

Likelihood Ratio Tests for a Change Point with Survival Data

By XIAOLONG LUO

*Department of Biostatistics, St. Jude Children's Research Hospital, P.O. Box 318,
Memphis, Tennessee 38101, U.S.A.*

BRUCE W. TURNBULL

*School of Operations Research and Industrial Engineering, Cornell University,
Ithaca, New York 14853, U.S.A.*

and LARRY C. CLARK

University of Arizona Cancer Center, Tucson, Arizona 85724, U.S.A.

SUMMARY

We consider the likelihood ratio test for a change point or lag of the effect of some covariates (e.g. treatment) in regression modeling of survival data subject to right censoring. Asymptotic distributions under null hypotheses are derived. Numerical examples are used to illustrate the implementation of the procedure. Finally we apply the theory to data from a large randomized cancer prevention trial.

Some key words: Asymptotic theory; Change point inference; Clinical trial; Cox regression model; Kaplan-Meier estimator; Lag effect; Partial likelihood; Proportional hazard; Survival analysis; Time-dependent covariate.

1. INTRODUCTION

In life-testing problems, it is often reasonable to assume that any effect of a treatment is not immediate but affects the risk of failure only after a time lag. This commonly occurs in the analysis of clinical trial data for disease prevention or disease treatment. Zucker & Lakatos (1990) described several such situations. For example, they describe how, in the Lipid Research Clinics Coronary Primary Prevention Trial, it was expected that cholesterol lowering therapy would only gradually reduce the risk of coronary events, since the therapy is initiated only after many years of high cholesterol and associated plaque development. Zucker & Lakatos (1990) described the planned analysis of the Physician's Health Study for testing the effect of beta-carotene on cancer incidence. Since the investigators thought that treatment would not affect pre-existing tumors, and because of the time required for new tumors to become detectable, it was decided not to include cancers detected in the first two years following randomization. This amounts to a "threshold" lag time effect. The same two-year "waiting period" was proposed by Clark *et al.* (1991, p.522) with respect to the proposed analysis of a placebo controlled randomized trial to explore the benefits of a nutritional supplement of selenium for the prevention of cancer. The delay in observable treatment effect may be because it takes several months of daily

ingestion of selenium tablets for plasma levels to rise; another explanation may be that selenium affects only initiation not promotion of tumors, so tumors initiated but not detected before the start of treatment might not be affected. The impact of a lag time in the design of clinical trials has been described by Meinert (1986, Sec. 9.3.10).

Zucker & Lakatos (1990) studied two-sample weighted logrank statistics for comparing survival curves that are robust against a class of unknown treatment lag effects. Matthews & Farewell (1982) and Matthews, Farewell & Pyke (1985) considered the problem of testing for a null hypothesis of a constant hazard against a changepoint alternative. Liang, Self, & Liu (1990) present a variant of the Cox (1972) proportional hazards model which includes a change point. They present a test of the null hypothesis of no changepoint effect based on a maximal score statistic. Simon & Altman (1994) and George (1994) discuss related issues. For estimation of lag effects, Luo, Turnbull, Cai, & Clark (1994) and Luo (1996) have obtained the asymptotic distributions of maximum partial likelihood estimators of the change point and regression coefficients.

In this paper, we consider a modification of the proportional hazards failure time regression model, also used by Liang *et al.* (1990), which includes a threshold lag effect for some of the covariates. The likelihood ratio test is proposed for testing the hypothesis that the changepoint is equal to some specified value. Asymptotic properties of the test statistic are derived. Examples and an application are given. For simplicity of exposition, we start with the two-sample problem where there is a single binary covariate, assigned treatment say. The more general case is described in Section 5. Proofs appear in the Appendix.

2. THE TWO-SAMPLE PROBLEM

Suppose there are n independent subjects or individuals, each with an associated p dimensional covariate column vector $X_i = (x_{i1}, \dots, x_{ip})'$ ($1 \leq i \leq n$). Let T_i , $1 \leq i \leq n$, be the survival time of the i -th individual and C_i be the censoring time for T_i . We assume that the n triplets $\{(T_i, C_i, X_i), 1 \leq i \leq n\}$ are independently identically distributed; also that the $\{T_i\}$ and $\{C_i\}$ are conditionally independent of each other, given the $\{X_i\}$. We observe pairs $(Y_1, \delta_1), \dots, (Y_n, \delta_n)$ where $Y_i = T_i \wedge C_i$ and $\delta_i = 1_{(T_i \leq C_i)}$. The data consist of n triplets $\{(Y_i, \delta_i, X_i), 1 \leq i \leq n\}$. In the two sample problem, $p = 1$ and the covariate X_i is binary, taking on the value either 0 or 1. We write $\lambda_x(t)$ as the hazard function for a subject with covariate $X = x$ and $\lambda_0(t)$ as the hazard function for baseline case when $X = 0$. The lag model has the form

$$\lambda_x(t) = \{(1 - 1_{(t > \theta)}) + 1_{(t > \theta)} e^{\beta x}\} \lambda_0(t). \quad (1)$$

