

A CLASS OF WEIGHTED LOG-RANK TESTS FOR SURVIVAL DATA WHEN THE EVENT IS RARE

STEVEN BUYSKE, RICHARD FAGERSTROM, AND ZHILIANG YING

SUMMARY. In many epidemiological and medical follow-up studies a majority of study subjects do not experience the event of interest during the follow-up period. An important example is the on-going prostate, lung, colorectal, and ovarian cancer screening trial of the National Cancer Institute. In such a situation, the widely used G^ρ family of weighted log-rank statistics essentially reduces to the special case of the (unweighted) log-rank statistics. We propose a simple modification to the G^ρ family that adapts to survival data with rare events, a concept which we formulate in terms of a small number of events at the study endpoint relative to the sample size. The usual asymptotic properties, including convergence in distribution of the standardized statistics to the standard normal, are obtained under the rare event formulation. Semiparametric transformation models forming sequences of contiguous alternatives are considered and, for each ρ , a specific such model is identified so that the corresponding modified G^ρ statistic is asymptotically efficient. Simulation studies show that the proposed statistics do behave differently from the original G^ρ statistics when the event rate during the study period is low and the former could lead to a substantial efficiency gain over the latter. Extensions to the $G^{\rho,\gamma}$ family and to the regression problem are also given.

1. INTRODUCTION

Weighted log-rank statistics have been widely used in medical and epidemiological follow-up studies to test for treatment or exposure effects on survival. The original (unweighted) log-rank test was proposed by Mantel and Haenszel (1959); see also Mantel (1966). It coincides with the partial likelihood score of Cox (1972) for the proportional hazards regression model. On the other hand, efforts to extend the Wilcoxon rank-sum test to censored failure time data led to the work of Gehan (1965), Peto and Peto (1972), Tarone and Ware (1977), and Prentice (1978) among others. Both the log-rank statistic and the extensions of the Wilcoxon statistic can be incorporated into the class of weighted log-rank statistics, whose theoretic properties can be derived easily via martingale theory; cf. Gill (1980), Fleming and Harrington (1991), Andersen, Borgan, Gill and Keiding (1993).

A particularly useful sub-family within the class of weighted log-rank statistics is the so-called G^ρ family, proposed by Fleming and Harrington (1981) and Harrington and

Key words and phrases. G -rho family, censoring, Peto-Prentice test, cancer screening.

Fleming (1982). For this sub-family, the weight function is chosen to be the Kaplan-Meier estimate of the survival function raised to a specified power. The family contains as special cases the log-rank statistic ($\rho = 0$) and the Peto-Prentice extension of the Wilcoxon statistic ($\rho = 1$), and bridges the two widely used statistics in a smooth and natural way. When the ratio of the two hazard functions for the treatment and the control groups is constant over time, the log-rank test is asymptotically efficient, whereas if the ratio decreases, then a proper choice from the G^ρ family, with $\rho > 0$, is likely to produce a test more efficient than the log-rank. The popularity of the G^ρ family in testing survival difference is manifested by the implementation of the family in major software packages, including SAS, Splus, and BMDP.

In some medical and epidemiological follow-up studies, the event rate during the study period can be very low. This will generally be the case for cancer screening and prevention trials. An important such example, which is in fact the main motivation for the present research, is the on-going prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial conducted by the National Cancer Institute; cf. Gohagan, Prorok, Kramer, and Cornett (1994). This is a multicenter, randomized trial of enormous scale which includes 74,000 men and 74,000 women. The prostate aspect of the trial, which includes 74,000 men from 55 to 74 years old who are allocated to either the treatment (screening) group or the control group, is aimed at determining whether screening is of substantial benefit in terms of reducing mortality. The study is designed to have one-sided type I error rate of 0.05 and power of 0.90 for a 20% mortality reduction with a ten year follow-up.

With the possible exception of non-melanoma skin cancer, prostate cancer has the highest incidence rate among all cancers in men. It occurs mostly in men 50 years or older. Although screening can lead to early detection of the cancer, its benefit, including mortality reduction, is still largely unknown (Gohagan et al. 1994). This is due primarily to the fact that prostate cancer mostly attacks older men and progresses slowly. According to Hanks and Scardino (1996), “a third of men over age 50 harbor some form of the cancer, but only between 6 and 10 percent will acquire the type likely to lead to death or disability. And only about 3 percent eventually die of it.”

As will be demonstrated in the next section, the G^ρ family of weighted log-rank tests is essentially the same as the (unweighted) log-rank test when the event rate is low, such as in the PLCO trial. Thus, the G^ρ family of tests, although simple and elegant in its form and easy to use, does not fulfill its role to adapt, through different values of the parameter ρ , to a range of alternatives. It is well known that in many situations, properly choosing a weight function can lead to a substantial increase in efficiency over the unweighted log-rank test. The efficiency gain is especially important for follow-up studies with low event rates, which by definition require very large sample sizes. For example, a 10% efficiency gain for the PLCO trial would mean a net reduction of about 14,500 individuals, each with a multi-year follow-up.

The purpose of the present paper is to propose a modification to the G^ρ family of statistics to accommodate rare event (low event rate) survival data so that the versatility

of the G^ρ family in the usual situations can be maintained. The modified G^ρ family of statistics has similar simplicity in its form and it is intuitively easy to identify situations under which a particular member should be preferred. In order to provide a formal treatment, a precise mathematical definition for rare event survival data will be given. Asymptotic distributions of the modified G^ρ -statistics are then derived under the null hypothesis of no treatment difference and under a sequence of contiguous alternatives. These results enable us to evaluate their asymptotic efficiencies. Differential equations, similar to those given in Fleming and Harrington (1991, Chapter 7), are derived to identify families of distributions for the underlying local alternatives so that a particular member of the G^ρ family is asymptotically optimal.

The rest of the paper is organized as follows. Section 2 reviews the G^ρ family of weighted log-rank statistics and introduces a new modification that handles rare event data. A precise mathematical formulation for a rare event is given in Section 3, where the main theoretic properties, including consistency, asymptotic normality and asymptotic efficiency, are obtained under such a formulation. Simulation results comparing the proposed modification of the G^ρ family with the original versions are summarized in Section 4. Section 5 applies the proposed test to data from the BHAT medical study. Section 6 contains extensions to a more general $G^{\rho,\gamma}$ family (Fleming and Harrington, 1991) and to regression data. Section 7 contains some concluding remarks. Most technical derivations and proofs are put together in the Appendix.

2. A MODIFIED G^ρ FAMILY OF TEST STATISTICS

We will consider in this section two-sample comparison tests for follow-up studies. In this connection, let $T_{1i}, C_{1i}, i = 1, \dots, n_1$ be the failure and censoring times, respectively, from the control group and $T_{2i}, C_{2i}, i = 1, \dots, n_2$ be the corresponding times from the treatment group. For simplicity, simultaneous entry (at time 0) is assumed. Let τ be the time that the results are collected or the time at the end of the study. Thus observations consist of $\tilde{T}_{ki} = \min\{T_{ki}, C_{ki}, \tau\}$, $\delta_{ki} = I(T_{ki} \leq \min\{C_{ki}, \tau\})$, $i = 1, \dots, n_k, k = 1, 2$.

Define

$$\begin{aligned}
N_{ki}(t) &= \delta_{ki} I(\tilde{T}_{ki} \leq t) \\
Y_{ki}(t) &= I(\tilde{T}_{ki} \geq t) \\
N_{k\cdot}(t) &= \sum_{i=1}^{n_k} N_{ki}(t) \\
Y_{k\cdot}(t) &= \sum_{i=1}^{n_k} Y_{ki}(t) \\
N_{\cdot\cdot}(t) &= \sum_{k=1}^2 N_{k\cdot}(t) \\
Y_{\cdot\cdot}(t) &= \sum_{k=1}^2 Y_{k\cdot}(t).
\end{aligned}$$

We will use $\hat{S}(t)$ to denote the Kaplan-Meier estimate from the pooled data, i.e.,

$$\hat{S}(t) = \prod_{s \leq t} \left(1 - \frac{\Delta N_{\cdot\cdot}(s)}{Y_{\cdot\cdot}(s)} \right),$$

where $\Delta N_{\cdot\cdot}(s) = N_{\cdot\cdot}(s) - N_{\cdot\cdot}(s-)$. We estimate $F(t)$ with $\hat{F}(t) = 1 - \hat{S}(t)$.

