Statistica Sinica **12**(2002), 707-723

# TESTS FOR LIFETIME UTILITY OR COST VIA CALIBRATING SURVIVAL TIME

Yijian Huang and Laura Lovato

*Fred Hutchinson Cancer Research Center and Wake Forest University*

*Abstract:* Testing for lifetime utility or cost is complicated with incomplete follow-up data. First, the marginal distribution in each sample is potentially nowhere identifiable. Second, the associated survival time may distribute differently across samples, whereas the difference is a nuisance. To overcome these difficulties, we propose to test the equivalence of the joint distributions of the variable of interest and survival time after calibrating the latter under the accelerated failure time model. Formulating the problem in the marked point process framework, we build upon and extend the well-known weighted log-rank statistics. Asymptotic theory has been developed and optimal weight functions derived. These tests are applied to a randomized clinical trial. Simulations show that they perform well with practical sample sizes.

*Key words and phrases:* Accelerated failure time model, dependent censoring, estimating equation, identifiability, lifetime medical cost, log-rank statistic, marked point process, minimum dispersion statistic, optimization, Pitman efficiency, quality adjusted survival time, semiparametric inference.

## 1. Introduction

In many medical studies, comparison of interventions is of interest with respect to lifetime medical cost or quality adjusted survival time. Being a mark of death (see Huang and Louis (1998)), such a lifetime measure with incomplete follow-up data poses a unique statistical challenge. One prominent phenomenon is the induced dependent censoring pattern on the scale of the mark. Much effort has been devoted to the one-sample nonparametric estimation problem in recent years, including Glasziou, Simes and Gelber (1990), Zhao and Tsiatis (1997, 2000), Lin, Feuer, Etzioni and Wax (1997), Huang and Louis (1998, 1999), and Bang and Tsiatis (2000). However, due to the fact that the marginal distribution of the mark may be largely unidentifiable, these one-sample estimation results do not readily lend themselves to the construction of hypothesis tests.

To be specific, let $T$ be the time-to-event, or survival time, and $U$ the mark of interest (e.g., lifetime medical cost) in a one-sample setting. The censoring time $C$ operates on $T$ and the following variables are observed: $X := T \wedge C$,

$Y := U\, I(T \leq C)$ and $\Delta := I(T \leq C)$, where $\wedge$ is the minimization operator and $I(.)$ is the indicator function. It is clear that, when the maximum support point of $C$, $\tau_C$, is less than that of $T$, the joint distribution of $(T, U)$ is not observable on $(\tau_C, \infty) \times (-\infty, \infty)$. Thus, the marginal cumulative distribution function $\text{pr}(U \leq u)$ can be nowhere identifiable. This issue is of practical concern since the duration of a medical study is typically finite and shorter than the maximum support point of survival time $T$.

The above data structure, as identified by Huang and Louis (1998), is basic and general to various applications in which a mark is of interest. Note that, even with additional information available in certain situations, the identifiability issue would not be completely resolved. For instance, instead of $Y$ only, suppose that one observes the accumulation process of the lifetime utility or cost up to follow-up time $X$. Huang (1999) showed that this additional information can be used to examine a sufficient condition for the marginal cumulative distribution function of $U$ to be identifiable up to, say $\iota$; that is, $\text{pr}(U > \iota | T \geq \tau_C) = 1$. However the existence of reasonably large $\iota$ can not be guaranteed, and thus the applicability of the nonparametric two-sample tests in Huang (1999) may be somewhat limited.

To avoid the identifiability issue, the notion of time-restricted variable has long been introduced in the literature of quality adjusted survival time and lifetime medical cost. For instance, five-year restricted lifetime cost is the cost accumulated up to death or five years, whichever happens earlier. While tests based on the time-restricted variable have been suggested (e.g., Glasziou, Simes and Gelber (1990), Zhao and Tsiatis (2000)), attempts to interpret them in terms of $U$ itself, as desired, are not appropriate and can be misleading. In particular for a two-sample problem with the same distributions of $U$, the time-restricted variable is in general differently distributed, as $T$ may have different distributions in the two populations.

Recognizing that testing for $U$ may not be intuitively based on its marginal distribution, we suggest considering the joint distribution of $T$ and $U$. We propose to calibrate the difference in the survival time $T$ and, in turn, to test the equivalence of the joint distributions. The accelerated failure time model will be used for the purpose of calibration. For generality, we consider the aforementioned basic data structure setting: each sample consists of realizations of $(X, Y, \Delta)$. In Section 2, we present some results on marked point processes, which are instrumental in the formulation of our proposed two-sample tests as presented in Section 3. We investigate the optimality of our tests in Section 4. The proposed tests are applied to a cancer clinical trial in Section 5 and simulation studies are reported in Section 6. Section 7 concludes with discussion.

## 2. Results on Marked Point Process

Marked point process may be viewed as an extension of the intensively investigated counting process (cf. Andersen, Borgan, Gill and Keiding (1993)), In this section, we establish a few general results in the one-sample setting; they are analogous to those well-known on counting process and will be the building blocks of our proposed tests. Hereafter, we suppose that all time variables are measured on a logarithmic scale.

Consider a sample: $(X_i, Y_i, \Delta_i)$, $i = 1, \ldots, n$, are $n$ independent and identically distributed replicates of $(X, Y, \Delta)$. For individual $i$, define the following processes:

$$R_i(t) = I(X_i \geq t), \qquad N_i(t,u) = I(X_i \leq t, Y_i \leq u)\Delta_i,$$
$$N_{mi}(t) = \int_{-\infty}^{\infty} u^m N_i(t, du) = I(X_i \leq t)\Delta_i Y_i^m, \qquad m = 0, 1, 2.$$

Note that $R_i$ and $N_{0i}$ are the familiar at-risk process and counting process, whereas $N_i$ was earlier introduced in Huang and Louis (1998). We refer to $N_{1i}$ as the marked process. Note that $N_{mi}$, $m = 1, 2$, differs from counting process $N_{0i}$ in the jump size, $Y_i^m$, which is random rather than the constant 1.

### 2.1. Hazard functions

Huang and Louis (1998) introduced the cumulative mark-specific hazard function

$$A(t,u) = \int_{-\infty}^{t} \lim_{h \downarrow 0} \frac{\mathrm{pr}(T < s + h, U \leq u | T \geq s)}{h}\, ds.$$

We now define

$$\Lambda_m(t) = \int_{-\infty}^{\infty} u^m A(t, du), \qquad \lambda_m(t) = d\Lambda_m(t)/dt, \qquad m = 0, 1, 2.$$

Note that $\Lambda_0$ is the cumulative hazard function and $\lambda_0$ is the hazard function of $T$. Further, $\lambda_m(t) = \lambda_0(t)E(U^m | T = t)$, $m = 1, 2$. We term $\lambda_1$ the marked hazard function and, in turn, $\Lambda_1$ the cumulative marked hazard function.