Here $\lambda_0(t)$ is an arbitrary unknown baseline hazard rate function, β is an unknown regression coefficient and θ denotes the threshold or lag parameter, also unknown, which is the quantity of primary interest in this paper. We will assume throughout

that $\beta \neq 0$, since otherwise θ is not identifiable and the problem not meaningful: if treatment has no effect, it makes no sense to discuss a lag.

We will assume that there are no tied uncensored failure times. If there are ties, Peto's modification can be used (Miller, 1981, p.137); we assume no ties to avoid theoretical complications. The partial likelihood function for β and θ is given by

$$\prod_{i:\delta_i=1} \frac{\lambda_{x_i}(Y_i)}{\sum_{j \in \mathcal{R}_i} \lambda_{x_j}(Y_i)} = \prod_{i:\delta_i=1} \frac{(1 - 1_{(Y_i > \theta)}) + 1_{(Y_i > \theta)} e^{\beta X_i}}{\sum_{j \in \mathcal{R}_i} [(1 - 1_{(Y_i > \theta)}) + 1_{(Y_i > \theta)} e^{\beta X_j}]} .$$

Here, as usual, $\mathcal{R}_i = \{j : Y_j \geq Y_i\}$ denotes the risk set at time Y_i and $j \in \mathcal{R}_i$ means that $Y_j \geq Y_i$. The log partial likelihood is given by

$$L(\beta, \theta) = \sum_{i:\delta_i=1} \left[\log \left\{ (1 - 1_{(Y_i > \theta)}) + 1_{(Y_i > \theta)} e^{\beta X_i} \right\} - \log \sum_{j \in \mathcal{R}_i} \left\{ (1 - 1_{(Y_i > \theta)}) + 1_{(Y_i > \theta)} e^{\beta X_j} \right\} \right] . \quad (2)$$

For some specified value of interest of the lag parameter, θ_0 , say, we wish to test the null hypothesis $H_0 : \theta = \theta_0$ versus the alternative $H_1 : \theta \neq \theta_0$. We propose to use the likelihood ratio test statistic

$$T = \sup_{\theta, \beta} L(\beta, \theta) - \sup_{\beta} L(\beta, \theta_0) \quad (3)$$

The computation of T is not difficult. First, denote the maximum partial likelihood estimators $(\hat{\beta}, \hat{\theta})$ as the maximizing values of $L(\beta, \theta)$. Also, for each fixed θ , we denote $\tilde{\beta}(\theta)$ as the value of β that maximizes $L(\beta, \theta)$, and $\tilde{\beta} = \tilde{\beta}(\theta_0)$. Now $\tilde{\beta}(\theta)$ can be obtained by noting that, for fixed θ , the model (1) can be written as a Cox regression model with a time dependent covariate $\tilde{X}(t) = X 1_{(t > \theta)}$. Hence standard software can be used to compute $\tilde{\beta}(\theta)$. We used PROC PHREG in SAS Rel. 6.10. With a grid of θ -values, a search technique can be used to find $\hat{\theta}$ and $\hat{\beta} = \tilde{\beta}(\hat{\theta})$. Thus the test statistic T given by (3) can be computed.

We derive the asymptotic distribution of T under $H_0 : \theta = \theta_0$ as the sample size $n \rightarrow \infty$. We need to define some further random quantities. First suppose $\{U_j, j = 1, 2, 3, \dots\}$ and $\{V_j, j = 1, 2, 3, \dots\}$ are independently identically distributed exponential random variables with unit mean. We define $\{Z_j^-, j = 1, 2, 3, \dots\}$ to be independently identically distributed Bernoulli random variables with distribution

$$\text{pr}(Z_j^- = i) = \text{pr}(Y \geq \theta_0, X = i) / \text{pr}(Y \geq \theta_0) , \quad (4)$$

for $i = 0, 1$, where (Y, X) has the same distribution as the pairs $\{(Y_i, X_i), 1 \leq i \leq n\}$. Similarly, we define $\{Z_j^+, j = 1, 2, 3, \dots\}$ to be independently identically distributed Bernoulli random variables with distribution

$$\text{pr}(Z_j^+ = 0) = 1 - \text{pr}(Z_j^+ = 1) = \frac{\text{pr}(Y \geq \theta_0, X = 0)}{\text{pr}(Y \geq \theta_0, X = 0) + e^\beta \text{pr}(Y \geq \theta_0, X = 1)}. \quad (5)$$