It will be assumed throughout the paper that $\{T_{ki}, i = 1, \dots, n_k\}$, $\{C_{ki}, i = 1, \dots, n_k\}$, $k = 1, 2$ are sequences of independent and identically distributed (iid) random variables and they are independent of one another. Denote by F_1 and F_2 the failure time distributions for the first and the second sample, i.e., $F_k(t) = P(T_{ki} \leq t)$, $k = 1, 2$. For simplicity, we assume that the F_k are continuous. The corresponding survival and cumulative hazard functions will be denoted by $S_k(t) = 1 - F_k(t)$ and $\Lambda_k(t) = -\log S_k(t)$, $k = 1, 2$. Let $G_k(t) = P(C_{ki} \leq t)$, $k = 1, 2$, be the censoring distributions, which need not be continuous, and $\bar{G}_k(t) = 1 - G_k(t-)$.

To test equality of the two distributions, $F_1 = F_2$, Fleming and Harrington (1981) and Harrington and Fleming (1982) proposed the following class of statistics, known as the G^ρ family:

$$U_\rho = \int_0^\tau \hat{S}^\rho(t-) \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{\cdot\cdot}(t)} \left[\frac{dN_{1\cdot}(t)}{Y_{1\cdot}(t)} - \frac{dN_{2\cdot}(t)}{Y_{2\cdot}(t)} \right], \quad (2.1)$$

where $0 \leq \rho < \infty$ is a fixed constant. The family is included in a more general class, known as the K -class (Gill, 1980), of weighted log-rank statistics. When $\rho = 0$, U_ρ is the well-known log-rank statistic, while for $\rho = 1$, U_ρ becomes the Peto-Prentice extension of the Wilcoxon statistic.

Harrington and Fleming (1982) found semiparametric transformation models that generate sequences of local alternatives for which the G^ρ statistics are asymptotically

efficient. Specifically, define H_ρ by

$$H_\rho(t) = \begin{cases} \exp(-e^t), & \text{if } \rho = 0 \\ (1 + \rho e^t)^{-1/\rho} & \text{if } \rho > 0. \end{cases} \quad (2.2)$$

For each $\rho \geq 0$, the semiparametric model they used is

$$S_1(t) = H_\rho(g(t)), \quad S_2(t, \theta) = H_\rho(g(t) + \theta), \quad (2.3)$$

where g is any (unspecified) monotone increasing and smooth function, and θ is the parameter indicating treatment difference. The model has an equivalent characterization:

$$g(T_{ki}) = -z_k \theta + \epsilon_{ki},$$

where $z_1 = 0$, $z_2 = 1$, and ϵ_{ki} are iid random variables with survival function H_ρ . Note that for $\rho = 0$, the ϵ_{ki} follow the extreme value distribution while for $\rho = 1$ they follow the logistic distribution. Harrington and Fleming found that the one-sided test of rejecting $H_0 : F_1 = F_2$, (i.e. $\theta = 0$) whenever

$$\frac{U_\rho}{\sqrt{\langle U_\rho \rangle}} > z_\alpha$$

is asymptotically efficient against the local alternatives $\theta = \theta_n = b/\sqrt{n}$ for a fixed $b > 0$. Here z_α is the $(1 - \alpha) \times 100$ th percentile of the standard normal and

$$\langle U_\rho \rangle = \int_0^\tau \hat{S}^{2\rho}(t-) \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{\cdot\cdot}(t)} [dN_{1\cdot}(t) + dN_{2\cdot}(t)]$$

estimates the variance of U_ρ .

When the event rate is low, the G^ρ family of weighted log-rank statistics does not have as good a range of flexibility as one would expect. This is because the low event rate means that $F(\tau) \approx 0$ or $S(\tau) \approx 1$, so that $\hat{S}^\rho(t) \approx S^\rho(t) \approx 1$ for all $t \leq \tau$ and reasonable values for ρ such as $0 \leq \rho \leq 1$. Therefore, any member in the family is essentially the same as the unweighted log-rank statistic, as the weight function changes little over the entire follow-up period $[0, \tau]$.

Obviously, the reason that the G^ρ weight function remains stationary for all values of ρ is that the survival function (and thus its Kaplan-Meier estimate) does not change much relative to its initial value. A simple modification that forces the weight function to decrease substantially is to subtract from the Kaplan-Meier estimator its value at the endpoint τ . We therefore propose the following modification to the G^ρ statistics:

$$\tilde{U}_\rho = \int_0^\tau [\hat{S}(t-) - \hat{S}(\tau-)]^\rho \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{\cdot\cdot}(t)} \left[\frac{dN_{1\cdot}(t)}{Y_{1\cdot}(t)} - \frac{dN_{2\cdot}(t)}{Y_{2\cdot}(t)} \right]. \quad (2.4)$$

Since $\hat{S}(t)$ decreases from 1 to $\hat{S}(\tau)$, the new weight function $[\hat{S}(t-) - \hat{S}(\tau-)]^\rho$, when scaled by $[1 - \hat{S}(\tau-)]^\rho$, decreases from 1 to 0 for any $\rho > 0$. A natural estimator for

the variance of \tilde{U}_ρ is

$$\hat{\sigma}_{\tilde{U}_\rho}^2 = \int_0^\tau [\hat{S}(t-) - \hat{S}(\tau-)]^{2\rho} \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{\cdot\cdot}^2(t)} [dN_{1\cdot}(t) + dN_{2\cdot}(t)]. \quad (2.5)$$

One slight technical complication of using the new weight function is that it is no longer predictable with respect to the natural σ -filtration generated by the counting processes $\{N_{ki}, Y_{ki}\}$. Thus, strictly speaking, \tilde{U}_ρ with τ replaced by $t \in [0, \tau]$, is not a martingale and the standard martingale central limit theorem (Rebolledo, 1980) cannot be applied directly to prove the asymptotic normality under suitable scaling. However, as will be shown in the Appendix, the random weight function $[\hat{S}(t-) - \hat{S}(\tau-)]^\rho$ can be replaced by its non-random limit $[S(t) - S(\tau)]^\rho$ without altering the limiting distribution of \tilde{U}_ρ under both the null and contiguous alternatives. In fact, we have the following results, which will be proved in the Appendix.

Theorem 2.1. *Suppose that for $k = 1, 2$, $n_k/n \rightarrow r_k \in (0, 1)$ and $Y_k(t)/n_k \rightarrow y_k(t)$ with $y_k(\tau) > 0$.*

1. *Under $H_0 : S_2(t) = S_1(t)$, $t \leq \tau$, we have $\tilde{U}_\rho/\hat{\sigma}_{\tilde{U}_\rho} \xrightarrow{\mathcal{L}} N(0, 1)$.*
2. *Under a sequence of contiguous alternatives, $S_2(t) = S_2^{(n)}(t) = (1+n^{-1/2}\Gamma_n(t))S_1(t)$, with $\Gamma_n(t)$ differentiable and $\Gamma_n'(t) = \gamma_n(t) \rightarrow \gamma(t)$, uniformly in $t \in [0, \tau]$, for a continuous function γ , we have $\tilde{U}_\rho/\hat{\sigma}_{\tilde{U}_\rho} \xrightarrow{\mathcal{L}} N(\mu_\rho, 1)$, where*

$$\mu_\rho = \frac{\int_0^\tau [F_1(\tau) - F_1(t)]^\rho [r_1 y_1(t) r_2 y_2(t)] [r_1 y_1(t) + r_2 y_2(t)]^{-1} \gamma(t) dt}{\left\{ \int_0^\tau [F_1(\tau) - F_1(t)]^{2\rho} [r_1 y_1(t) r_2 y_2(t)] [r_1 y_1(t) + r_2 y_2(t)]^{-1} d\Lambda_1(t) \right\}^{1/2}}. \quad (2.6)$$

When the entire survival curve is estimable, (or, equivalently, $F(\tau) = 1$), it is clear from the preceding theorem that the proposed modification is asymptotically equivalent to the original G^ρ family. The effect of the modification becomes noticeable with rare event survival data, which, as stated above, motivated the modified statistics. The next section derives asymptotic properties of the modified G^ρ statistics for such rare event survival data.

