{n} X_j(1-Z_j) = \int_0^\infty dG_{0n}(t) / \int_0^\infty tdH_{0n}(t)$$

and $$\hat{\lambda}_1 = \sum_{j=1}^{n} \delta_j Z_j / \sum_{j=1}^{n} X_j Z_j = \int_0^\infty dG_{1n}(t) / \int_0^\infty tdH_{1n}(t) \quad .$$

Thus, the maximum likelihood estimator of θ is $\hat{\lambda}_1/\hat{\lambda}_0$. From (1) it can be seen that, in general,

$$\hat{\lambda}_i \overset{P}{\to} \int_0^\infty dG_i(t) \;/\; \int_0^\infty tdH_i(t)$$

$$= \int_0^\infty \lambda_i(t)\ (1-H_i(t))dt \;/\; \int_0^\infty (1-H_i(t))dt \quad ,$$

so that $\hat{\lambda}_1/\hat{\lambda}_0 \overset{P}{\to} \theta$ under the exponential model, but not necessarily otherwise.

Asymptotic normality of $\hat{\lambda}_1/\hat{\lambda}_0$ can be shown in general using the representation (1) and follows from likelihood theory under the exponential model (see Gardiner (1982) for details.) In the latter case, the large sample variance of $\sqrt{n}\,(\hat{\lambda}_1/\hat{\lambda}_0 - \theta)$ is

(2) $$\frac{\theta^2}{\alpha_0\alpha_1}\left[\frac{\int_0^\infty [\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t))\,\theta\,]\ \lambda_0(t)\ dt}{\int_0^\infty (1-H_0(t))\ \lambda_0(t)dt \cdot \int_0^\infty (1-H_1(t))\,\theta\ \lambda_0(t)dt}\right] \quad .$$

3.2 Maximum Partial Likelihood

Cox (1972) presented a statistical procedure for inference from censored survival data which depended on the model of proportional hazards, and used a likelihood which did not depend on the form of $\lambda_0(t)$. This is not a likelihood in the standard sense but was later shown by Cox (1975) to be a partial likelihood.

In the two sample case, and using $\ell n\theta = \beta$, the Cox partial likelihood is

$$L(\beta) = \prod_{j=1}^{n} [\exp(\beta Z_j) / (n_{0j} + n_{1j} \exp(\beta))]^{\delta_j} ,$$

and ln (L(β)) =

$$\ell(\beta) = \sum_{j=1}^{n} \delta_j[\beta Z_j - \ln(n_{0j} + n_{1j} \exp(\beta))] , \tag{3}$$

The proposed estimator of θ is then $e^{\hat{\beta}}$, where $\hat{\beta}$ is the solution to

$$0 = \ell'(\beta) = \sum_{j=1}^{n} \delta_j \left[Z_j - \frac{n_{1j} \exp(\beta)}{n_{0j} + n_{1j} \exp(\beta)} \right]$$

$$= n_1 \int_0^\infty dG_{1n}(t) - n \int_0^\infty \frac{n_1(1-H_{1n}(t^-)) \exp(\beta)\, dG_n(t)}{n_0(1-H_{0n}(t^-)) + n_1(1-H_{1n}(t^-)) \exp(\beta)} . \tag{4}$$

McRae and Thomas (1972) show there is a solution to (4) corresponding to a maximum of (3) with arbitrarily high probability as $n \to \infty$.

The estimated variance is $-[e^{\hat{\beta}}]^2/\ell''(\hat{\beta})$, where

$$\ell''(\beta) = - \sum_{j=1}^{n} \frac{\delta_j n_{0j} n_{1j} \exp(\beta)}{[n_{0j} + n_{1j} \exp(\beta)]^2}$$

$$= - \frac{n_1 n_2}{n} \int_0^\infty \frac{(1-H_{0n}(t^-))(1-H_{1n}(t^-)) \exp(\beta)\, dG_n(t)}{\left[\frac{n_0}{n} (1-H_{0n}(t^-)) + \frac{n_1}{n} [1-H_{1n}(t^-)) \exp(\beta)\right]^2} , \tag{5}$$

Thus, $\sqrt{n}\,(e^{\hat{\beta}}-\theta)$ is taken to be asymptotically normal, with mean 0 and variance

$$\frac{\theta}{\alpha_0 \alpha_1} \left[\int_0^\infty \frac{[1-H_0(t)][1-H_1(t)]\, \theta\ \lambda_0(t)dt}{\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t))\, \theta} \right]^{-1} , \tag{6}$$

the denominator arising from the limit in probability of $-\frac{\ell''(\hat{\beta})}{n}$.