Furthermore, all sequences are independent of each other. Define

$$\begin{aligned} J_j &= \beta Z_j^- - \log \frac{E e^{\beta X} 1_{(Y > \theta_0)}}{E 1_{(Y > \theta_0)}} U_j \\ A_m &= \sum_{j=1}^m J_j, \quad A_0 = 0 \\ K_j &= -\beta Z_j^+ - \log \frac{E 1_{(Y > \theta_0)}}{E e^{\beta X} 1_{(Y > \theta_0)}} V_j \\ B_m &= \sum_{j=1}^m K_j, \quad B_0 = 0 \end{aligned}$$

Thus $\{A_m : m \geq 0\}$ and $\{B_m : m \geq 0\}$ are two independent random walks. Note that, since we have assumed $\beta \neq 0$, Jensen's inequality leads to

$$\begin{aligned} E J_j &= \beta \frac{E\{X 1_{(Y > \theta_0)}\}}{E\{1_{(Y > \theta_0)}\}} - \log \frac{E\{e^{\beta X} 1_{(Y > \theta_0)}\}}{E\{1_{(Y > \theta_0)}\}} < 0 \\ E K_j &= -\beta \frac{E\{X e^{\beta X} 1_{(Y > \theta_0)}\}}{E\{e^{\beta X} 1_{(Y > \theta_0)}\}} - \log \frac{E\{1_{(Y > \theta_0)}\}}{E\{e^{\beta X} 1_{(Y > \theta_0)}\}} < 0 \end{aligned}$$

and thus $\lim_{m \rightarrow \infty} A_m = -\infty$ and $\lim_{m \rightarrow \infty} B_m = -\infty$ almost surely (Feller, 1971, Chap. XII). Finally define random variables A, B , by

$$A = \sup_{m \geq 0} (A_m), \quad B = \sup_{m \geq 0} (B_m) \quad (6)$$

It is shown in the Appendix that the null distribution of the test statistic T in (3) is given asymptotically as $n \rightarrow \infty$ by

$$T \approx \begin{cases} \max\{A, B\} & \text{if } \theta_0 > 0 \\ B & \text{if } \theta_0 = 0. \end{cases} \quad (7)$$

To implement the test procedure, p -values for testing null hypotheses should be calculated from the exact distribution of (7) through Wiener-Hopf integral equations (Hinkley, 1970, Atkinson 1974). However, since the distributions of J_j 's or K_j 's are unknown, the exact Wiener-Hopf integral equations cannot be specified. Instead we can use the Monte Carlo method to obtain p -values.

We proceed as follows. We need to obtain some number, N , say, of simulated values of T under H_0 . In our examples we used $N = 500$, which choice a sensitivity check showed gave stable results. The observed value of T can then be compared with the simulated ones to obtain the p -value. For the case $\theta_0 > 0$, each simulated T value depends on the maximum of the two random walks $\{A_m, B_m; m \geq 0\}$. Suppose n_A and n_B are the number of uncensored Y_i which are less than and greater than

θ_0 , respectively. In practice, when we simulate the random walks, we truncate at n_A and n_B steps for $\{A_m, m \geq 0\}$ and $\{B_m; m \geq 0\}$ respectively and use $\{A_m, 0 \leq m \leq n_A; B_m, 0 \leq m \leq n_B\}$; see Luo (1996, Section 6). Since β is unknown, we replace it by $\tilde{\beta} = \tilde{\beta}(\theta_0)$, computed as described above; an asymptotic justification is given in the Appendix. The step sizes $\{J_j\}, \{K_j\}$ in the random walks $\{A_m\}, \{B_m\}$ also depend on the unknown distributions (4, 5) of $\{Z_j^-\}$ and $\{Z_j^+\}$. However, we can replace these by their empirical counterparts based on the observed data. For example, the probability $\text{pr}(Z_j^- = 0)$ in (4) is replaced by the sample proportion of individuals in Sample 0 that have failure or censoring times $\{Y_i\}$ no less than θ_0 . We can now simulate N pairs of random walks, $\{A_m\}, \{B_m\}$, and hence N realizations of $\max\{A, B\}$. The p -value is then computed as the proportion of these realizations that exceed the observed value of T . Of course, in the case $\theta_0 = 0$ when we are testing the null hypothesis of no lag, we need only consider the value of B and the single random walk $\{B_m\}$.

We now give two illustrations – a simulated example and an application from a clinical trial.