3. ASYMPTOTIC PROPERTIES OF MODIFIED G^ρ STATISTICS WITH RARE EVENT DATA

In this section, we will study asymptotic properties of the proposed modification to the G^ρ family of statistics for rare event survival data. Section 3.1 is devoted to a mathematical definition for the notion of rare event. Section 3.2 derives asymptotic distributions for the modified G^ρ statistics under both the null hypothesis and the contiguous alternatives. In doing so, we also derive the consistency of the properly scaled Kaplan-Meier estimate for the rare event survival data. We present the information on contiguous alternatives not only to match the historic development of the usual weighted log-rank statistic, but also to assess the asymptotic efficiency of the proposed

test and because the contiguous alternative for a given test can be used to determine the appropriate sample size for any given study. Finally in Section 3.3, the asymptotic efficiency of the proposed test is studied under semiparametric transformation models and some optimality results, analogous to those presented by Fleming and Harrington (1982, 1991) for their original G^ρ statistics, are given.

3.1. A mathematical formulation for rare event survival data. To study the asymptotic properties of the proposed modification to the G^ρ family under the assumption of rare events, it is necessary to provide a mathematically well-defined environment. This subsection is devoted to setting up the notion of rare event survival data in precise mathematical terms.

The notion of rare event must be relative to the sample size n . In this respect, it is natural to let the two distribution functions for the two populations go to zero over the study period $[0, \tau]$ as $n \rightarrow \infty$. Furthermore, the amount of “information” accumulated from the data must tend to infinity as $n \rightarrow \infty$, since, otherwise, the underlying true distributions(s) cannot be identified even as $n \rightarrow \infty$. To this end, we define the low event rate condition for the two populations $F_k = F_k^{(n)}$, $k = 1, 2$, to be

$$F_k^{(n)}(t) = m_n^{-1} \tilde{F}_k(t), \quad t \in [0, \tau] \quad (3.1)$$

for some $m_n \rightarrow \infty$ and increasing functions \tilde{F}_k , $k = 1, 2$. Note that the F_k naturally depend on n and thus we have written $F_k^{(n)}$ to indicate this dependency explicitly. We will also assume $n/m_n \rightarrow \infty$ to ensure that the number of events (which, intuitively speaking, is approximately proportional to the “information”) will tend to ∞ . In addition, the G_k may also depend on n and we require $G_k(t) \rightarrow \tilde{G}_k(t)$ and $1 - \tilde{G}_k(\tau) > 0$. If the censoring due to causes other than termination of study is also “rare,” then we expect $1 - \tilde{G}_k(t) \approx 1$ for all $t \leq \tau$.

From (3.1), we clearly have as $n \rightarrow \infty$,

$$m_n \Lambda_k(t) = -m_n \log(1 - F_k^{(n)}(t)) \rightarrow \tilde{F}_k(t), \quad t \in [0, \tau], \quad k = 1, 2. \quad (3.2)$$

In other words, the cumulative hazard functions of the two samples are approximated by $m_n^{-1} \tilde{F}_k(t)$, $k = 1, 2$.

We may replace (3.1) with a slightly less stringent requirement that

$$m_n F_k^{(n)}(t) \rightarrow \tilde{F}_k(t) \quad \text{as } n \rightarrow \infty. \quad (3.3)$$

It can be shown that all subsequent asymptotic results can also be developed under (3.3). For notational convenience, however, we will stick to formulation (3.1) rather than (3.3) for low event rate.

The null hypothesis, $F_1(t) = F_2(t)$, $0 \leq t \leq \tau$, is, in view of (3.1), equivalent to $\tilde{F}_1(t) = \tilde{F}_2(t)$, $0 \leq t \leq \tau$. The contiguous alternatives can also be defined in terms of \tilde{F}_k . Specifically, let $b(t)$ be a continuous function on $[0, \tau]$. Then writing $\tilde{f}_k =$

\tilde{F}'_k , $k = 1, 2$, we define the contiguous alternatives to be, in conjunction with (3.1),

$$\tilde{f}_2(t) = \tilde{f}_1(t) + \sqrt{\frac{m_n}{n}}b(t)\tilde{f}_1(t) + o\left(\sqrt{\frac{m_n}{n}}\right), \quad (3.4)$$

where $o(\sqrt{m_n/n})$ is uniform over $t \in [0, \tau]$. The scaling factor reflects that the Fisher information under (3.1) is of order n/m_n .

Integrating (3.4) over t , we get

$$\tilde{F}_2(t) = \tilde{F}_1(t) + \sqrt{\frac{m_n}{n}}B(t) + o\left(\sqrt{\frac{m_n}{n}}\right),$$

where $B(t) = \int_0^t b(s) d\tilde{F}_1(s)$. Furthermore,

$$\begin{aligned} \Lambda_2(t) - \Lambda_1(t) &= -\log \frac{1 - F_2(t)}{1 - F_1(t)} \\ &= -\log \left[1 - \frac{(m_n n)^{-1/2}(B(t) + o(1))}{1 - m_n^{-1}\tilde{F}_1(t)} \right] \\ &= (m_n n)^{-1/2}B(t) + o((m_n n)^{-1/2}). \end{aligned}$$

Formally through differentiation, but rigorously via (3.1), we can likewise get

$$\begin{aligned} \lambda_2(t) - \lambda_1(t) &= (m_n n)^{-1/2}b(t)\tilde{f}_1(t) + o((m_n n)^{-1/2}) \\ &= (m_n n)^{-1/2}b(t)\tilde{\lambda}_1(t) + o((m_n n)^{-1/2}). \end{aligned} \quad (3.5)$$

Here $\tilde{\lambda}_1(t) = \lim_{n \rightarrow \infty} m_n \lambda_1(t)$, which is the same as $\tilde{f}_1(t)$ by (3.2).

3.2. Asymptotic distributions under the null hypothesis and the contiguous alternatives. We will study asymptotic distributions of the proposed modification of the G^ρ family of statistics under both the null hypothesis and the contiguous alternatives under the rare event assumption as described in Section 3.1. The results will allow us to use normal approximations to get asymptotic size and power of the corresponding tests. Note that Theorem 2.1 is not formulated for rare events. Because the statistics involve the Kaplan-Meier estimate of the survival function, it is necessary to derive the consistency of this estimate first.

Recall that \hat{S} is the Kaplan-Meier estimate from the pooled sample, and that $\hat{F} = 1 - \hat{S}$. Let \hat{S}_k be the Kaplan-Meier estimate from the k -th sample, $k = 1, 2$, and $\hat{F}_k = 1 - \hat{S}_k$. In view of the rare event rate setup, $F_k^{(n)} = \tilde{F}_k/m_n$, it is natural to view consistency as $m_n \hat{F}_k$ approaching \tilde{F}_k .

Theorem 3.1. *Under the low event rate assumption (3.1), for $k = 1, 2$ and $t \in [0, \tau]$, $m_n \hat{F}_k(t) \xrightarrow{\mathcal{P}} \tilde{F}_k(t)$ and $m_n \hat{\Lambda}_k(t) \xrightarrow{\mathcal{P}} \tilde{F}_k(t)$ as $n_k \rightarrow \infty$, where $\hat{\Lambda}_k(t) = \int_0^t dN_{k \cdot}(s)/Y_{k \cdot}(s)$ is the Nelson-Aalen estimate of the cumulative hazard function. Furthermore, under either the null hypothesis $\tilde{F}_2 = \tilde{F}_1$ or the contiguous alternatives (3.4), $m_n \hat{F}_k(t) \xrightarrow{\mathcal{P}} \tilde{F}_1(t)$, $k = 1, 2$ and $m_n \hat{F}(t) \xrightarrow{\mathcal{P}} \tilde{F}_1(t)$ as $n \rightarrow \infty$.*

The proof of Theorem 3.1 involves standard counting process-martingale techniques and is given in the Appendix. From Theorem 3.1, we see that the weight function $[\hat{F}_k(\tau) - \hat{F}_k(t)]^\rho$ is approximately proportional to $[\tilde{F}_1(\tau) - \tilde{F}_1(t)]^\rho$, which is crucial to the proof of Theorem 3.2.