The log-likelihood $\ell(\beta)$ is composed of random variables which are neither independent nor identically distributed, so standard likelihood theory fails to justify the large sample moments and distribution given above. However, an approach is possible following the outline of the proofs for standard likelihoods, as suggested in Cox (1975). There are several articles covering the large sample theory for random covariates (Tsiatis, 1981, Andersen and Gill, 1981, for example), but none which explicitly cover the two-sample case arguing directly from the partial likelihood using classical methods. We sketch such a proof below.

Expanding $\ell'(\beta)$ around the true value β_0, we have

$$\ell'(\hat{\beta}) = 0 = \ell'(\beta_0) + (\hat{\beta} - \beta_0)\, \ell''(\beta^*) \quad ,$$

where β^* is in $(\beta_0, \hat{\beta})$. Thus

$$\sqrt{n}\,(\hat{\beta} - \beta_0) = \frac{-n^{-\frac{1}{2}}\,\ell'(\beta_0)}{\ell''(\beta^*)/n} \quad .$$

The result follows if $n^{-\frac{1}{2}}\,\ell'(\beta_0)$ is asymptotically normal with mean 0 and variance estimated by $-\frac{1}{n}\,\ell''(\beta_0)$, if $\hat{\beta} \xrightarrow{P} \beta_0$, and if $\frac{1}{n}\,(\ell''(\beta^*) - \ell''(\beta_0)) \xrightarrow{P} 0$.

From (4) we can see that $\ell'(\beta_0)$ is just the logrank test, with the addition of the constant $\exp(\beta_0)$ multiplying n_{2j}, and thus trivial extensions of existing proofs show that

$$\frac{n^{-\frac{1}{2}}\,\ell'(\beta_0)}{[\ell''(\beta_0)/n]^{\frac{1}{2}}} \xrightarrow{D} N(0,1) \quad ,$$

under proportional hazards and random censorship (Crowley and Thomas, 1975) or more general censoring (Aalen, 1978). Further, it is clear from (5) that $\frac{1}{n}\,[\ell''(\beta^*) - \ell''(\beta_0)]$ will converge in probability to 0 if $\hat{\beta} \xrightarrow{P} \beta_0$.

To show the consistency of $\hat{\beta}$, note that $\ell''(\beta) < 0$, so that $\ell'(\beta)$ is decreasing, and that $\frac{\ell'(\beta)}{n}$ can be seen from (4) to converge in probability to

$$\alpha_1 \int_0^\infty dG_1(t) - \int \frac{\alpha_1(1-H_1(t)) \exp(\beta)\, dG(t)}{\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t)) \exp(\beta)} \quad . \tag{7}$$

If the true parameter is β_0, and setting $\beta = \beta_0 + \Delta$, (7) decomes

$$\int_0^\infty \alpha_1(1-H_1(t)) \exp(\beta_0)\, \lambda_0(t)dt - \int \frac{\alpha_1(1-H_1(t)) \exp(\beta_0 + \Delta)(1-F_0(t))^{\exp(\Delta)}\, dG(t)}{\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t)) \exp(\beta_0 + \Delta)} \quad . \tag{8}$$

Writing $dG(t) = [\alpha_0(1-H_0(t) + \alpha_1(1-H_1(t)) \exp(\beta_0)]\, \lambda_0(t)dt$, (8) simplifies after a little algebra to

$$= \alpha_0\alpha_1 \int_0^\infty \frac{(1-H_0(t))(1-H_1(t)) \exp(\beta_0)(1-\exp(\Delta))\lambda_0(t)dt}{\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t)) \exp(\beta_0 + \Delta)} \quad . \tag{9}$$

The expression (9) is 0 when $\Delta = 0$, and negative (positive) when Δ is positive (negative); there is thus with high probability a single root of $\ell'(\beta) = 0$ in an arbitrary neighborhood of β_0, so that $\hat{\beta} \xrightarrow{P} \beta_0$.