### 2.2. Martingale structure

We make the following assumptions.

(a) Censoring time $C$ is independent of the pair $(T, U)$.
(b) Survival time $T$ is continuously distributed.
(c) Mark $U$ is bounded.

These assumptions are largely satisfied in practical applications.

Let

$$M_{mi}(t) = N_{mi}(t) - \int_{-\infty}^{t} R_i(s)\, d\Lambda_m(s), \qquad m = 0, 1.$$

With respect to the filtration $\mathcal{F}_t = \sigma\{N_{mi}(s), I(X_i \le s, \Delta_i = 0) : m = 0, 1,\ s \le t,\ i = 1, \ldots, n\}$, both $M_{0i}(t)$ and $M_{1i}(t)$ are local martingales. Furthermore, the predictable variance process is given as

$$\left\langle \begin{pmatrix} M_{0i} \\ M_{1i} \end{pmatrix} \right\rangle (t) = \int_{-\infty}^{t} R_i(s)\, d \begin{Bmatrix} \Lambda_0(s)\ \Lambda_1(s) \\ \Lambda_1(s)\ \Lambda_2(s) \end{Bmatrix}. \tag{2.1}$$

These results extend those well-known on $M_{0i}(t)$ (cf. Andersen et al. (1993)); they are obtained with similar arguments.

### 2.3. Empirical hazard functions

Huang and Louis (1998) derived the nonparametric maximum likelihood estimator of the cumulative mark-specific hazard function:

$$\widehat{A}(t, u) = \int_{-\infty}^{t} \frac{N(ds, u)}{R(s)},$$

where $R = \sum_i R_i$ and $N = \sum_i N_i$. In view of the mapping from $A$ to $\Lambda_m$,

$$\widehat{\Lambda}_m(t) = \int_{-\infty}^{t} \frac{dN_m(s)}{R(s)}, \qquad m = 0, 1, 2,$$

is the nonparametric maximum likelihood estimator of $\Lambda_m$, where $N_m = \sum_i N_{mi}$. As easily recognized, $\widehat{\Lambda}_0$ is the Nelson−Aalen estimator of $\Lambda_0$.

Huang and Louis ((1998), proofs of Theorems 4 and 5) showed that, on $(-\infty, L] \times (-\infty, \infty)$, $\widehat{A}$ is uniformly and strongly consistent for $A$ and $n^{1/2}(\widehat{A} - A)$ is asymptotically normal, where constant $L$ satisfies $\mathrm{pr}(X > L) > 0$. With condition (c), it follows that $\widehat{\Lambda}_m$ is uniformly and strongly consistent for $\Lambda_m$ and $n^{1/2}(\widehat{\Lambda}_m - \Lambda_m)$ is asymptotically normal, on $(-\infty, L]$ and for $m = 0, 1, 2$.

In our later development, we make use of the properties of $\widehat{\Lambda}_0$ and $\widehat{\Lambda}_1$ on multiple time scales. Specifically, considering a sequence of location changes $b_n \to 0$ as $n \to \infty$, we are interested in the asymptotic behavior of $\widehat{\Lambda}_m(t + b_n)$ relative to $\widehat{\Lambda}_m(t)$, $m = 0, 1$. Note that $N_0$, $N_1$, and $R$ are empirical-type processes as studied by Lai and Ying (1988). In addition to conditions (a)−(c), we assume the following.

(d) Survival time $T$ has a bounded and a continuously differentiable density function $f_T$ which satisfies $\int_{-\infty}^{\infty} \sup_{t \le s \le t+d} |f_T'(s)|\, dt < \infty$ for some $d > 0$.
(e) Censoring time $C$ has a bounded density function.
(f) $E(|X|^r) < \infty$ for some $r > 0$.

Then, similarly to Theorems 1, 2, and 3 of Lai and Ying (1988), one can show that, almost surely,

$$\sup_{t \le L} |\widehat{\Lambda}_m(t + b_n) - \Lambda_m(t + b_n) - \widehat{\Lambda}_m(t) + \Lambda_m(t)| = o(n^{-1/2} + |b_n|), \qquad m = 0, 1, \tag{2.2}$$

as $n \to \infty$ and $b_n \to 0$. Yang (1998) gave a similar result for the Nelson−Aalen estimator $\widehat{\Lambda}_0$.

## 3. The Proposed Two-Sample Tests

Now considering the two-sample problem, we use the notation introduced earlier for one sample and add an asterisk to indicate the other. Since the marginal distributions of $U$ and $U^*$ may be nowhere identifiable, we construct tests on the basis of the joint distributions of survival time and the mark after calibrating the former across the two samples. For the purpose of calibration, we adopt the commonly-used accelerated failure time model for the difference of $T$ and $T^*$:

$$T \stackrel{\mathcal{D}}{=} T^* + \beta \text{ for some } \beta, \tag{3.1}$$

where $\stackrel{\mathcal{D}}{=}$ denotes equivalence in distribution; some justification of the model in clinical studies can be found in Kalbfleisch and Prentice (1980). Recall that $T$ and $T^*$ are on a logarithmic scale and thus the above location-shift model is a scale-change model on the original scale.

The test of interest is on the difference of $U$ and $U^*$. The null hypothesis is specified as $\mathcal{H}_0 : (T, U)^T \stackrel{\mathcal{D}}{=} (T^* + \beta, U^*)^T$. Apparently, $U \stackrel{\mathcal{D}}{=} U^*$ under $\mathcal{H}_0$. This hypothesis is relevant particularly in the clinical trial setting; cf. other null hypotheses considered by, for example, Zhou, Melfi and Hui (1997) for the uncensored medical cost problem. Writing $\mu(t) = E(U|T = t)$ and $\mu^*(t) = E(U^*|T^* = t)$, we consider the alternative hypothesis $\mathcal{H}_A$ : Given (3.1), $\mu(t) \geq \mu^*(t - \beta)$ or $\mu(t) \leq \mu^*(t - \beta)$, where the inequality is strict for at least some $t$. Clearly $\mathcal{H}_A$ implies different marginal distributions of $U$ and $U^*$. For a large sample study, $n/n^*$ converges to a finite constant bounded away from 0 as $n_\bullet = n + n^* \to \infty$. To investigate the power of the proposed tests, we appeal to the notion of Pitman efficiency under the local alternative $\mathcal{H}_A^L$ : Given (3.1), $n_\bullet^{1/2}\{\mu(t) - \mu^*(t - \beta)\} \to \delta(t)$ uniformly on $t \in (-\infty, \infty)$, for some bounded function $\delta(t)$. (The dependence of $\mu$ and $\mu^*$ on $n_\bullet$ has been suppressed for notational convenience.)