Let

$$\tilde{\mu}_\rho = - \frac{\int_0^\tau [\tilde{F}_1(\tau) - \tilde{F}_1(t)]^\rho b(t) d\mu(t)}{\left\{ \int_0^\tau [\tilde{F}_1(\tau) - \tilde{F}_1(t)]^{2\rho} d\mu(t) \right\}^{1/2}}, \quad (3.6)$$

where $d\mu(t) = r_1 \bar{G}_1(t) r_2 \bar{G}_2(t) [r_1 \bar{G}_1(t) + r_2 \bar{G}_2(t)]^{-1} d\tilde{F}_1(t)$. For the modified G^ρ statistics \tilde{U}_ρ defined by (2.4), and its variance estimator $\hat{\sigma}_{\tilde{U}_\rho}^2$ given by (2.5), we have the following theorem giving asymptotic type I error and power.

Theorem 3.2. *Suppose $n_k/n \rightarrow r_k \in (0, 1)$, $k \in (0, 1)$. Under the rare event formulation of Section 3.1,*

$$\tilde{U}_\rho / \hat{\sigma}_{\tilde{U}_\rho} \xrightarrow{\mathcal{L}} N(0, 1)$$

if the null hypothesis $\tilde{F}_1 = \tilde{F}_2$ holds; and

$$\tilde{U}_\rho / \hat{\sigma}_{\tilde{U}_\rho} \xrightarrow{\mathcal{L}} N(\tilde{\mu}_\rho, 1)$$

if the contiguous alternative (3.4) holds.

Suppose $\tilde{\mu}_\rho > 0$. Then it follows from the preceding theorem that a one-sided level α test rejects H_0 whenever $\tilde{U}_\rho / \hat{\sigma}_{\tilde{U}_\rho} > z_\alpha$, where z_α is the $(1 - \alpha) \times 100$ th percentile of the standard normal distribution. The asymptotic power is then $1 - \Phi(z_\alpha - \tilde{\mu}_\rho)$.

The weight function in \tilde{U}_ρ is essentially $[\tilde{F}_1(t) - \tilde{F}_1(\tau)]^\rho$, in view of Theorem 3.1 and the fact that its standardized version is scale invariant. Theorem 3.2 can easily be extended to cover general weight functions. Suppose that $K_n(t)$ is a sequence of weight functions, converging to a deterministic limit $K(t)$. Let \tilde{U}_K and $\hat{\sigma}_{\tilde{U}_K}^2$ be the same as \tilde{U}_ρ and $\hat{\sigma}_{\tilde{U}_\rho}^2$ except with $[\tilde{F}_1(t) - \tilde{F}_1(\tau)]^\rho$ replaced by K_n . Then $\tilde{U}_K / \hat{\sigma}_{\tilde{U}_K}$ converges in distribution to $N(0, 1)$ under the null hypothesis and to $N(\mu_K, 1)$ under contiguous alternatives as defined by (3.4). Here

$$\mu_K = \frac{\int_0^\tau K(t) b(t) d\mu(t)}{\left(\int_0^\tau K^2(t) d\mu(t) \right)^{1/2}}. \quad (3.7)$$

3.3. Asymptotic efficiency. We will now discuss the asymptotic efficiency of the modified G^ρ tests. This is closely related to the asymptotic mean, defined by (3.6) or (3.7), of the standardized test statistic under contiguous alternatives. In fact, $\tilde{\mu}_\rho$ of (3.6), or μ_K of (3.7) is the asymptotic efficacy; cf. Randles and Wolfe (1979, p. 149).

Among the general K -class of weighted log-rank statistics described in Section 3.2, the most efficient one maximizes μ_K^2 . In view of the definition of μ_K in (3.7), we can apply the Cauchy-Schwartz inequality to get

$$\mu_K^2 \leq \int_0^\tau b^2(t) d\mu(t).$$

Since it does not involve K , the right-hand side of the above inequality serves as an upper bound. The upper bound can be attained by taking $K(t) = \text{constant} \times b(t)$, which corresponds to an optimal choice of the weight function. Therefore, the asymptotic efficiency of a G^ρ statistic relative to the optimal weighted log-rank statistic is

$$\frac{\{\int_0^\tau [\tilde{F}(\tau) - \tilde{F}(t)]^\rho b(t) d\mu(t)\}^2}{\int_0^\tau [\tilde{F}(\tau) - \tilde{F}(t)]^{2\rho} d\mu(t) \int_0^\tau b^2(t) d\mu(t)}. \quad (3.8)$$

Following Fleming and Harrington (1991), we define a semiparametric transformation model by

$$\begin{aligned} \tilde{F}_1(t) &= \tilde{H}(g(t)) \\ \tilde{F}_2(t) &= \tilde{H}(g(t) + \theta) \end{aligned} \quad (3.9)$$

for some monotone increasing functions \tilde{H} and g . The contiguous alternative requirement is to set $\theta = c_1 \sqrt{m_n/n}$ for a constant c_1 . Thus, under (3.9) with $\theta = c_1 \sqrt{m_n/n}$,

$$\tilde{f}_2(t) = \tilde{f}_1(t) + c_1 \frac{\tilde{H}''(g(t))}{\tilde{H}'(g(t))} \tilde{f}_1(t) \sqrt{m_n/n} + o(\sqrt{m_n/n}). \quad (3.10)$$

Using the notation of (3.4), $b(t) = c_1 \tilde{H}''(g(t))/\tilde{H}'(g(t))$. We know that the optimal weight function is $b(t)$. Therefore, in order for $[\hat{F}(\tau) - \hat{F}(t)]^\rho$ to be asymptotically optimal, it is necessary and sufficient that $[\tilde{F}(\tau) - \tilde{F}(t)]^\rho = cb(t)$ for some constant c , or

$$[\tilde{H}(g(\tau)) - \tilde{H}(g(t))]^\rho = c \frac{\tilde{H}''(g(t))}{\tilde{H}'(g(t))}. \quad (3.11)$$

Because we can introduce a new variable $u = g(t)$ to eliminate g , equation (3.11) is essentially a second order ordinary differential equation. In the Appendix, we will solve it to get the following result.

Theorem 3.3. *Under the rare event setup as defined by (3.1), U_ρ is asymptotically optimal provided that (3.9) and (3.10) are satisfied with*

$$\tilde{H}(u) = a_2 - L_{\rho, a_2}^{-1}(a_1(u^* - u)),$$

for some $a_2 > 0$ and $a_1 > 0$, where $u^* = g(\tau)$ and

$$L_{\rho, a_2}(v) = \int_0^v \frac{dx}{a_2^{\rho+1} - x^{\rho+1}} \quad 0 \leq v < a_2.$$

Note that $L_{\rho, a_2}(0) = 0$ and $L_{\rho, a_2}(a_2) = \infty$ so that $\tilde{H}(u^*) = a_2$ and $\tilde{H}(-\infty) = 0$. In particular, for $\rho = 1$, the optimal function is

$$\tilde{H}(u) = \frac{a_2^*}{1 + \exp(a_1^*(u^* - u))}, \quad -\infty < u \leq u^*,$$

where $a_1^* > 0$ and $a_2^* > 0$, whereas for $\rho = 0$, the optimal function is

$$\tilde{H}(u) = \tilde{a}_2 \exp(\tilde{a}_1 u), \quad -\infty < u \leq u^*,$$

where $\tilde{a}_1 > 0$ and $\tilde{a}_2 > 0$.