3. SIMULATION EXAMPLE

The dataset was generated using sample sizes and parameter values that might be typical of a large disease prevention trial such as the one described in Section 1 and which will be analyzed further in the next section. The dataset has the following features. There is a single covariate X which has a Bernoulli distribution $B(1, \frac{1}{2})$. The regression coefficient $\beta = \ln(0.5) = -0.693$. This represents the case when a treatment indicator X is such that treatment $X = 1$ represents an eventual halving of the hazard rate after the lag period when compared to the control group ($X = 0$). A subject is equally likely to be assigned to either group. The baseline hazard rate $\lambda_0(t) = 0.1$ is assumed constant and the lag value θ set equal to 3. Thus, but for the lag effect, failure times would all be exponentially distributed. Note that, at $\theta_0 = 3$, the cumulative mortality for the control group ($X = 0$) is $1 - \exp(-.3) = 26\%$. A random sample of $n = 1000$ failure times was generated using a hazard rate of the form (1). The censoring times were chosen to be independent and uniformly distributed over $[2,6]$. This mimics a clinical trial with a 6 year horizon which starts with a 4 year patient recruitment period during which accrual occurs at a uniform rate and there is only end-of-study censoring.

The simulated dataset generated 505 observations from the control group, of which 337 (168) were censored (failed), and 495 from the treatment group with 346 (149) censored (failed). The Kaplan-Meier curves (not given here), when overlaid, clearly show the three-year-lag feature. The maximum likelihood estimates of θ and β were $\hat{\theta} = 2.868$ and $\hat{\beta} = -0.733$. This latter value is within one standard error (0.26) of the true value (-0.693); see Luo *et al.* (1994) for details on point estimation.

With the simulated dataset, we apply the procedure described in Section 2 to test

various null hypotheses concerning θ . The results are summarized in Table 1.

Table 1. *Results of testing hypothesis $H_0 : \theta = \theta_0$ for selected θ_0 using the simulated dataset.*

Lag value θ_0	p -value	Estimated regression coefficient $\tilde{\beta} = \tilde{\beta}(\theta_0)$ (with standard error)
0.0	0.000	-0.103 (0.123)
1.0	0.000	-0.117 (0.133)
2.0	0.248	-0.321 (0.174)
3.0	0.538	-0.626 (0.267)
4.0	0.178	-0.662 (0.422)
5.0	0.032	-0.218 (0.817)

Thus the hypotheses $\theta_0 = 0.0, 1.0, 5.0$ would all be rejected at a 5% significance level, whereas the true value $\theta_0 = 3$ would not be. In fact the p -value for each value of θ_0 on a grid from 0 to 5 in multiples of 0.05 was computed. The results (shown as a continuous plot) are displayed in Fig. 1. Note that the plot appears quite rough. This is a necessary consequence of the discrete nature of the problem and follows from the fact that both the observed value of the test statistic T and its null distribution depend on the hypothesized value θ_0 . In fact, it can be shown that the plot behaves like a smooth function plus a stochastic integral with respect to Brownian motion, which is nowhere differentiable: this is analogous to a result of Miller and Siegmund (1982, Eqn (7)). The p -value plot can serve as a useful complement to a plot of the profile likelihood function, given by $L(\theta) = \sup_{\beta} L(\beta, \theta)$, as an indicator of θ -values that can be supported by the data.

[Figure 1 about here]

A 95% (say) confidence set for θ could be obtained by inverting the hypothesis test and including all θ values for which the p -values exceed 0.05. Fig. 1 suggests this may not be an interval, but a conservative interval could be formed by taking convex hull of the set.

Of course, Table 1 and Fig. 1 represent just one simulated dataset. In fact we simulated 100 such datasets. This simulation experiment, involving repeated use of a Cox regression procedure with time varying covariates on a large generated dataset and replicated 100 times, was extremely computer intensive. It was necessary to carry this out on the Cornell SP2 supercomputer. Table 2 below shows proportions of these datasets which reject the null hypothesis $\theta_0 = 0, 1, 2, 3, 4, 5$ at significance levels 0.05, 0.1, and 0.2.

Table 2. *Percentages of 100 simulated datasets which reject hypothesis $H_0 : \theta = \theta_0$.*

Lag value θ_0	$p\text{-value} \leq 0.05$	$p\text{-value} \leq 0.1$	$p\text{-value} \leq 0.2$
0.0	71%	76%	86%
1.0	61%	69%	79%
2.0	36%	44%	59%
3.0	2%	7%	14%
4.0	19%	34%	57%
5.0	46%	65%	79%

For $\theta_0 = 3$, since this was the value used to generate the data, those proportions can be seen as estimates of actual significance levels. In this particular finite sample situation, the likelihood ratio test appears to be conservative, yielding significance levels below the nominal values. For θ_0 -values other than 3.0, the entries in Table 2 give estimates of the power of the procedure. Of course, a conservative procedure will adversely affect the power.