This result can be generalized to vector-valued covariates, not necessarily i.i.d., with a finite number of possible outcomes. The approach requires establishing that

$$\lim_{n \to \infty} P\{(\ell(\beta_0) - \sup_{\beta \in N_\Delta(\beta_0)} \ell(\beta)) > 0\} = 1 \quad ,$$

for $N_\Delta(\beta_0) = \{\beta : |\beta - \beta_0| = \Delta\}$. This rests on careful consideration of the terms $U_j(\beta)$ of the log-partial likelihood, centered by their conditional expectations under $\beta = \beta_0$ given the information on the times of all previous censored and uncensored observations and the associated covariates, and on the fact that an uncensored observation occurs at X_j. Details are given in Liu and Crowley (1978).

4. Ad Hoc Methods

4.1 Standardized Mortality Ratio

Peto, et al. (1977) provide an excellent review of the analysis of clinical trials and point out the need for simple, closed-form test statistics and estimators. They suggest that by analogy with the standardized mortality ratio, an estimator of relative risk is provided by the ratio of observed deaths in sample 1 to that expected under the null hypothesis. This is

$$O/E = \frac{\sum_{j=1}^{n} \delta_j Z_j}{\sum_{j=1}^{n} \delta_j \frac{n_{1j}}{n-j+1}} \quad ;$$

$n-j+1$ being the total number at risk at X_j and $n_{1j}/n-j+1$ thus the estimated number of deaths from sample 1. This can be rewritten as

$$\frac{\int_0^\infty dG_{1n}(t)}{\int_0^\infty \frac{1-H_{1n}(t^-)\, dG_n(t)}{1-H_n(t^-)}} ,$$

from which it can be seen that

$$O/E \xrightarrow{P} \frac{\int_0^\infty dG_1(t)}{\int_0^\infty \frac{(1-H_1(t))\ dG(t)}{(1-H(t))}} .$$

Under proportional hazards this is

$$(10) \qquad \frac{\theta \int_0^\infty (1-H_1(t))\ \lambda_0(t)dt}{\int_0^\infty \frac{(1-H_1(t))\ [\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t))\,\theta\,]\ \lambda_0(t)dt}{\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t))}} .$$

The estimator O/E is thus biased towards 1. For example, with the exponential model and no censoring, (10) is equal to 1.24 when $\theta = 1.5$, and 1.44 when $\theta = 2$. However, Bernstein, Anderson, and Pike (1981) present some Monte Carlo results which indicate that O/E behaves fairly well for moderate sample sizes and moderate departures from the null hypothesis.

4.2 The Mantel-Haenszel Estimator

An analogy with the analysis of case-control studies could also be drawn, as was done for the logrank (or Mantel-Haenszel) test by Mantel (1966). This would suggest that the Mantel-Haenszel (1959) estimator of the log odds ratio (relative risk for rare diseases) from sets of 2×2 tables be used as an estimator of relative risk for survival studies as well. This is given by

$$\hat{\theta}_R = \frac{\sum_{j=1}^{n} \frac{\delta_j Z_j (n_{0j} - (1-Z_j))}{n-j+1}}{\sum_{j=1}^{n} \frac{\delta_j (1-Z_j)(n_{1j} - Z_j)}{n-j+1}}$$

$$= \frac{\displaystyle\int_0^\infty \frac{(1-H_{0n}(t))\, dG_{1n}(t)}{(1-H_n(t^-))}}{\displaystyle\int_0^\infty \frac{(1-H_{1n}(t))\, dG_{0n}(t)}{(1-H_n(t^-))}} .$$

In establishing the large sample properties of $\hat{\theta}_R$ it is convenient to use the results on counting processes, outlined, for example, by Andersen, Borgan, Gill and Keiding (1981). Thus, $n_i\, dG_{in}(t)$ are counting processes with compensators $\int_0^t \lambda_i(s)\, n_i(1-H_{in}(s^-))ds$, and

$$M_i(t) = n_i\, G_{in}(t) - \int_0^t \lambda_i(s)\, n_i(1-H_{in}(s^-))ds \quad ,$$

are orthogonal square-integrable martingales. Also, as $\sqrt{n}\,(1-H_{1n}(t^-))/n_0(1-H_n(t^-))$ and $\sqrt{n}\,(1-H_{0n}(t^-))/n_1(1-H_n(t^-))$ are left-continuous and therefore predictable processes, and are bounded (interpreting 0/0 as 0), we have that

$$\sqrt{n} \int_0^\infty \frac{(1-H_{1n}(t^-))}{n_0(1-H_n(t^-))}\, dM_0(t)$$

and

$$\sqrt{n} \int_0^\infty \frac{(1-H_{0n}(t^-))}{n_1(1-H_n(t^-))}\, dM_1(t) \quad ,$$