Remark. If we recall that $\tilde{F}_k(t)$ is the limit of the scaled cumulative hazard function $m_n \Lambda_k(t)$, as shown in equation (3.2), we can easily interpret this theorem. For example, for $\rho = 0$, when the test is asymptotically optimal we have the ratio of the hazards equal to

$$\frac{\lambda_2(t)}{\lambda_1(t)} = \frac{d\tilde{H}(g(t) + \theta)/dt}{d\tilde{H}(g(t))/dt} = \exp(\tilde{a}_1 \theta);$$

that is, the hazards are proportional. For $\rho > 0$, we have

$$\begin{aligned} \frac{\lambda_2(t)}{\lambda_1(t)} &= \frac{d\tilde{H}(g(t) + \theta)/dt}{d\tilde{H}(g(t))/dt} \\ &= \frac{\tilde{a}_2^{\rho+1} - \left(L_{\rho, a_2}^{-1}(a_1(g(\tau) - g(t) - \theta))\right)^{\rho+1}}{\tilde{a}_2^{\rho+1} - \left(L_{\rho, a_2}^{-1}(a_1(g(\tau) - g(t)))\right)^{\rho+1}} \\ &= \frac{\tilde{a}_2^{\rho+1} - (\tilde{a}_2 - \tilde{F}_2(t))^{\rho+1}}{\tilde{a}_2^{\rho+1} - (\tilde{a}_2 - \tilde{F}_1(t))^{\rho+1}}. \end{aligned}$$

By looking at the derivative of the next-to-last expression, we find that the ratio of hazards is increasing when θ is negative (and so the ratio is less than one) and decreasing when θ is positive (when the ratio is greater than one). In other words, when the test is asymptotically optimal the hazard ratio moves towards one.

It is also interesting to compare the preceding theorem with the corresponding findings of Harrington and Fleming (1982) for the non-rare situation as specified by (2.2) and (2.3). For $\rho = 1$, both H_ρ and \tilde{H} are logistic functions which have the same simple form. But for $\rho \neq 1$, they are completely different functions. In particular, \tilde{H} in general does not have explicit form whereas H_ρ always has as simple and explicit form as seen from (2.2).

4. SIMULATION RESULTS

		Sample size=60,000							
		$\tilde{F}_k \sim \log \text{logistic}$				Linear hazards			
		$m_n = 100$		$m_n = 20$		$m_n = 100$		$m_n = 20$	
		150 expected control events		750 expected control events		150 expected control events		750 expected control events	
		type I	power	type I	power	type I	power	type I	power
Proposed ($\rho = 1$)		0.050	0.51	0.051	0.98	0.049	0.51	0.049	0.98
Log-rank		0.050	0.45	0.052	0.96	0.049	0.31	0.049	0.84
Peto-Prentice		0.050	0.45	0.052	0.96	0.050	0.31	0.050	0.84
		Sample size=20,000							
		$\tilde{F}_k \sim \log \text{logistic}$				Linear hazards			
		$m_n = 100$		$m_n = 20$		$m_n = 100$		$m_n = 20$	
		50 expected control events		250 expected control events		50 expected control events		250 expected control events	
		type I	power	type I	power	type I	power	type I	power
Proposed ($\rho = 1$)		0.050	0.25	0.051	0.70	0.050	0.25	0.048	0.70
Log-rank		0.050	0.22	0.050	0.63	0.051	0.16	0.050	0.45
Peto-Prentice		0.050	0.22	0.050	0.63	0.050	0.16	0.050	0.46
		Sample size=2,000							
		$F_k \sim \log \text{logistic}$				Linear hazards			
		$m_n = 2.5$		$m_n = 1.67$		$m_n = 2.5$		$m_n = 1.67$	
		360 expected control events		539 expected control events		360 expected control events		539 expected control events	
		type I	power	type I	power	type I	power	type I	power
Proposed ($\rho = 1$)		0.049	0.59	0.050	0.82	0.049	0.99	0.050	1.00
Log-rank		0.049	0.23	0.049	0.43	0.042	0.36	0.045	0.76
Peto-Prentice		0.050	0.32	0.049	0.63	0.044	0.64	0.047	0.98

TABLE 1. Summary of Simulation Results

For some comparisons of the modified G^ρ family with the original G^ρ family, we ran six sets of simulations for each of two underlying distributions. One distribution was the log-logistic. For the second distribution, we compared the exponential distribution, which has a constant hazard function λ_0 , against the alternative distribution with hazard $\lambda_1(t) = \lambda_0(.47t/\tau + .40)$. In both cases, the parameters of the alternative were chosen so that average mortality over the study period $[0, \tau]$ under the alternative would be 80% of the control distribution. Here the average mortality over $[0, \tau]$ is defined to be $\tau^{-1} \int_0^\tau F_k(t) dt$. The truncation time τ was selected so that with $m_n = 1$ there would be 10% censoring due to truncation. In all cases, the censoring due to truncation in the treatment group is 96% that of the control group. For each distribution, we used sample sizes of 60,000, 20,000, and 2,000 equally divided between the control and treatment arms. With the notation of (3.1) we used 2 different levels of m_n for each sample size. The table shows the expected number of events in the control group for each level of m_n . Finally, at each level of m_n we ran 100,000 simulations under the null and under the alternative hypothesis. The results are summarized in Table 1. Results for one-sided testing are shown; the results for two-sided testing are comparable.

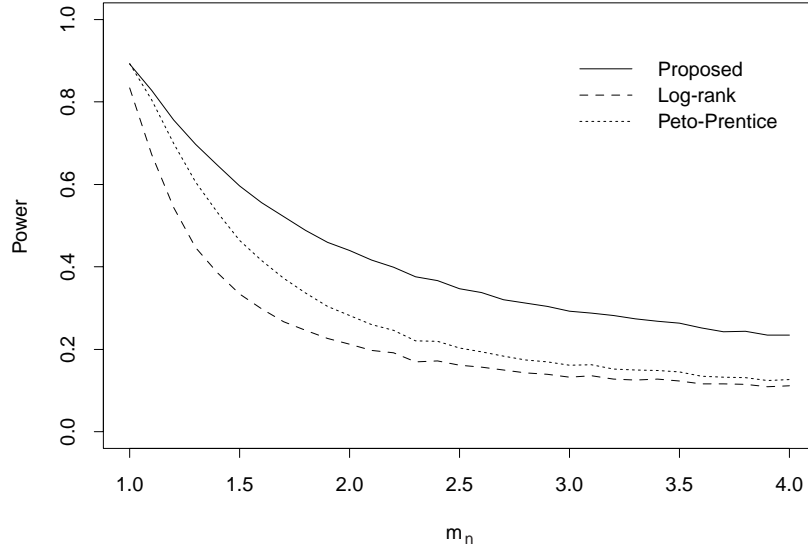


FIGURE 1. Power functions from a simulation study with sample size 200 and varying m_n . The distribution of survival times is log-logistic with parameters chosen to given power of .90 for the Prentice-Peto ($\rho = 1$) test with 10% censoring at $m_n = 1$.

Three aspects of the table should be apparent: the log-rank and Peto-Prentice tests do not differ appreciably for rare events; the proposed test has type I error rates comparable to that of the G^ρ family; and the power of the proposed test exceeds that of the original tests in all the cases simulated, in some cases by a great deal.

To indicate the performance of the proposed test for small samples, we performed a simulation study with sample size 200. The distribution here is again log-logistic, but this time the size of the effect was chosen so that the power of the test at $m_n = 1$ would be 0.90 with 10% censoring. Figure 1 shows the power of the proposed and usual G^ρ tests with $\rho = 1$, as well as the power of the log-rank test. The power is based on 30,000 simulations for each value of m_n . With $m_n = 1$, the proposed and usual G^ρ tests have almost identical power (indeed, the tests are almost identical, since $\hat{S}(\tau-)$ is close to zero), but as m_n increases the power of the usual G^ρ tests falls to close to that of the log-rank test. This simulation suggests that the proposed test may be preferable to the usual G^ρ test even when only a small proportion of the events are censored.