4. CANCER PREVENTION TRIAL DATA

We now apply the procedure to data from the Nutritional Prevention of Cancer (NPC) trial. Clark *et al.* (1996) have described the trial and reported its results. Aspects of some earlier interim analyses have been discussed by Abu-Libdeh *et al.* (1990), Clark *et al.* (1991), and by Luo *et al.* (1994). Starting in 1983, 1312 patients entering this double blind study were randomized either to a daily nutritional supplement of selenium (Se) or to placebo. Selenium, an anti-oxidant, was thought to have cancer prevention potential. Full details of the background are given in Clark *et al.* (1996).

To illustrate the statistical methods proposed here, we consider just two of the many outcomes of interest reported in the study.

- (a) Total cancer incidence (194 cases). The responses $\{T_i\}$ are times to first diagnosis of an internal cancer. This category includes mostly lung cancer (48 cases), prostate cancer (48 cases) and colorectal cancer (27 cases), but also some small number of other cancers, *e.g.* bladder, breast, brain, leukaemia. The more numerous but less serious squamous cell carcinomas (SCC) and basal cell carcinomas (BCC) of the skin (408 and 726 cases, respectively) are excluded from this category.
- (b) SCC incidence (408 cases). The responses $\{T_i\}$ are times to first new diagnosis of squamous cell carcinoma (SCC) of the skin.

All response times were measured in months from entry to the study. The frequencies are summarized in Table 3.

Table 3. *Summary data for NPC Trial*

Group	Placebo (X=0)	Se (X=1)
Patients entered	659	653
Mean followup (months at risk)	74.4	76.8
(a) Cancer incidence	118	76
(b) SCC incidence	190	218

The trial investigators originally thought that the effect of nutritional supplementation on tumour occurrence rates might not be immediate but would be subject to a lag for reasons described in §1. Lag periods of the order of 12 or 24 months were suggested as being biologically plausible. The model (1) with a single covariate X was fitted where $X = 1$ indicated the Se group and $X = 0$ placebo. The maximum partial likelihood estimators were as follows

- (a) Cancer incidence: $\hat{\theta} = 0.1$ months, $\hat{\beta} = -0.480$ (s.e. = 0.148)
- (b) SCC incidence: $\hat{\theta} = 5.3$ months, $\hat{\beta} = 0.174$ (s.e. = 0.108).

The $\hat{\beta}$ values do suggest a significant protective effect of the Se intervention for the total cancer incidence endpoint, and a non-significant negative benefit for the SCC endpoint. These findings agree with the results of the standard logrank tests used in Clark *et al.* (1996) which do not take lags into account. There the two-sided p -values were $p < 0.001$ for (a), $p = 0.14$ for (b).

Turning our attention to the lag parameter θ , Table 4 gives results for the likelihood ratio test for θ_0 -values of 0, 6, 12 and 24 months. Alternatively, p -value plots, similar to Fig. 1 could be constructed but these are not shown here. For the cancer incidence endpoint, there does not appear to be sufficient evidence to rule out any of the lag values tested, including zero. For SCC incidence, note that, although the maximum partial likelihood estimator and logrank test indicate a nonsignificant treatment effect, it is still of interest to investigate if there are any hypothesized lag values for which the treatment is significant. However Table 4 shows that this is not the case here for SCC. Furthermore, if there were any true treatment effect on SCC, there is strong evidence against any lag greater than six months.

Table 4. *Results of testing hypothesis $H_0 : \theta = \theta_0$ for selected θ_0 using the NPC Trial data.*

Lag value θ_0 (months)	Cancer incidence		SCC incidence	
	p -value	$\tilde{\beta}(\theta_0)$ (s.e.)	p -value	$\tilde{\beta}(\theta_0)$ (s.e.)
0	0.782	-0.466 (0.147)	0.792	0.143 (0.099)
6	0.698	-0.456 (0.150)	0.842	0.150 (0.109)
12	0.794	-0.482 (0.159)	0.082	0.064 (0.121)
24	0.428	-0.441 (0.172)	0.000	-0.057 (0.141)

Although these results concerning any lag may not have been what the investigators had been looking for, this illustration does demonstrate the utility of the methodology proposed here. It was able to quantify the strength or weakness of the evidence concerning lag effects extant in the data. It also permits a check of the robustness of the significance of the treatment effect to the presence of a lag.