are orthogonal square-integrable martingales. Further, the conditions for asymptotic normality given by Andersen, Borgan, Gill and Keiding (1981) are satisfied, and we have that

$$\sqrt{n} \left[\int_0^\infty \frac{(1-H_{1n}(t^-))}{(1-H_n(t^-))}\, dG_{0n}(t) - \int_0^\infty \frac{(1-H_{1n}(t^-))}{(1-H_n(t^-))}\, \lambda_0(t)(1-H_{0n}(t^-))dt \right] ,$$

and

$$\sqrt{n}\left[\int_0^\infty \frac{(1-H_{0n}(t^-))}{(1-H_n(t^-))}\,dG_{1n}(t) - \int_0^\infty \frac{(1-H_{0n}(t^-))}{(1-H_n(t^-))}\,\lambda_1(t)(1-H_{1n}(t^-))dt\right] ,$$

are in the limit, mean 0, independent normal random variables with variances

$$\sigma_0^2 = \frac{1}{\alpha_0}\int_0^\infty \frac{(1-H_0(t))(1-H_1(t))^2}{(1-H(t))^2}\,\lambda_0(t)dt$$

$$\sigma_1^2 = \frac{1}{\alpha_1}\int_0^\infty \frac{(1-H_0(t))^2(1-H_1(t))}{(1-H(t))^2}\,\lambda_1(t)dt \quad .$$

Thus, we have a result of the form

$$\sqrt{n}\,(A_{0n} - B_{0n}, A_{1n} - B_{1n}) \xrightarrow{D} N\left(0, \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}\right) ,$$

and our estimator $\hat{\theta}_R$ can be seen to be $\frac{A_{1n}}{A_{0n}}$ (except for the presence of t^- instead of t in the respective numerators, which will not matter in the limit because of the assumed continuity of $F_i(t)$). Asymptotic normality of $\hat{\theta}_R$ follows from repeated application of the δ-method, expanding A_{0n} and B_{0n} around their common limit

$$\mu_0 = \int_0^\infty \frac{(1-H_1(t))}{(1-H(t))}\,dG_0(t) \quad ,$$

and A_{1n} and B_{1n} around

$$\mu_1 = \int_0^\infty \frac{(1-H_0(t))}{(1-H(t))}\,dG_1(t) \quad .$$

This gives

$$\sqrt{n}\left(\frac{A_{1n}}{A_{0n}} - \frac{B_{1n}}{B_{0n}}\right) \xrightarrow{D} N\left(0, \left(\frac{\mu_1}{\mu_0}\right)^2 \cdot \left(\frac{\sigma_0^2}{\mu_0^2} + \frac{\sigma_1^2}{\mu_1^2}\right)\right).$$

For proportional hazards, $\frac{B_{1n}}{B_{0n}} = \frac{\mu_1}{\mu_0} = \theta$, and we have established that $\hat{\theta}_R \xrightarrow{P} \theta$, and that $\sqrt{n}\,(\hat{\theta}_R - \theta)$ is asymptotically normal, with variance which reduces to

$$\frac{\theta^2}{\alpha_0\alpha_1}\left\{\frac{\int_0^\infty \frac{(1-H_0(t))(1-H_1(t))}{(1-H(t))^2}\,[\alpha_0(1-H_0(t)) + \alpha_1(1-H_1(t))\theta]\,\lambda_0(t)dt}{\theta\left[\int_0^\infty \frac{(1-H_0(t))(1-H_1(t))}{1-H(t)}\,\lambda_0(t)dt\right]^2}\right\} \tag{11}$$

The estimator $\hat{\theta}_R$ has been generalized to statistics of the form

$$\frac{\int_0^\infty J(H_{0n}(t),\, H_{1n}(t))\,\frac{dG_{1n}(t)}{1-H_{1n}(t^-)}}{\int_0^\infty J(H_{0n}(t),\, H_{1n}(t))\,\frac{dG_{0n}(t)}{1-H_{0n}(t^-)}}, \tag{12}$$

which will also be consistent for θ under the proportional hazards model and will be asymptotically normal for a certain class of functions J. The choice of $J = \frac{(1-H_{0n}(t^-))(1-H_{1n}(t^-))}{1-H_n(t^-)}$ corresponds to $\hat{\theta}_R$. This is discussed from the point of view of the resulting test statistics by Andersen (1981), and from the point of view of estimation by Begun and Reid (1981). That some restrictions on J are necessary can be seen from the choice of $J \equiv 1$, resulting in the ratio of cumulative hazards