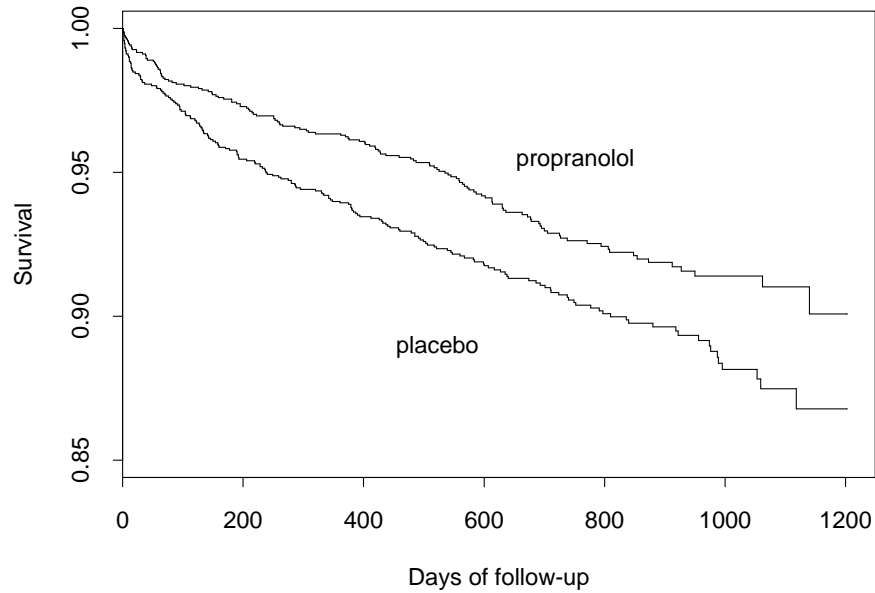


FIGURE 2. The Kaplan-Meier curves for the BHAT study.

5. THE β -BLOCKER HEART ATTACK TRIAL

In this section we apply our proposed test to the data from the β -Blocker Heart Attack Trial (BHAT), which established the value of β -blockers in reducing total mortality of survivors of myocardial infarction (β -Blocker Heart Attack Trial Research Group, 1982). The trial involved 3837 patients randomized to a placebo or propranolol. The endpoint was death. Eleven patients with unknown mortality status are excluded here. The event of interest, death, occurred in just 8.5% of the patients. Figure 2 shows the survival curves for the control and treatment groups.

The original BHAT Research Group analyzed the data using the log-rank statistic. In fact, because of the high survival rate, the log-rank and the Peto-Prentice (that is, $\rho = 1$) statistic differ only slightly. For the full set of data, the p -value for the log-rank test is $p = 0.00351$, while the Peto-Prentice test gives $p = 0.00314$. For comparison, the proposed test with $\rho = 1$ gives $p = 0.00152$. While these values are all so small that a difference in a decision is unlikely, Figure 3 shows the running values of the three test statistics over the full course of the study. The figure clearly shows that the log-rank and usual Peto-Prentice statistics track each other very closely. Since the actual BHAT trial was monitored using group-sequential methods, the proposed G^ρ statistic might have lead to the same conclusion at a much earlier time, with a savings of both resources and lives. We will report on the use of the proposed G^ρ family with group-sequential methods in a future article.

Of course, there might be scientific reasons to prefer the log-rank statistic to a G^ρ statistic, but as this example illustrates, the Peto-Prentice statistic might not offer much

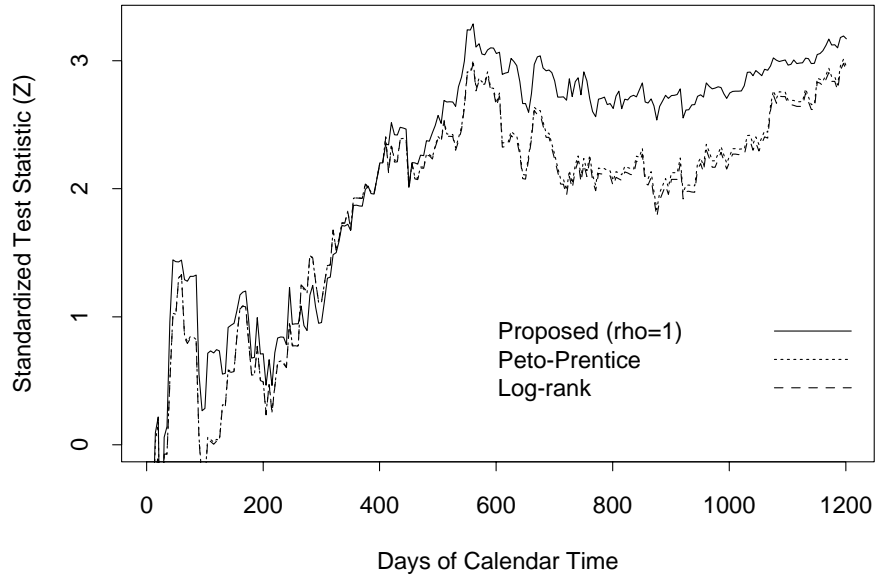


FIGURE 3. The values of the 3 standardized test statistics over the course of the BHAT trial. Negative values at the beginning of the trial are truncated.

of a real alternative when the survival rate is high. This example also illustrates that proposed G^ρ does offer a flexible alternative.

6. EXTENSIONS

The modified G^ρ family will be extended here in two directions. First, an extension will be given in analogy with the $G^{\rho,\gamma}$ family, which extends the G^ρ family and is more efficient for testing survival difference when the difference is more pronounced in the middle; cf. Fleming and Harrington (1991). Second, we will extend the G^ρ family to regression data, where the covariate may take arbitrary values. The regression extension can also be used in the k-sample testing problem since we can view the k-sample problem as a special case of the regression setup by defining a suitable covariate vector.

The $G^{\rho,\gamma}$ family of weighted log-rank statistics is defined, for $\rho \geq 0$ and $\gamma \geq 0$, as

$$U_{\rho,\gamma} = \int_0^\tau \hat{S}^\rho(t-) \hat{F}^\gamma(t-) \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{\cdot\cdot}(t)} \left[\frac{dN_{1\cdot}(t)}{Y_{1\cdot}(t)} - \frac{dN_{2\cdot}(t)}{Y_{2\cdot}(t)} \right]. \quad (6.1)$$

Clearly $U_\rho = U_{\rho,0}$. To see how the $G^{\rho,\gamma}$ statistics work, suppose that we set $\rho = \gamma = 1$. Then the limiting weight function is monotone increasing in t up to the median of $F = F_1$ and decreases afterwards. Therefore, the test statistic emphasizes differences of two survival curves around the median rather than for t near 0 as the U_ρ with $\rho = 1$ does. By adjusting values of ρ and γ , one can select a suitable member to make the resulting test more efficient.

In view of (2.4), we can naturally modify $U_{\rho,\gamma}$ to handle rare events by introducing

$$\tilde{U}_{\rho,\gamma} = \int_0^\tau [\hat{S}(t-) - \hat{S}(\tau-)]^\rho \hat{F}^\gamma(t-) \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{\cdot\cdot}(t)} \left[\frac{dN_{1\cdot}(t)}{Y_{1\cdot}(t)} - \frac{dN_{2\cdot}(t)}{Y_{2\cdot}(t)} \right]. \quad (6.2)$$

Under the rare event formulation of Section 3.1 and the conditions of Theorem 3.2, we can show that $\tilde{U}_{\rho,\gamma}/\hat{\sigma}_{\tilde{U}_{\rho,\gamma}} \xrightarrow{\mathcal{L}} N(0, 1)$ under $\tilde{F}_1 = \tilde{F}_2$ and that $\tilde{U}_{\rho,\gamma}/\hat{\sigma}_{\tilde{U}_{\rho,\gamma}} \xrightarrow{\mathcal{L}} N(\tilde{\mu}_{\rho,\gamma}, 1)$ under the contiguous alternatives as defined by (3.4). Here

$$\hat{\sigma}_{\tilde{U}_{\rho,\gamma}}^2 = \int_0^\tau [\hat{S}(t-) - \hat{S}(\tau-)]^{2\rho} \hat{F}^{2\gamma}(t-) \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{\cdot\cdot}^2(t)} [dN_{1\cdot}(t) + dN_{2\cdot}(t)]$$

and

$$\tilde{\mu}_{\rho,\gamma} = - \frac{\int_0^\tau [\tilde{F}_1(\tau) - \tilde{F}_1(t)]^\rho \tilde{F}^\gamma(t) b(t) d\mu(t)}{\int_0^\tau [\tilde{F}_1(\tau) - \tilde{F}_1(t)]^{2\rho} \tilde{F}^{2\gamma}(t) d\mu(t)}.$$

Based on $\tilde{\mu}_{\rho,\gamma}$, we can derive, for each pair (ρ, γ) , a differential equation similar to (3.11). The solution to that equation then defines a semiparametric family of the form (3.9) so that the modified $G^{\rho,\gamma}$ test is asymptotically optimal.