5. A GENERAL LAGGED REGRESSION PROBLEM

The two-sample problem can be extended to a more general regression setting where there are several covariates, some of which are moderated by a lag effect. For each of the n independent subjects there is an associated $(p+q)$ -dimensional covariate column vector $X_i = (X_i^{(1)}, X_i^{(2)}) = (X_{i1}^{(1)}, \dots, X_{iq}^{(1)}, X_{i1}^{(2)}, \dots, X_{ip}^{(2)})'$ ($1 \leq i \leq n$). The lag model (1) becomes

$$\lambda_x(t) = e^{\alpha x^{(1)}} \{ (1 - 1_{(t>\theta)}) + 1_{(t>\theta)} e^{\beta x^{(2)}} \} \lambda_0(t). \quad (8)$$

Here $\phi = (\alpha, \beta) = (\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)$ is a row vector of regression coefficients and, as before, we assume that $\beta \neq 0$ for identifiability of θ . A more general model would allow each covariate to have a different lag effect by adding appropriate terms in the model (8); we shall not consider this here since often, as in our example, a single lag will suffice. Note that we can include the situation where a covariate can have an effect before the changepoint θ and another effect after by simply including that covariate in both $X^{(1)}$ and $X^{(2)}$. Thus model (8) is equivalent to that considered by Liang *et al.* (1990). For that situation, they developed a different procedure for testing the specific hypothesis that $\beta = 0$, whereas our interest lies in testing hypotheses of the form $H : \theta = \theta_0$ for particular θ_0 , while assuming $\beta \neq 0$. For us, the special case of no lag effect is represented by taking $\theta_0 = 0$.

By analogy with (2), the loglikelihood $L(\alpha, \beta, \theta)$ is given by

$$\sum_{\delta_i=1} \left[\log e^{\alpha X_i^{(1)}} \left\{ (1 - 1_{(Y_i>\theta)}) + 1_{(Y_i>\theta)} e^{\beta X_i^{(2)}} \right\} - \log \sum_{j \in \mathcal{R}_i} e^{\alpha X_j^{(1)}} \left\{ (1 - 1_{(Y_j>\theta)}) + 1_{(Y_j>\theta)} e^{\beta X_j^{(2)}} \right\} \right]. \quad (9)$$

We define $\hat{\phi} = (\hat{\alpha}, \hat{\beta})$ and $\hat{\theta}$ to be the maximizing values of ϕ and θ , respectively, and $\tilde{\phi}(\theta_0) = (\tilde{\alpha}(\theta_0), \tilde{\beta}(\theta_0))$ to be the maximizing value of ϕ when θ is held fixed at θ_0 . The test statistic T is now

$$T = \sup_{\theta, \alpha, \beta} L(\alpha, \beta, \theta) - \sup_{\alpha, \beta} L(\alpha, \beta, \theta_0) = L(\hat{\alpha}, \hat{\beta}, \hat{\theta}) - L(\tilde{\alpha}(\theta_0), \tilde{\beta}(\theta_0), \theta_0). \quad (10)$$

The value of this statistic can be obtained by repeated use of a standard Cox (1972) regression computer program with time dependent covariate $\tilde{X}(t) = (X^{(1)}, X^{(2)} 1_{(t>\theta)})$ in a manner similar to that in Section 2. For the asymptotic distribution of T under H_0 , analogous to (4, 5), we define independently identically distributed random

variables $\{Z_j^-, j = 1, 2, 3, \dots\}$ with distribution μ_- given by

$$\mu_-(C) = \frac{E\{e^{\alpha X^{(1)}} 1_{(Y \geq \theta_0, X^{(2)} \in C)}\}}{E\{e^{\alpha X^{(1)}} 1_{(Y \geq \theta_0)}\}} \quad \text{for } C \subseteq \mathcal{R}^p. \quad (11)$$

Similarly $\{Z_j^+, j = 1, 2, 3, \dots\}$ are independently and identically distributed with distribution μ_+ given by

$$\mu_+(C) = \frac{E\{e^{\alpha X^{(1)} + \beta X^{(2)}} 1_{(Y \geq \theta_0, X^{(2)} \in C)}\}}{E\{e^{\alpha X^{(1)} + \beta X^{(2)}} 1_{(Y \geq \theta_0)}\}} \quad \text{for } C \subseteq \mathcal{R}^p. \quad (12)$$