Accommodating possibly different censoring in the two samples, we build upon and extend the well-known weighted log-rank statistics. Define

$$\begin{aligned}\xi_m(b) &= \int_{-\infty}^{\tau} w_m(t, b) \left\{ \frac{dN_m(t)}{R(t)} - \frac{dN_m^*(t - b)}{R^*(t - b)} \right\} \\ &= \int_{-\infty}^{\tau} w_m(t, b)\{d\widehat{\Lambda}_m(t) - d\widehat{\Lambda}_m^*(t - b)\}, \qquad m = 0, 1. \end{aligned} \tag{3.2}$$

Here, we have imposed an upper integration limit $\tau < \infty$ such that $\text{pr}(X > \tau)\text{pr}(X^* + \beta > \tau) > 0$ to avoid complications with tail instability; in practice, one

may take $\tau$ large enough to cover all the follow-up times for a specific dataset. The weight function $w_m(t,b)$ is a nonnegative process that vanishes whenever $R(t)=0$ or $R^*(t-b)=0$. Note that $\xi_0$ is the familiar rank statistic for survival time, where $w_0(t,b)=n_\bullet^{-1}R(t)R^*(t-b)/\{R(t)+R^*(t-b)\}$ and $w_0(t,b)=n_\bullet^{-2}R(t)R^*(t-b)$ correspond to the log-rank and Gehan statistics, respectively. In comparison, $\xi_1$ is obtained with the marked processes in place of the counting processes in $\xi_0$.

One should note that $\beta$ is a nuisance parameter. There are situations in which $\beta$ is known *a priori*: An intervention may reduce the lifetime medical cost, say, but it is known to have no impact on the survival time; in this case, $\beta$ is known to be 0. Of course, $\beta$ is usually unknown. In the following, we construct tests under both situations.

### 3.1. Tests when $\beta$ is known

To study the properties of $\xi_m(\beta)$, we need the following.

(g) Weight function $w_m(t,\beta)$, $m=0,1$, converges uniformly in $t\in(-\infty,\tau]$ to a nonrandom function $W_m(t,\beta)$, in probability.

Clearly, the log-rank and Gehan weight functions satisfy this condition.

We establish properties of $\xi_m(\beta)$ based on martingale theory. The weight function $w_m(t,\beta)$ will be implicitly taken as $\mathcal{F}_t\bigvee\mathcal{F}^*_{t-\beta}$-predictable. This is true for most commonly chosen weight functions; otherwise, one can always find a $\mathcal{F}_t\bigvee\mathcal{F}^*_{t-\beta}$-predictable replacement (e.g., $W_m(t,\beta)$) such that the resulting $\xi_m(\beta)$ is asymptotically equivalent.

Under $\mathcal{H}_0$, a standard technique in survival analysis yields

$$\xi_m(\beta)=\int_{-\infty}^{\tau}w_m(t,\beta)\left\{\frac{dM_m(t)}{R(t)}-\frac{dM^*_m(t-\beta)}{R^*(t-\beta)}\right\},\qquad m=0,1,$$

where $M_m=\sum_i M_{mi}$ and $M^*_m=\sum_i M^*_{mi}$. Rebolledo's Central Limit Theorem for local martingales then asserts that $n_\bullet^{1/2}\{\xi_0(\beta),\xi_1(\beta)\}^T$ is asymptotically normal with mean $(0,0)^T$ and variance

$$\mathbf{\Sigma}(\beta)=\int_{-\infty}^{\tau}B(t)\begin{Bmatrix}W_0(t,\beta)^2\,d\Lambda_0(t) & W_0(t,\beta)W_1(t,\beta)\,d\Lambda_1(t)\\ W_0(t,\beta)W_1(t,\beta)\,d\Lambda_1(t) & W_1(t,\beta)^2\,d\Lambda_2(t)\end{Bmatrix},$$

where $B(t)$ is the limit of $n_\bullet I\{R(t)R^*(t-\beta)>0\}\{R(t)^{-1}+R^*(t-\beta)^{-1}\}$. The asymptotic variance is derived from (2.1), and it can be consistently estimated by

$$\widehat{\mathbf{\Sigma}}(\beta)=n_\bullet\int_{-\infty}^{\tau}\left\{\frac{1}{R(t)}+\frac{1}{R^*(t-\beta)}\right\}\times\begin{Bmatrix}w_0(t,\beta)^2\,d\widehat{\Lambda}_{\bullet0}(t,\beta) & w_0(t,\beta)w_1(t,\beta)\,d\widehat{\Lambda}_{\bullet1}(t,\beta)\\ w_0(t,\beta)w_1(t,\beta)\,d\widehat{\Lambda}_{\bullet1}(t,\beta) & w_1(t,\beta)^2\,d\widehat{\Lambda}_{\bullet2}(t,\beta)\end{Bmatrix},$$

where $\widehat{\Lambda}_{\bullet m}(t,\beta) = \int_{-\infty}^{t}\{R(t)+R^*(t-\beta)\}^{-1}\, d\{N_m(t)+N_m^*(t-\beta)\}$ is the pooled estimator of cumulative hazard function $\Lambda_m$, $m = 0, 1, 2$. One can easily show that, under local alternative $\mathcal{H}_A^L$, we have the same asymptotic results except that the mean of $n_\bullet^{1/2}\{\xi_0(\beta), \xi_1(\beta)\}^T$ is then $\{0, \int_{-\infty}^{\tau} W_1(t,\beta)\delta(t)\lambda_0(t)\, dt\}^T$.

Although $\xi_1(\beta)$ is an obvious test statistic for $\mathcal{H}_0$, a class of tests can be constructed based on $\xi_0(\beta)$ and $\xi_1(\beta)$, and more powerful tests are of interest. Write $\varepsilon_{pq}$, $p, q = 1, 2$, as the $(p, q)$th element of $\mathbf{\Sigma}(\beta)$ and $\widehat{\varepsilon}_{pq}$ as its counterpart in $\widehat{\mathbf{\Sigma}}(\beta)$.

**Theorem 1.** *Assume* (a)−(c), *for both samples, and* $(g)$ *hold. The test based on*

$$\Phi = n_\bullet^{1/2}\frac{\xi_1(\beta) - \widehat{\varepsilon}_{11}^{-1}\widehat{\varepsilon}_{12}\xi_0(\beta)}{(\widehat{\varepsilon}_{22} - \widehat{\varepsilon}_{11}^{-1}\widehat{\varepsilon}_{12}^2)^{1/2}}$$

*achieves the maximum asymptotic local power among those based on* $\xi_0(\beta)$ *and* $\xi_1(\beta)$. *Under* $\mathcal{H}_0$, $\Phi$ *is asymptotically standard normal.*

**Proof.** With the asymptotic distribution of $\{\xi_0(\beta), \xi_1(\beta)\}^T$ under $\mathcal{H}_0 \bigcup \mathcal{H}_A^L$, algebraic derivations show that an asymptotically and locally efficient likelihood ratio test is based on statistic $n_\bullet^{1/2}\{\xi_1(\beta) - \varepsilon_{11}^{-1}\varepsilon_{12}\xi_0(\beta)\}$, the asymptotic variance of which is $\varepsilon_{22} - \varepsilon_{11}^{-1}\varepsilon_{12}^2$. The desired result follows since $\Phi$ is asymptotically equivalent to $n_\bullet^{1/2}\{\xi_1(\beta) - \varepsilon_{11}^{-1}\varepsilon_{12}\xi_0(\beta)\}/(\varepsilon_{22} - \varepsilon_{11}^{-1}\varepsilon_{12}^2)^{1/2}$.