$$\frac{\displaystyle\int_0^\infty \frac{dG_{1n}(t)}{1-H_{1n}(t^-)}}{\displaystyle\int_0^\infty \frac{dG_{0n}(t)}{1-H_{0n}(t^-)}} = \frac{\displaystyle\sum_{j=1}^n \frac{\delta_j z_j}{n_{1j}}}{\displaystyle\sum_{j=1}^n \frac{\delta_j(1-z_j)}{n_{0j}}} = \frac{\hat{\Lambda}_1(\infty)}{\hat{\Lambda}_0(\infty)} \quad ,$$

where $\hat{\Lambda}_i(t)$ is sometimes referred to as Nelson's estimator (Nelson, 1969) of the cumulative or integrated hazard $\int_0^t \lambda_i(s)ds$. The ratio may well provide a useful estimate as a function of time, for

$$\frac{\hat{\Lambda}_1(t)}{\hat{\Lambda}_0(t)} = \frac{\hat{\Lambda}_1(t)/t}{\hat{\Lambda}_0(t)/t} \quad ,$$

has the interpretation of the ratio of time-averaged hazards, but choice of t will be important in its use as an estimator, as the asymptotic variance of $\hat{\Lambda}(t)$ can increase without bound (cf Breslow and Crowley, 1974). A similar point was made by Kalbfleisch and Prentice (1981), who study estimates of the average hazard ration, defined to by

$$\int_0^t \lambda_1(s)/\lambda_0(s)\ dW(s) \quad ,$$

for suitably chosen weight functions W.

5. Some Numerical Comparisons

As a large sample measure of efficiency we can compare asymptotic variances for those estimators which are consistent. With the exponential model this includes the maximum likelihood estimator $\hat{\lambda}_1/\hat{\lambda}_0$, the maximum partial likelihood estimator $e^{\hat{\beta}}$, and the Mantel-Haenszel estimator $\hat{\theta}_R$. Since it is maximum likelihood, we have that A.Var. $\sqrt{n}\ (\hat{\lambda}_1/\hat{\lambda}_0 - \theta) \leq$ A.Var. $\sqrt{n}\ (e^{\hat{\beta}} - \theta)$, where A.Var. stands for asymptotic variance. For the general proportional hazards model we can see from (6) and (11) and the Cauchy-Schwarz inequality that A.Var.

$\sqrt{n}\,(e^{\hat{\beta}} - \theta) \leq$ A.Var. $\sqrt{n}\,(\hat{\theta}_R - \theta)$, with equality with $\theta=1$. With equal exponential survival distributions and equal censoring, comparison of (2) with (6) and (11) reveals that all three estimators have the same asymptotic variance.

Table 1 gives large sample variances for the three estimators under an exponential model with $\lambda_0 = 1$, covering the cases $\theta = 1, 1.5$, and 2.0 for four different censoring patterns. Conditions 1 and 2 correspond to a cohort entering the study at time 0, with staggered entry of other subjects from 0 to 1 and analysis at time 1; case 1 having an equal size cohort in each sample and case 2, unequal. The third censoring condition is equal, type I censoring at time 1, and the fourth is no censoring. Also given in Table 1 are comparisons of $e^{\hat{\beta}}$ and $\hat{\theta}_R$ for Weibull distributions under the same censoring conditions. The most remarkable feature of the table is the high relative efficiency of the simple, closed-form Mantel-Haenszel estimator. This small loss can be regained entirely in a two-step procedure suggested by Begun and Reid (1981), using the statistic (12). They show that the optimal J depends on θ, but that use of this J with θ replaced by $\hat{\theta}_R$ (or any consistent estimator) provides full efficiency relative to the maximum partial likelihood estimator.

Further numerical comparisons covering the case of grouped survival times are given in Crowley (1975).