The sequences $\{U_j, V_j; j \geq 1\}$ are independently unit exponential as before and the random walks are now defined by

$$\begin{aligned} J_j &= \beta Z_j^- - \log \frac{E\{e^{\alpha X^{(1)} + \beta X^{(2)}} 1_{(Y \geq \theta_0)}\}}{E\{e^{\alpha X^{(1)}} 1_{(Y \geq \theta_0)}\}} U_j \\ A_m &= \sum_{j=1}^m J_j, \quad A_0 = 0 \\ K_j &= -\beta Z_j^+ - \log \frac{E\{e^{\alpha X^{(1)}} 1_{(Y \geq \theta_0)}\}}{E\{e^{\alpha X^{(1)} + \beta X^{(2)}} 1_{(Y \geq \theta_0)}\}} V_j \\ B_m &= \sum_{j=1}^m K_j, \quad B_0 = 0 \end{aligned}$$

The asymptotic null distribution of (10) is then given by (6) and (7). This distribution depends on unknown quantities so the Monte Carlo method is used to perform the test and obtain the p -value. In the definitions of the step sizes J_j and K_j , unknowns α and β are replaced by $\tilde{\alpha}(\theta_0)$ and $\tilde{\beta}(\theta_0)$, respectively. Corresponding empirical estimates are used in place of the distributions for the $\{Z_j^-\}$ and $\{Z_j^+\}$ in (11, 12). In this way the p -value of the test based on statistic T (10) can be obtained, directly generalizing the approach in §2.

The more general method was applied to the NPC Trial data of Section 4. Various unlagged covariates were considered, including gender, age and, of particular interest, plasma Se status upon study entry. It turned out that the inclusion of these covariates caused little difference in the lag p -values and the significance of treatment effect from those shown in Table 4, and so are not shown here. However it was important to have available methods to handle the generalized model (8), as described in this section, as it was certainly possible that inclusion of other covariates could have had an effect.

ACKNOWLEDGEMENTS

The first author's work was supported in part by Cancer Center CORE Grant CA 21765, by the American Lebanese Syrian Associated Charities(ALSAC). The second and third authors were supported by grants from the U.S. National Institutes of Health. This research was conducted using the resources of the Cornell Theory

Center, which receives major funding from the National Science Foundation and New York State with additional support from Advanced Research Projects Agency, the National Center for Research Resources at the National Institutes of Health, IBM Corporation and members of the Corporate Research Institute.

APPENDIX
Proofs

We outline the proof of the asymptotic distribution of the test statistic T defined by (10) in §5. Define the counting process $N_i(t) = 1_{(Y_i \leq t, \delta_i = 1)}$ and let

$$M_i(t) = N_i(t) - \int_0^t e^{\alpha X_i^{(1)}} \{1 - 1_{(Y_i > \theta_0)} + 1_{(Y_i > \theta_0)} e^{\beta X_i^{(2)}}\} \lambda_0(s) 1_{(Y_i \geq s)} ds$$

be the associated locally square integrable martingale (Fleming & Harrington, 1991). Recall that $\hat{\theta}$ and $\hat{\phi} = (\hat{\alpha}, \hat{\beta})$ maximize the loglikelihood (9) and $\tilde{\phi} = \tilde{\phi}(\theta_0) = (\tilde{\alpha}, \tilde{\beta}) = (\tilde{\alpha}(\theta_0), \tilde{\beta}(\theta_0))$ maximizes (9) under the constraint $\theta = \theta_0$. Under the null hypothesis, the true value $\theta = \theta_0$ and

$$\begin{aligned} T &= L(\hat{\phi}, \hat{\theta}) - L(\tilde{\phi}, \theta) = \{L(\hat{\phi}, \hat{\theta}) - L(\hat{\phi}, \theta)\} + \{L(\hat{\phi}, \theta) - L(\tilde{\phi}, \theta)\} \\ &= F(\hat{\alpha}, \hat{\beta}) + R_{1,n} = \hat{T} + R_{1,n} + R_{2,n}, \end{aligned}$$

where

$$F(\alpha, \beta) = \hat{T} = \sum_{i=1}^n \int_{\theta}^{\hat{\phi}} \left(-\beta X_i^{(2)} - \log \frac{\sum_{j=1}^n e^{\alpha X_j^{(1)}} 1_{(Y_j \geq s)}}{\sum_{j=1}^n e^{\alpha X_j^{(1)} + \beta X_j^{(2)}} 1_{(Y_j \geq s)}} \right) dN_i(s), \quad (13)$$

$$R_{1,n} = L(\hat{\phi}, \theta) - L(\tilde{\phi}, \theta) \quad \text{and} \quad R_{2,n} = F(\hat{\alpha}, \hat{\beta}) - F(\alpha, \beta).$$

In Luo (1996), Theorem 3.1(4), it is shown that the asymptotic distribution of (13) is given by (6) and (7) with random walks defined in Section 5. It suffices to show that

$$R_{1,n} = o_P(n^{-1/3}) \quad \text{and} \quad R_{2,n} = o_P(n^{-1/3}). \quad (14)$$