### 3.2. Tests when $\beta$ is unknown

Before proceeding, we impose a stronger condition on $w_m(t, b)$ to facilitate large-sample arguments.

(h) The weight function $w_m(t, b)$ converges to $W_m(t, b)$ uniformly in $t \in (-\infty, \tau]$ and in $b$, $|b - \beta| \le B$ for some $B > 0$, $m = 0, 1$, in probability. The limit $W_m(t, b)$ is continuous at $b = \beta$ uniformly in $t \in (-\infty, \tau]$. Further, in probability,

$$\limsup_{n_\bullet\to\infty} \sup_{|b-\beta|\le B} \int_{-\infty}^{\tau} |w_m(dt, b)| < \infty.$$

This condition can be easily verified for the log-rank and Gehan weight functions.

Louis (1981) and Wei and Gail (1983) suggested a consistent estimator $\widehat{\beta}_0$, a zero-crossing of $\xi_0(b)$. Another estimator $\widehat{\beta}_1$ can be constructed to be a zero-crossing of $\xi_1(b)$ in a neighborhood of $\widehat{\beta}_0$. By the arguments of Gill and Schumacher (1987), $\widehat{\beta}_0 - \widehat{\beta}_1$ can be used as a test statistic with both $\widehat{\beta}_0$ and $\widehat{\beta}_1$ being consistent for $\beta$ under $\mathcal{H}_0$. In view of (3.2), and from (2.2), we have, for $b \to \beta$ under $\mathcal{H}_0 \bigcup \mathcal{H}_A^L$,

$$\xi_m(b) = \xi_m(\beta) + (b - \beta)r_m + o_p(n^{-1/2} + |b - \beta|), \qquad m = 0, 1, \tag{3.3}$$

where $r_m = \int_{-\infty}^{\tau} W_m(t,\beta)\, d\lambda_m(t)$. Given the asymptotic results on $\xi_m(\beta)$ from Section 3.1, it is easily shown that $\widehat{\beta}_m - \beta = -r_m^{-1}\xi_m(\beta) + o_p(n^{-1/2})$, $m = 0, 1$. Therefore, $\widehat{\beta}_0 - \widehat{\beta}_1$ is asymptotically normal with mean 0 under $\mathcal{H}_0$.

Since calculating $\widehat{\beta}_0 - \widehat{\beta}_1$ requires solving two discrete equations, one may use $\xi_1(\widehat{\beta}_0)$ as a computationally efficient test statistic. Expression (3.3) leads to $\xi_1(\widehat{\beta}_0) = r_1(\widehat{\beta}_0 - \widehat{\beta}_1) + o_p(n^{-1/2})$. Indeed, the test based on $\xi_1(\widehat{\beta}_0)$ is asymptotically equivalent to that on $\widehat{\beta}_0 - \widehat{\beta}_1$ under $\mathcal{H}_0 \bigcup \mathcal{H}_A^L$.

Despite $\widehat{\beta}_0 - \widehat{\beta}_1$ and $\xi_1(\widehat{\beta}_0)$ being reasonable test statistics, neither of their variances can be easily estimated. For example, under $\mathcal{H}_0 \bigcup \mathcal{H}_A^L$,

$$\begin{aligned}\mathrm{var}\{n_\bullet^{1/2}\xi_1(\widehat{\beta}_0)\} = \int \{&r_0^{-2}r_1^2 W_0(t,\beta)^2\lambda_0(t) - 2r_0^{-1}r_1 W_0(t)W_1(t,\beta)\lambda_1(t) \\ &+ W_1(t,\beta)^2\lambda_2(t)\}B(t)dt, \end{aligned} \tag{3.4}$$

and $r_m$ involves $\lambda_m$. To overcome this difficulty, we suggest a different test statistic.

**Theorem 2.** *Let* $\Psi = \min\ n_\bullet\{\xi_0(b), \xi_1(b)\}\widehat{\mathbf{\Sigma}}(\widehat{\beta}_0)^{-1}\{\xi_0(b), \xi_1(b)\}^T$, *where the minimization is taken over* $b$ *in a neighborhood of* $\widehat{\beta}_0$. *Assume that* (a)–(f), *for both samples, and* (h) *hold. Under* $\mathcal{H}_0$, $\Psi$ *is asymptotically* $\chi^2(1)$. *The test based on* $\Psi$ *is locally equivalent to that based on* $\widehat{\beta}_0 - \widehat{\beta}_1$ *or* $\xi_1(\widehat{\beta}_0)$.

**Proof.** First one can show that $\widehat{\mathbf{\Sigma}}(\widehat{\beta}_0)$ is consistent for $\mathbf{\Sigma}(\beta)$ under $\mathcal{H}_0 \bigcup \mathcal{H}_A^L$, since $\widehat{\beta}_0$ converges to $\beta$. Then by arguments similar to those of Wei, Ying and Lin ((1990), Appendix 2), $\Psi$ is asymptotically equivalent to $n_\bullet(\widehat{\beta}_0 - \widehat{\beta}_1)^2/\mathrm{var}\{n_\bullet^{1/2}(\widehat{\beta}_0 - \widehat{\beta}_1)\}$.

Computationally, one may perform a grid search to calculate the minimum dispersion statistic $\Psi$.

## 4. Optimality Consideration of the Weight Functions

The proposed tests based on $\Phi$ or $\Psi$, with $\beta$ known or unknown, have wide choices of weight functions $w_0$ and $w_1$. One natural question is how to choose them for test efficiency. We address this optimality issue under local alternative $\mathcal{H}_A^L$.

### 4.1. Test based on $\Phi$ when $\beta$ is known

Under $\mathcal{H}_A^L$, $\Phi$ is asymptotically normal with mean $\int_{-\infty}^{\tau} W_1(t,\beta)\delta(t)\lambda_0(t)\, dt / (\varepsilon_{22} - \varepsilon_{11}^{-1}\varepsilon_{12}^2)^{1/2}$ and unit variance. Thus, the efficacy of the test is

$$\mathrm{eff}(\Phi; W_0, W_1) = \frac{\left\{\int_{-\infty}^{\tau} W_1(t,\beta)\delta(t)\lambda_0(t)\, dt\right\}^2}{\varepsilon_{22} - \varepsilon_{11}^{-1}\varepsilon_{12}^2}.$$

Note that $\varepsilon_{11}$, $\varepsilon_{12}$ and $\varepsilon_{22}$ are functions of $W_0$ or $W_1$ and that efficacy is invariant under a scalar multiplication of either weight function.