TABLE 1. Large Sample Variances

I. $1-F_0(t) = e^{-t}$ $\quad 1-F_1(t) = e^{-\theta t}$

1) $C_i(t) = t/1.25$, $t \varepsilon [0,1];1,\ t > 1$

θ	$\hat{\lambda}_1/\hat{\lambda}_0$	$e^{\hat{\beta}}$	$\hat{\theta}_R$
1	9.51	9.51	9.51
1.5	19.01	19.08	19.08
2	31.77	32.20	32.25

2) $C_0(t) = t/1.25$ $\quad C_1(t) = t/1.5$, $t \varepsilon [0,1];1,\ t > 1$

θ	$\hat{\lambda}_1/\hat{\lambda}_0$	$e^{\hat{\beta}}$	$\hat{\theta}_R$
1	9.14	9.16	9.16
1.5	18.45	18.46	18.46
2	31.02	31.21	31.24

3) $C_0(t) = C_1(t) = 0$, $t \varepsilon [0,1];1,\ t > 1$

θ	$\hat{\lambda}_1/\hat{\lambda}_2$	$e^{\hat{\beta}}$	$\hat{\theta}_R$
1	6.33	6.33	6.33
1.5	12.91	12.97	12.98
2	21.91	22.31	22.36

4) No censoring

θ	$\hat{\lambda}_1/\hat{\lambda}_2$	$e^{\hat{\beta}}$	$\hat{\theta}_R$
1	4	4	4
1.5	9	9.35	9.36
2	16	17.75	17.87

Table 1 (continued)

II. $1-F_0(t) = e^{-t^2}$ $\quad 1-F_1(t) = e^{-\theta t^2}$

1) $C_i(t) = t/1.25 \quad , \quad t\varepsilon[0,1];1, \ t > 1$

θ	$e^{\hat{\beta}}$	$\hat{\theta}_R$
1	12.16	12.16
1.5	24.37	24.37
2	41.07	41.14

2) $C_0(t) = t/1.25 \quad , \quad C_1(t) = t/1.5 \quad , \quad t\varepsilon[0,1];1, \ t > 1$

θ	$e^{\hat{\beta}}$	$\hat{\theta}_R$
1	11.38	11.38
1.5	22.99	22.99
2	38.90	38.92

Results for cases 3) and 4) are the same as for corresponding entries in Part I.

REFERENCES

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. Annals of Statistics **6**, 374-391.

Andersen, P.K. (1981). Comparing survivals via hazard ratio estimates. Research Report 81/7, Statistical Research Unit, Danish Medical and Social Science Research Councils. To appear in the Scandinavian Journal of Statistics, 1983.

Anderson, P.K. and Gill, R. (1981). Cox's regression model for counting processes: a large sample study. Research Report 81/6, Statistical Research Unit, Danish Medical and Social Science Research Councils. To appear in the Annals of Statistics, 1982.

Anderson, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1981). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. Research Report 81/4, Statistical Research Unit, Danish Medical and Social Science Research Councils. To appear in the International Statistical Review, 1982.

Bartholomew, D.J. (1957). A problem in life testing. Journal of the American Statistical Association **52**, 350-355.

Begun, J.M. and Reid, N. (1981). Estimating the relative risk with censored data. Technical Report No. 81/16, The Institute of Applied Mathematics and Statistics, the University of British Columbia. To appear in the Journal of the American Statistical Association, 1983.

Bernstein, L., Anderson, J., and Poke, M.C. (1981). Estimation of the proportional hazard in two-treatment-group clinical trials. Biometrics **37**, 513-519.

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.

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

Cox, D.R. (1975). Partial likelihood. Biometrika **62**, 269-276.

Crowley, J. and Thomas, D.R. (1975). Large sample theory for the log rank test. Technical Report No. 415, University of Wisconsin-Madison, Department of Statistics.

Crowley, J. (1975). Estimation of relative risk in survival studies. Technical Report No. 423, University of Wisconsin-Madison, Department of Statistics.

Gardiner, J. (1982). The asymptotic distribution of mortality rates in competing risks analysis. Scandinavian Journal of Statistics **9**, 31-36.

Kalbfleisch, J.D. and Prentice, R.L. (1981). Estimation of the average hazard ratio. Biometrika **68**, 105-112.

Liu, P.Y. and Crowley, J. (1978). Large sample theory of the MLE based on Cox's regression and life model for survival data. Technical Report No. 1, Wisconsin Clinical Cancer Center - Biostatistics.

Mantel, N., and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute **22**, 719-748.

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

McRae, K.B. and Thomas, D.R. (1972). Inference for pairs of distributions with proportional failure rate functions. Paper presented at the Joint Statistical Meetings in Montreal.

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

Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J., and Smith, P.G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. British Journal of Cancer **35**, 1-39.

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