The second inequality in (14) can be shown by using Theorem 3.1(1,2) of Luo (1996) and a simple Taylor expansion. For the first, we can write

$$L(\hat{\phi}, \theta) - L(\tilde{\phi}, \theta) = \frac{1}{2}(\hat{\phi} - \tilde{\phi}) H_{L(\cdot, \theta)}(\phi^{**})(\hat{\phi} - \tilde{\phi})',$$

by Taylor's theorem for some ϕ^{**} between $\tilde{\phi}$ and $\hat{\phi}$. Note that $\nabla_{\phi}L(\cdot, \theta)|_{\phi=\hat{\phi}} - \nabla_{\phi}L(\cdot, \theta)|_{\phi=\tilde{\phi}}$ can be expanded at $\tilde{\phi}$ using Taylor's theorem and also can be written as a stochastic integral from $\hat{\theta}$ to θ , we have $\hat{\phi} - \tilde{\phi} = o_P(n^{-1/3})$ by applying some standard regularity conditions on the Hessian matrix of $L(\cdot, \theta)$ and Theorem 3.1(1,2). Therefore, $L(\hat{\phi}, \theta) - L(\tilde{\phi}, \theta) = o_P(n^{-1/3})$.

Define $\tilde{T} = F(\tilde{\alpha}, \tilde{\beta})$. We have $\tilde{T} - \hat{T} = o_P(n^{-1/3})$ similar to $R_{2,n}$. Hence T , \hat{T} and \tilde{T} are asymptotically identically distributed.

REFERENCES

- ABU-LIBDEH, H., TURNBULL, B. W. & CLARK, L. C. (1990). Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. *Biometrics*, **46**, 1017-34.
- ATKINSON, C. (1974). A Wiener-Hopf integral equation arising in some inference and queueing problems. *Biometrika* **61**, 277-283.
- CLARK, L.C., PATTERSON, B.H., WEED, D.L. & TURNBULL, B.W. (1991). Design issues in cancer chemoprevention trials using micronutrients: Application to skin cancer. *Cancer Bulletin* **43**, 519-524.
- CLARK, L.C., COMBS, G. F., TURNBULL, B.W., SLATE, E.H., CHALKER, D.K., CHOW, J., CURTIS, D., DAVIS, L.S., GLOVER, R.A., GRAHAM, G.F., GROSS, E.G., KRONGRAD, A., LESHER, J.L., PARK, H.K., SANDERS, B.B., SMITH, C.L., TAYLOR, J.R. AND THE NUTRITIONAL PREVENTION OF CANCER STUDY GROUP. (1996). Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: A randomized clinical trial. *Journal of the American Medical Association*. **276**(24), 1957-1963.
- COX D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc.* **B 34**, 187-220.
- FELLER, W. (1971). *An Introduction to Probability Theory and its Applications, Vol. II*. Wiley, New York.
- FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.
- GEORGE, S. L. (1994). Statistical considerations and modeling of clinical utility of tumor markers. *Hematology/Oncology Clinics of North America* Vol. **8**, No.3, 457-470.
- HINKLEY, D.V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1-17

- LIANG, K. Y., SELF, S. & LIU, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics* **46**, 783-793.
- LUO, X. (1996). The asymptotic distribution of MLE of treatment lag threshold. *J. Statist. Planning Inf.* **53**, 33-61.
- LUO, X., TURNBULL, B. W., CAI, H. & CLARK, L. C. (1994). Regression for censored survival data with lag effects. *Commun. Statist.-Theor. Meth.* **23**(12), 3417-3438.
- MATTHEWS, D. E. & FAREWELL, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics* **38**, 463-8.
- MATTHEWS, D. E., FAREWELL, V. T. & PYKE, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *Ann. Statist.* **13**, 583-91.
- MEINERT, C. L. (1986). *Clinical Trials: Design, Conduct, and Analysis*, Oxford University Press.
- MILLER, R. G. JR. (1981). *Survival Analysis*, Wiley, New York.
- MILLER, R. AND SIEGMUND, D. (1982). Maximally selected chi square statistics. *Biometrics* **38**, 1011-1016.
- SIMON, R. & ALTMAN, D. G. (1994). Statistical aspects of prognostic factor studies in oncology. *Br. J. Cancer* **69** (6), 979-985.
- ZUCKER, D. M. & LAKATOS, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment, *Biometrika* **77**, 853-864.

CAPTION FOR FIGURE

Figure 1: p -values for testing $\theta = \theta_0$ for simulated dataset. True lag $\theta = 3$.