In the Appendix, by applying the Cauchy–Schwartz inequality, we obtain optimal weight functions and associated optimal efficacy as

$$\left.\begin{aligned} W_{0\,\mathrm{opt}}(t) &\propto \delta(t)\sigma(t)^{-2}\lambda_0(t)^{-1}\lambda_1(t)B(t)^{-1}, \\ W_{1\,\mathrm{opt}}(t) &\propto \delta(t)\sigma(t)^{-2}B(t)^{-1}, \\ \mathrm{eff}(\Phi; W_{0\,\mathrm{opt}}, W_{1\,\mathrm{opt}}) &= \int_{-\infty}^{\tau} \delta(t)^2\sigma(t)^{-2}\lambda_0(t)B(t)^{-1}\,dt, \end{aligned}\right\} \tag{4.1}$$

where $\sigma(t)^2 = \mathrm{var}(U|T=t) = \lambda_2(t)/\lambda_0(t) - \lambda_1(t)^2/\lambda_0(t)^2$.

### 4.2. Test based on $\Psi$ when $\beta$ is unknown

From Theorem 2, the efficacy of the test based on $\Psi$ is

$$\mathrm{eff}(\Psi; W_0, W_1) = \frac{\left\{\int_{-\infty}^{\tau} W_1(t,\beta)\lambda_0(t)\delta(t)\,dt\right\}^2}{\mathrm{var}\{n_\bullet^{1/2}\xi_1(\widehat{\beta}_0)\}},$$

where $\mathrm{var}\{n_\bullet^{1/2}\xi_1(\widehat{\beta}_0)\}$ as given in (3.4) is a function of $W_0$ and $W_1$. The efficacy can be optimized over $W_0$ and $W_1$, similar to the study of the $\Phi$ test but algebraically more tedious. The resulting optimal weight functions and the associated efficacy are complicated, and thus omitted.

### 4.3. Special case: Independence between mark and survival time

We have shown that optimal $W_0$ and $W_1$ can be derived for tests based on $\Phi$ and $\Psi$. Unfortunately, this does not imply that optimal $w_0$ and $w_1$ can be feasibly constructed from the data. In particular, even for the test based on $\Phi$, the rather simple expressions for $W_{0\,\mathrm{opt}}$ and $W_{1\,\mathrm{opt}}$ in (4.1) involve $\lambda_0$, $\lambda_1$ and $\lambda_2$, and these are difficult to estimate. Therefore, the preceding results may not be sufficient to provide practical guidelines for choosing weight functions. For this purpose, we consider a special scenario where $U$ is independent of $T$ and $U^*$ is independent of $T^*$. Optimal $w_0$ and $w_1$ for both tests can be constructed in this case, and they may serve as working weight functions in general.

Under the independence scenario, $\delta(t) = \delta$ and $\sigma(t)^2 = \mathrm{var}(U) = \sigma^2$ are constant. For the test based on $\Phi$, from (4.1) it is easily shown that the optimal $W_0$ and $W_1$ are proportional to $B(t)^{-1}$. Therefore, using the log-rank weight function for both $w_0$ and $w_1$ achieves optimality. For the test based on $\Psi$, interestingly, the optimal $w_0$ and $w_1$ are also the log-rank weight function (see the

Appendix). Furthermore, the two tests have the same optimal efficacy:

$$\delta^2\sigma^{-2}\int_{-\infty}^{\tau}\lambda_0(t)B(t)^{-1}\,dt. \tag{4.2}$$

Notice that uncensored $\{Y_i : \Delta_i = 1, i = 1,\cdots,n\}$ and $\{Y_i^* : \Delta_i^* = 1, i = 1,\cdots,n^*\}$ are unbiased samples of $U$ and $U^*$, respectively. Therefore, the $t$-test is valid and also asymptotically efficient when the mark is normally distributed. To compare the efficiency of our tests to that of the $t$-test, we note that the optimal efficacy (4.2) can be rewritten as

$$\delta^2\sigma^{-2}\int_{-\infty}^{\tau}\frac{p(1-p)S_C(t)S_{C^*}(t-\beta)}{pS_C(t)+(1-p)S_{C^*}(t-\beta)}\,dF_T(t),$$

where $p = \lim n/n_\bullet$, $S_C(t) = \mathrm{pr}(C > t)$, $S_{C^*}(t) = \mathrm{pr}(C^* > t)$, and $F_T(t) = \mathrm{pr}(T \le t)$. It is interesting to observe that, if $C \overset{\mathcal{D}}{=} C^* + \beta$, the efficacy becomes $\delta^2\sigma^{-2}p(1-p)\mathrm{pr}(T \le C\wedge\tau)$, which is the same as that of the $t$-test. Nevertheless, in general and as expected, the $t$-test is more efficient.

## 5. Illustration: Application to a Cancer Clinical Trial

Our research is motivated by a recently-completed randomized clinical trial of the Southwest Oncology Group (SWOG). The study was designed to investigate Vinorelbine plus Cisplatin versus Paclitaxel plus Carboplatin therapies in earlier untreated patients with advanced non-small cell lung cancer (Kelly et al. (2001)). One objective, among others, is to compare lifetime medical costs of the two treatments.

During the study, resource utilization was tracked for each participant and reported at months 3, 6, 12, 18, and 24 for the five corresponding previous intervals. These resources consisted of supportive care medications, blood products, medical procedures, protocol and non-protocol related treatments, and other medical care inpatient days or outpatient visits. Costs were assigned to them using national databases, upon adjustment to 1998 US dollars according to the medical care component of the Consumer Price Index. For the analysis of lifetime medical cost, the follow-up duration of a participant is referred and confined to that of the cost accumulation. Of the total 408 participants eligible for the study, 10 were excluded from our analysis due to insufficient documentation. Among the remaining 398 participants, 198 were randomized to receive Vinorelbine plus Cisplatin. The median follow-up time was 6 months and 36% of the participants had their lifetime cost censored. The Nelson–Aalen estimator for the survival time and the estimated cumulative marked hazard function for the lifetime cost are presented in Figure 1 by treatment regimen. Indeed, a substantial portion of

the participants survived beyond 24 months, which suggests that the marginal distribution of the lifetime medical cost for each group may be nowhere identifiable, as discussed in Section 1.