Following Fleming and Harrington (1982), for the k-sample problem, we can cast it as a special case of the more general regression problem. Thus here we will only show how to extend the modified G^ρ statistics to tests of a covariate effect on survival with regression data.

Suppose there are $i = 1, \dots, n$ study subjects. Let T_i and C_i be the survival and censoring times, Z_i the p -dimensional covariate vector and $\delta_i = I(T_i \leq C_i)$. For the i th subject, define $N_i(t) = \delta_i I(\min\{T_i, C_i\} \leq t)$, the counting process, and $Y_i(t) = I(\min\{T_i, C_i\} \geq t)$, the ‘‘at risk’’ indicator. The extension of $\tilde{U}_{\rho,\gamma}$ to handle regression data is defined by

$$\tilde{U}_{\rho,\gamma,z} = \sum_{i=1}^n \int_0^\tau [\hat{S}(t-) - \hat{S}(\tau-)]^\rho [Z_i - \bar{Z}(t)] dN_i(t),$$

where $\bar{Z}(t) = \sum_{i=1}^n Z_i Y_i(t) / \sum_{i=1}^n Y_i(t)$. Its variance-covariance matrix may be estimated by

$$\hat{\Sigma}_{\tilde{U}_{\rho,\gamma,z}} = \sum_{i=1}^n \int_0^\tau [\hat{S}(t-) - \hat{S}(\tau-)]^{2\rho} \hat{F}^\gamma(t-) [Z_i - \bar{Z}(t)] [Z_i - \bar{Z}(t)]' dN_i(t).$$

Then $\tilde{U}'_{\rho,\gamma,z} \hat{\Sigma}_{\tilde{U}_{\rho,\gamma,z}}^{-1} \tilde{U}_{\rho,\gamma,z}$ can be used to test the null hypothesis that there is no overall covariate effect on survival, i.e., the survival distribution of T_i given Z_i are the same for all i . Under the rare event formulation of Section 3.1, the asymptotic distribution of $\tilde{U}'_{\rho,\gamma,z} \hat{\Sigma}_{\tilde{U}_{\rho,\gamma,z}}^{-1} \tilde{U}_{\rho,\gamma,z}$ under the null hypothesis is χ^2 with p degrees of freedom.

7. CONCLUDING REMARKS

Motivated by the ongoing PLCO trial of the National Cancer Institute, we have proposed a natural modification to the well-known G^ρ family of the weighted log-rank test statistics. The proposed modification incorporates rare event survival data so that the resulting family can effectively deal with different patterns of survival differences with suitable choices of the tuning index ρ . A mathematical formulation for rare event survival data has been given. Under this formulation, asymptotic distributions under both null hypothesis and contiguous alternatives, which are formulated with semi-parametric transformation models, have been derived. The limiting distribution under contiguous alternatives provided a way to study asymptotic efficiency. A semiparametric transformation model was identified for each ρ so that the modified G^ρ statistic is asymptotically optimal. Simulation studies indicated that the proposed testing statistics could lead to efficiency improvement over the original G^ρ statistics when the ratio of the two hazard functions decreases in time. Indeed, the simulations indicated improved efficiency even when the event is not particularly rare.

The proposed modification to the G^ρ family of statistics can be extended in many ways. We have discussed its extension to regression data, which include, as a special case, the k-sample problem. In addition, we presented an analogous modification to $G^{\rho,\gamma}$ family. A very important aspect in the application of weighted log-rank tests that we do not include in this paper is sequential data monitoring and repeated significance testing. This is particularly relevant in many follow-up studies, including the PLCO trial. We are currently investigating extensions of the modified $G^{\rho,\gamma}$ statistics for rare event survival data to handle sequential data monitoring and staggered entry. The findings will be communicated separately.

APPENDIX

Proof of Theorem 2.1: The Theorem would follow from the standard counting process-martingale argument for the weighted log-rank statistics (e.g., Fleming and Harrington (1991), Theorems 7.2.1 and 7.4.1) if the weight function $[\hat{S}(t-) - \hat{S}(\tau-)]^\rho$ were predictable. So, under the null hypothesis $F_2 = F_1$, we can apply the standard argument to get, for the case of $\rho = 0$, that

$$\xi(t) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \int_0^t \frac{Y_{1\cdot}(s)Y_{2\cdot}(s)}{Y_{\cdot\cdot}(s)} \left[\frac{dN_{1\cdot}(s)}{Y_{1\cdot}(s)} - \frac{dN_{2\cdot}(s)}{Y_{2\cdot}(s)} \right]$$

converges weakly to a Gaussian martingale W_ξ which has mean 0 and variance function

$$EW_\xi^2(t) = \int_0^t \frac{r_1 \bar{y}_1(s) r_2 \bar{y}_2(s)}{r_1 \bar{y}_1(s) + r_2 \bar{y}_2(s)} d\Lambda(s).$$

Furthermore, we know the Kaplan-Meier estimate is uniformly consistent over $[0, \tau)$, i.e., $\sup_{t \in [0, \tau)} |\hat{S}(t) - S_1(t)| \xrightarrow{\mathcal{P}} 0$ (Wang, 1987). So, by the Skorodhod strong embedding (Shorack and Wellner, 1986) we have that in some probability space both $\sup_{0 \leq t \leq \tau} |\xi(t) - W_\xi(t)| \rightarrow 0$ a.s., and $\sup_{t \in [0, \tau)} |\hat{S}(t) - S_1(t)| \rightarrow 0$ a.s.

Now,

$$\begin{aligned}
\frac{1}{\sqrt{n}} U_\rho &= \int_0^\tau [\hat{S}(t-) - \hat{S}(\tau-)]^\rho d\xi(t) \\
&= - \int_0^\tau \xi(t) d[\hat{S}(t-) - \hat{S}(\tau-)]^\rho \\
&= - \int_0^\tau W_\xi(t) d[\bar{F}_1(t-) - S_1(\tau-)]^\rho + o(1) \\
&= \int_0^\tau [S_1(t-) - S_1(\tau-)]^\rho dW_\xi(t) + o(1), \tag{A.1}
\end{aligned}$$

where the second and the last equalities come from integration by parts, and the third one from the Helly-Bray Lemma (Chow and Teicher, 1988, p. 256). In addition, it is straightforward to see via the law of large numbers that

$$\frac{1}{n} \hat{\sigma}_{U_\rho}^2 \rightarrow \int_0^\tau [\bar{F}_1(t-) - S_1(\tau-)]^{2\rho} dE W_\xi^2(t), \tag{A.2}$$

which is the variance of $\int_0^\tau [S_1(t-) - \bar{F}_1(\tau-)]^\rho dW_\xi(t)$. Combining (A.1) with (A.2), we get the asymptotic normality of $U_\rho / \hat{\sigma}_{U_\rho}$ under the null hypothesis.

Finally, the convergence of $U_\rho / \hat{\sigma}_{U_\rho}$ under the contiguous alternatives can be derived exactly the same way by applying the strong embedding and integration by parts argument. The details are omitted.