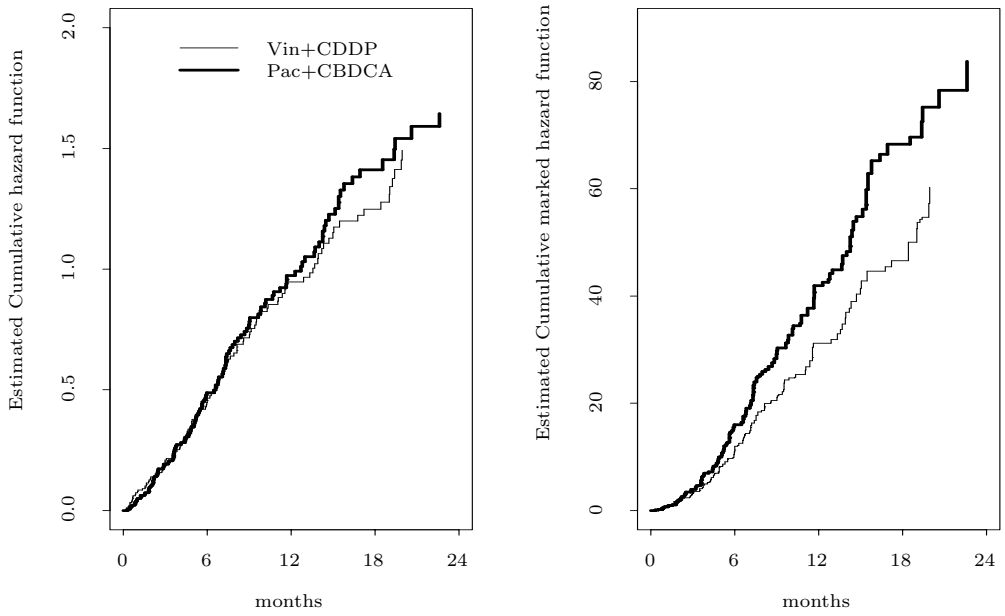


Figure 1. The Nelson–Aalen estimator for survival time and the estimated cumulative marked hazard function for the lifetime medical cost, in thousand US dollars, by treatment regimen: Vinorelbine plus Cisplatin (Vin+CDDP) versus Paclitaxel plus Carboplatin (Pac+CBDCA).

We applied our $\Psi$ test to the lifetime medical cost with $w_0 = w_1$. Using the log-rank and Gehan weight functions, we obtained the minimum dispersion statistics 6.37 and 5.76, respectively. In comparison to the 95th percentile of the $\chi^2(1)$ distribution, they show that the cost associated with Paclitaxel plus Carboplatin is significantly higher than that with Vinorelbine plus Cisplatin.

One should notice the inappropriateness of constructing tests based on $\Phi$ for this study, despite the fact that Figure 1 suggests little difference in survival time between the two treatments. Indeed, the difference was not known *a priori* and its determination was the primary objective of the trial.

## 6. Simulation Studies

Our proposed two-sample tests for lifetime utility or cost are justified on the basis of asymptotic theory. Extensive simulations have also been conducted to evaluate their finite-sample performance.

We considered $(T, U)^T \stackrel{\mathcal{D}}{=} (T^* + \beta, U^* + \eta)^T$ for constants $\beta$ and $\eta$, where $T$ and $U$ have standard extreme value and $[0, 1]$ uniform marginal distributions, respectively. Note that $\exp(T)$ is exponentially distributed with unit rate. Jointly, $(T, U)^T$ follows Frank's bivariate family (cf., Genest (1987)) and values $\gamma = 0, 10$ of the association parameter were chosen for independence and moderate association, with the corresponding Kendall's rank correlation coefficients being 0 and 0.67. Notice that our proposed tests in this setting have the same operating characteristics under $\gamma$ and $-\gamma$. We investigated situations both with and without censoring. In the presence of censoring, $\exp(C)$ and $\exp(C^*)$ were specified to have the same uniform distribution on $[0, c]$, where $c$ was selected to determine proportions of censoring. To contrast the proposed tests, we also investigated the $t$-test based on the uncensored mark values.

The first question we addressed was the accuracy of the size of our proposed tests under $\mathcal{H}_0$, i.e., $\eta = 0$. We investigated various configurations of sample sizes, associations between survival time and the mark, and censoring rates. We set $\beta = -0.2$ and chose $w_0 = w_1$ to be either the log-rank or Gehan weight function. The results are presented in Table 1 with 5000 iterations for each configuration. As seen, the proposed tests, with and without the knowledge of $\beta$, achieve reasonably accurate sizes. In comparison, the $t$-test is invalid unless the censoring is absent or the mark is independent of the survival time.

To assess the power of the proposed tests, we considered the same settings as in Table 1 except with $\eta = -0.1$. Shown in Table 2, with the same choice of weight functions, the test based on $\Phi$ is generally no less powerful than that based on $\Psi$. Notice that the difference is small when the survival time and the mark are independent. Further, between the log-rank and Gehan weight functions, under the independence scenario using the former is more powerful for both tests. These results conform with the earlier analysis given in Section 4. On the other hand, under moderate association, for the test based on $\Phi$ using Gehan weight function is more powerful, whereas for that based on $\Psi$ the relative performance of the two weight choices depends on censoring. As compared to the $t$-test when it is valid, the test based on $\Phi$ with log-rank wight function shows similar power, especially as the sample size increases.

Following a suggestion from a reviewer, we also studied the scenario when the distribution of the mark is highly skewed. This is particularly relevant when lifetime medical cost is under consideration. Our study showed that the proposed tests perform reasonably well, similarly to the setting presented earlier. For this reason, the results have been omitted.

In most practical situations, $\beta$ is unknown and tests based on $\Psi$ would be used. Our numerical experience suggests that the log-rank weight function is an appropriate choice.

Table 1. Empirical type **I** error, in percent, of 5000 random samples at the nominal level of 5%, with $\beta = -0.2$ and $\eta = 0$.

| size | assoc. | censoring | $t$-test on | $\Phi - \beta$ known | | $\Psi - \beta$ unknown | |
|---|---|---|---|---|---|---|---|
| $n = n^*$ | Kendall's | rate (%) | uncensored | log-rank | Gehan | log-rank | Gehan |
| 25 | 0 | 0 | 5.02 | 4.52 | 4.68 | 3.36 | 4.00 |
| | | 20 | 5.72 | 5.02 | 4.70 | 3.76 | 3.98 |
| | | 40 | 6.12 | 4.86 | 4.48 | 3.52 | 3.98 |
| | 0.67 | 0 | 5.86 | 5.68 | 4.54 | 4.48 | 4.56 |
| | | 20 | 6.08 | 5.46 | 4.50 | 4.22 | 4.42 |
| | | 40 | 6.92 | 5.08 | 4.60 | 4.92 | 5.00 |
| 50 | 0 | 0 | 4.98 | 4.94 | 4.98 | 4.14 | 4.70 |
| | | 20 | 5.28 | 4.90 | 4.70 | 3.92 | 4.38 |
| | | 40 | 5.16 | 4.68 | 5.28 | 4.06 | 4.84 |
| | 0.67 | 0 | 5.76 | 5.80 | 5.28 | 4.78 | 4.82 |
| | | 20 | 5.80 | 5.18 | 5.12 | 4.40 | 4.78 |
| | | 40 | 7.70 | 5.34 | 5.00 | 4.74 | 5.30 |
| 100 | 0 | 0 | 5.50 | 5.50 | 5.28 | 4.92 | 4.84 |
| | | 20 | 4.88 | 4.38 | 5.18 | 3.94 | 5.12 |
| | | 40 | 4.86 | 4.76 | 5.80 | 4.38 | 5.44 |
| | 0.67 | 0 | 4.92 | 5.58 | 6.02 | 4.52 | 5.12 |
| | | 20 | 6.62 | 5.42 | 4.98 | 5.10 | 5.14 |
| | | 40 | 8.74 | 5.34 | 5.38 | 4.74 | 5.16 |
| 200 | 0 | 0 | 6.12 | 5.10 | 4.86 | 4.92 | 4.80 |
| | | 20 | 4.88 | 4.12 | 5.70 | 3.90 | 5.68 |
| | | 40 | 4.56 | 4.08 | 6.04 | 3.82 | 5.96 |
| | 0.67 | 0 | 5.06 | 5.34 | 6.54 | 4.42 | 5.32 |
| | | 20 | 7.36 | 4.62 | 5.18 | 4.70 | 4.96 |
| | | 40 | 12.86 | 5.10 | 5.16 | 5.18 | 5.22 |

Note: Censoring rates of 20% and 40% correspond to the distributions of Uniform[0,5.2] and Uniform[0,2.4], respectively, for the exponential of the censoring time.

## 7. Remarks and Extension

In medical research today, utility and cost evaluation is becoming an accepted and often required adjunct to standard safety and efficacy analysis. Evaluations of outcomes, such as lifetime medical cost and quality adjusted survival time, are now frequently incorporated into controlled clinical trials of new medical therapies. However, statistical analysis of such outcomes with incomplete follow-up data is challenging due largely to the identifiability issue. In particular for two-sample comparison, most existing tests are based on a time-restricted measure instead of lifetime utility or cost itself. In this paper, we have proposed to

appropriately calibrate difference in survival time and further to construct tests for lifetime utility or cost on the basis of the joint distribution. The advantage of our proposal is its direct interpretability on the variable of interest.

Table 2. Empirical power, in percent, of 5000 random samples at the nominal level of 5%, with $\beta = -0.2$ and $\eta = -0.1$.

| size | assoc. | censoring | $t$-test on | $\Phi - \beta$ known | | $\Psi - \beta$ unknown | |
|---|---|---|---|---|---|---|---|
| $n = n^*$ | Kendall's | rate (%) | uncensored | log-rank | Gehan | log-rank | Gehan |
| 25 | 0 | 0 | 23.14 | 20.62 | 15.82 | 17.46 | 15.04 |
| | | 20 | 19.36 | 16.76 | 15.12 | 14.40 | 13.62 |
| | | 40 | 16.96 | 14.22 | 11.72 | 11.26 | 10.64 |
| | 0.67 | 0 | 23.60 | 22.38 | 25.10 | 14.40 | 12.80 |
| | | 20 | — | 19.40 | 24.14 | 11.76 | 11.68 |
| | | 40 | — | 18.24 | 22.94 | 11.78 | 12.66 |
| 50 | 0 | 0 | 40.00 | 37.96 | 30.64 | 35.60 | 29.70 |
| | | 20 | 33.66 | 30.90 | 25.52 | 28.44 | 24.70 |
| | | 40 | 26.02 | 23.76 | 20.42 | 21.30 | 19.72 |
| | 0.67 | 0 | 40.84 | 39.36 | 48.62 | 25.62 | 23.20 |
| | | 20 | — | 37.08 | 46.68 | 22.22 | 22.60 |
| | | 40 | — | 33.18 | 44.76 | 20.88 | 22.70 |
| 100 | 0 | 0 | 68.12 | 67.10 | 54.46 | 65.40 | 54.04 |
| | | 20 | 57.74 | 56.62 | 46.70 | 54.64 | 46.20 |
| | | 40 | 46.12 | 44.16 | 37.98 | 42.54 | 37.76 |
| | 0.67 | 0 | 68.86 | 67.82 | 78.16 | 48.12 | 42.44 |
| | | 20 | — | 62.00 | 77.14 | 41.60 | 41.60 |
| | | 40 | — | 59.52 | 76.20 | 37.80 | 41.60 |
| 200 | 0 | 0 | 92.60 | 92.34 | 84.38 | 92.18 | 84.18 |
| | | 20 | 87.44 | 87.56 | 76.84 | 86.88 | 76.64 |
| | | 40 | 75.96 | 75.14 | 64.66 | 74.50 | 64.44 |
| | 0.67 | 0 | 93.42 | 93.64 | 97.10 | 78.56 | 71.02 |
| | | 20 | — | 89.08 | 97.04 | 71.10 | 69.60 |
| | | 40 | — | 88.18 | 96.70 | 66.20 | 71.60 |

Note: Censoring rates of 20% and 40% correspond to the distributions of Uniform[0,5.2] and Uniform[0,2.4], respectively, for the exponential of the censoring time.

One remarkable feature of the proposed procedure is its minimal data requirement, only the uncensored lifetime measure of interest in addition to the standard survival data. This is appealing in terms of general applicability of this proposal. However, concern might arise on its efficiency when additional information becomes available. For instance, in the SWOG study the cost accumulation process is also observed at discrete follow-up times. In this regard, one

should recognize that the potential efficiency gain would be limited in situations with strong association between the lifetime cost and survival time; see Huang and Louis (1998). Moreover, further assumptions on, say the cost accumulation process, might be necessary in order to take advantage of the additional data. Nevertheless, this issue merits further and in-depth investigation.

While we have focused our discussion on the two-sample problem, the same idea naturally extends to other test problems. In particular, consider the $k$-sample problem with $k \geq 2$. Under the accelerated failure time model, there are $k-1$ parameters determining the differences among these samples for survival time. One can construct $k-1$ weighted log-rank estimating functions for survival time and, with the marked processes in place of the counting processes, obtain another $k-1$ estimating functions. The very same approach to the two-sample problem can then apply to testing for either global or trend difference on the mark among the $k$ samples.

Developing regression analysis for lifetime utility or cost is a current research topic. Due to the very same identifiability issue as discussed, existing regression models (e.g., Lin (2000a, 2000b)) do not render a direct interpretation on the lifetime measure of interest, particularly in the practical situation that the covariates of concern may also impact survival time. The results of this paper suggest that jointly modeling the lifetime outcome and survival time might be a promising approach.

**Acknowledgement**

Partial support was provided by funds from the US National Institutes of Health. The authors gratefully acknowledge that insightful and constructive comments from the reviewers have led to improvement in the presentation.

## Appendix. Derivation of Optimal Weight Functions and Efficacy

For the test based on $\Phi$ when $\beta$ is known, we first optimize $\mathrm{eff}(\Phi; W_0, W_1)$ over $W_0$ for fixed $W_1$. Thus maximize

$$\varepsilon_{12}^2/\varepsilon_{11} = \frac{\left\{\int_{-\infty}^{\tau} W_0(t,\beta)W_1(t,\beta)\lambda_1(t,\beta)B(t)\,dt\right\}^2}{\int_{-\infty}^{\tau} W_0(t,\beta)^2\lambda_0(t,\beta)B(t)\,dt}.$$

By the Cauchy–Schwartz inequality,

$$\varepsilon_{12}^2 \leq \int_{-\infty}^{\tau} W_0(t,\beta)^2\lambda_0(t,\beta)B(t)\,dt \int_{-\infty}^{\tau} W_1(t,\beta)^2\lambda_1(t,\beta)^2\lambda_0(t,\beta)^{-1}B(t)\,dt,$$

where the equality holds if and only if $W_0(t,\beta)\lambda_0(t) \propto W_1(t,\beta)\lambda_1(t)$. Therefore, $W_{0\,\mathrm{opt}|W_1}(t,\beta) \propto \lambda_0(t)^{-1}\lambda_1(t)W_1(t,\beta)$. Now, we have

$$\mathrm{eff}(\Phi; W_{0\,\mathrm{opt}|W_1}, W_1) = \frac{\left\{\int_{-\infty}^{\tau} W_1(t,\beta)\delta(t)\lambda_0(t)\,dt\right\}^2}{\int_{-\infty}^{\tau} W_1(t,\beta)^2\sigma(t)^2\lambda_0(t)B(t)\,dt}.$$

Again, by the Cauchy–Schwartz inequality,

$$\begin{aligned}&\left\{\int_{-\infty}^{\tau} W_1(t,\beta)\delta(t)\lambda_0(t)\,dt\right\}^2\\ &\le \int_{-\infty}^{\tau} W_1(t,\beta)^2\sigma(t)^2\lambda_0(t)B(t)\,dt \int_{-\infty}^{\tau} \delta(t)^2\sigma(t)^{-2}\lambda_0(t)B(t)^{-1}\,dt,\end{aligned}$$

where the equality holds if and only if $W_1(t,\beta) \propto \delta(t)\sigma(t)^{-2}B(t)^{-1}$. Combining the foregoing, we obtain (4.1).

For the test based on $\Psi$ when $\beta$ is unknown, rewrite (3.4) as

$$\begin{aligned}\mathrm{var}\{n_\bullet^{1/2}\xi_1(\widehat{\beta}_0)\} = \int_{-\infty}^{\tau} [\{r_0^{-1}r_1 W_0(t,\beta)\lambda_0(t)^{1/2} - W_1(t,\beta)\lambda_1(t)\lambda_0(t)^{-1/2}\}^2\\ + W_1(t,\beta)^2\sigma(t)^2\lambda_0(t)]B(t)\,dt.\end{aligned}$$

In the general situation, the first term in the integrand does not vanish for any choices of $W_0$ and $W_1$. Under the special independence scenario, the first term is 0 when and only when $W_0(t) \propto W_1(t)$. In this case, by applying the Cauchy–Schwartz inequality we further obtain that the efficacy is maximized if and only if $W_0(t) \propto W_1(t) \propto B(t)^{-1}$.

## References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Bang, H. and Tsiatis, A. A. (2000). Estimating medical costs with censored data. *Biometrika* **87**, 329-343.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74**, 549-555.

Gill, R. D. and Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika* **74**, 289-300.

Glasziou, P. P., Simes, R. J. and Gelber, R. D. (1990). Quality adjusted survival analysis. *Statist. Med.* **9**, 1259-1276.

Huang, Y. (1999). The two-sample problem with induced dependent censorship. *Biometrics* **55**, 1108-1113.

Huang, Y. and Louis, T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika* **85**, 785-798.

Huang, Y. and Louis, T. A. (1999). Expressing estimators of expected quality adjusted survival as functions of Nelson−Aalen estimators. *Lifetime Data Anal.* **5**, 199-212.

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

Kelly, K., Crowley, J., Bunn, P. A., Jr., Presant, C. A., Grevstad, P. K., Moinpour, C. M., Ramsey, S. D., Wozniak, A. J., Weiss, G. R., Moore, D. F., Israel, V. K., Livingston, R. B. and Gandara, D. R. (2001). Randomized phase III trial of Paclitaxel plus Carboplatin versus Vinorelbine plus Cisplatin in the treatment of patients with advanced non-small-cell lung cancer: A Southwest Oncology Group Trial. *J. Clin. Oncol.* **19**, 3210-3218.

Lai, T. L. and Ying, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Mult. Anal.* **27**, 334-358.

Lin, D. Y. (2000a). Linear regression analysis of censored medical costs. *Biostatistics* **1**, 35-47.

Lin, D. Y. (2000b). Proportional means regression for censored medical costs. *Biometrics* **56**, 775-778.

Lin, D. Y., Feuer, E. J., Etzioni, R. and Wax Y. (1997). Estimating medical costs from incomplete follow-up data. *Biometrics* **53**, 419-434.

Louis, T. A. (1981). Nonparametric analysis of an accelerated failure time model. *Biometrika* **68**, 381-390.

Wei, L. J. and Gail, M. H. (1983). Nonparametric estimation for a scale-change with censored observations. *J. Am. Statist. Assoc.* **78**, 382-388.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-851.

Yang, S. (1998). Some scale estimators and lack-of-fit tests for the censored two sample accelerated life model. *Biometrics* **54**, 1040-1052. Correction, **56**, 324.

Zhao, H. and Tsiatis, A. A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* **84**, 339-348.

Zhao, H. and Tsiatis, A. A. (2000). Estimating mean quality adjusted lifetime with censored data. *Sankhya Ser. B* **62**, 175-188.

Zhou, X.-H., Melfi, C. A. and Hui, S. L. (1997). Methods for comparison of cost data. *Ann. Intern. Med.* **127**, 752-756.

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N, MP-1195, P.O. Box 19024, Seattle, WA 98109-1024, U.S.A.

E-mail: eugene@fhcrc.org

Department of Public Health Sciences, Wake Forest University, School of Mdicine, Mecical Center Blvd., Winston−Salem, NC 27157-1063, U.S.A.

E-mail: llovato@wfubmc.edu

(Received September 2000; accepted October 2001)