Proof of Theorem 3.1: Let $\tilde{T}_k^* = \max_{1 \leq i \leq n_k} \tilde{T}_{ki}$, $k = 1, 2$. Then $\hat{\Lambda}_k(t) - \Lambda_k(t)$ is a martingale for $t \leq \tilde{T}_k^*$. Applying Lenglar's Inequality (Fleming and Harrington, 1991, p. 113), we get, for any $\epsilon > 0$ and $\eta > 0$,

$$\begin{aligned}
P[m_n \sup_{0 \leq t \leq \tilde{T}_k^*} |\hat{\Lambda}_k(t) - \Lambda_k(t)| \geq \epsilon] &\leq \frac{\eta}{\epsilon} + P[m_n \int_0^{\tilde{T}_k^*} \frac{d\Lambda_k(t)}{Y_{k\cdot}(t)} \geq \eta] \\
&\leq \frac{\eta}{\epsilon} + P\left[\frac{m_n \Lambda_k(\tau)}{n_k n_k^{-1} Y_{k\cdot}(\tau)} \geq \eta\right] \\
&\rightarrow \frac{\eta}{\epsilon}
\end{aligned}$$

as $n_k \rightarrow \infty$. Since η and ϵ are arbitrary, the left hand side must approach 0 as $n_k \rightarrow \infty$ and, therefore, $m_n(\hat{\Lambda}_k(t) - \Lambda_k(t)) \xrightarrow{\mathcal{P}} 0$, which in view of equation (3.2), implies

$$m_k \hat{\Lambda}_k(t) \xrightarrow{\mathcal{P}} \tilde{F}_k(t). \tag{A.3}$$

It is not difficult to show that $m_n[\hat{S}_k(t) - \exp(-\hat{\Lambda}_k(t))] \xrightarrow{\mathcal{P}} 0$. But $\exp(-\hat{\Lambda}_k(t)) = 1 - \hat{\Lambda}_k(t) + O(\hat{\Lambda}_k^2(t))$, which together with (A.3) implies $m_n \hat{F}_k(t) \rightarrow \tilde{F}_k(t)$.

Now under the sequence of contiguous alternatives, we can apply Lengart's Inequality the same way as before to show that

$$m_n \left[\hat{\Lambda}(t) - \int_0^t \frac{Y_{1\cdot}(s) d\Lambda_1(s) + Y_{2\cdot}(s) d\Lambda_2(s)}{Y_{1\cdot}(s) + Y_{2\cdot}(s)} \right] \xrightarrow{\mathcal{P}} 0. \quad (\text{A.4})$$

But the contiguous alternative assumption (3.4) entails that $\Lambda_2(s) = \Lambda_1(s) + o(m_n^{-1})$. Thus, (A.4) implies that $m_n[\hat{\Lambda}(t) - \Lambda_1(t)] \xrightarrow{\mathcal{P}} 0$. The same argument as used under the null hypothesis can be applied again to show $m_n(\hat{F}(t) - F_1(t)) \xrightarrow{\mathcal{P}} 0$ or $m_n \hat{F}(t) \xrightarrow{\mathcal{P}} \tilde{F}_1(t)$.

Proof of Theorem 3.2: It suffices to show convergence under the sequence of contiguous alternatives as specified by (3.4), since they include the null with $b(t) \equiv 0$. Convergence is guaranteed by the following:

$$\frac{m_n^{1/2+\rho}}{n^{1/2}} \tilde{U}_\rho \xrightarrow{\mathcal{L}} N(\sigma_\rho \tilde{\mu}_\rho, \sigma_\rho^2) \quad (\text{A.5})$$

and

$$\frac{m_n^{1+2\rho}}{n} \hat{\sigma}_{\tilde{U}_\rho}^2 \xrightarrow{\mathcal{P}} \sigma_\rho^2 = \int_0^\tau [\tilde{F}_1(\tau) - \tilde{F}_1(t)]^{2\rho} d\mu(t). \quad (\text{A.6})$$

By the definition of \tilde{U}_ρ ,

$$\begin{aligned} \frac{m_n^{1/2+\rho}}{n^{1/2}} \tilde{U}_\rho &= \sqrt{m_n n} \int_0^\tau [m_n \hat{F}(\tau) - m_n \hat{F}(t)]^\rho \\ &\quad \times \frac{n^{-1} Y_{1\cdot}(t) n^{-1} Y_{2\cdot}(t)}{n^{-1} Y_{\cdot\cdot}} [d(\hat{\Lambda}_1(t) - \Lambda_1(t)) - d(\hat{\Lambda}_2(t) - \Lambda_2(t))] \\ &\quad + \sqrt{m_n n} \int_0^\tau [m_n \hat{F}(\tau) - m_n \hat{F}(t)]^\rho \\ &\quad \times \frac{n^{-1} Y_{1\cdot}(t) n^{-1} Y_{2\cdot}(t)}{n^{-1} Y_{\cdot\cdot}} [d\Lambda_1(t) - d\Lambda_2(t)] \\ &= A + B, \end{aligned} \quad (\text{A.7})$$

where A and B are defined by the last equality.

Now

$$\zeta_n(t) \stackrel{\text{def}}{=} \sqrt{m_n n} \int_0^\tau \frac{n^{-1} Y_{1\cdot}(s) n^{-1} Y_{2\cdot}(s)}{n^{-1} Y_{\cdot\cdot}(s)} [d(\hat{\Lambda}_1(s) - \Lambda_1(s)) - d(\hat{\Lambda}_2(s) - \Lambda_2(s))]$$

is a martingale with respect to the natural σ -filtration generated by the processes Y_{ki} , N_{ki} , $k = 1, 2$, $i = 1, \dots, n_k$. By verifying the regularity conditions of the martingale central limit theorem (Rebolledo, 1980; Fleming and Harrington, 1991, pp. 203–204),

we can easily show that ζ_n converges weakly to a Gaussian martingale with mean 0 and variance function $\mu(t)$. Applying integration by parts and the strong embedding as in the proof of Theorem 3.1, we know that part A of (A.7) remains asymptotically the same if $m_n \hat{F}(\tau)$ and $m_n \hat{F}(t)$ are replaced by their respective limits $\tilde{F}_1(\tau)$ and $\tilde{F}_1(t)$. This and the weak convergence of ζ_n imply that $A \xrightarrow{\mathcal{L}} N(0, \sigma_\rho^2)$.

For the other term in (A.7), we use approximation (3.5) to get $B \xrightarrow{\mathcal{P}} \int_0^\tau [\tilde{F}_1(\tau) - \tilde{F}_1(t)]^\rho b(t) d\mu(t)$, which together with the convergence of A shows (A.5) holds.

Finally, (A.6) follows directly from replacing $m_n \hat{F}_k$ and $n^{-1} Y_k$ by their limits.

Proof of Theorem 3.3: Define a new variable $u = g(t)$ and set $u^* = g(\tau)$. Since $g(0) = -\infty$, the range of u is $(-\infty, u^*)$. Equation (3.10) gives

$$(\tilde{H}(u^*) - \tilde{H}(u))^\rho \tilde{H}'(u) = c \tilde{H}''(u),$$

which, upon integration over $(-\infty, u)$, results in

$$\frac{\tilde{H}'(u)}{\tilde{H}^{\rho+1}(u^*) - [\tilde{H}(u^*) - \tilde{H}(u)]^{\rho+1}} = \frac{1}{c(1 + \rho)}. \quad (\text{A.8})$$

Let $a = \tilde{H}(u^*)$ and define

$$L_{a,\rho}(v) = \int_0^v \frac{dx}{a^{\rho+1} - x^{\rho+1}},$$

which can be expressed in terms of hypergeometric functions but is not in general an elementary function. Then integrating (A.8) over (u, u^*) gives

$$\begin{aligned} \frac{u^* - u}{c(1 + \rho)} &= \int_u^{u^*} \frac{d\tilde{H}(s)}{a^{\rho+1} - (a - \tilde{H}(s))^{\rho+1}} \\ &= - \int_{a - \tilde{H}(u)}^0 \frac{dv}{a^{\rho+1} - v^{\rho+1}} \quad (\text{setting } v = a - \tilde{H}(s)) \\ &= L_{a,\rho}(a - \tilde{H}(u)). \end{aligned}$$

Hence $\tilde{H}(u) = a - L_{a,\rho}^{-1}((u^* - u)/c(1 + \rho))$. The result follows by setting $a_2 = a$ and $a_1 = 1/(c(1 + \rho))$.

Now for the special case of $\rho = 1$, $L_{a,\rho}(v) = (2a)^{-1} \log[(a + v)/(a - v)]$, or $L_{a,\rho}^{-1}(u) = a - 2a/(1 + \exp(2au))$. Therefore, $\tilde{H}(u) = a_2 - L_{a_2,\rho}^{-1}(a_1(u^* - u)) = 2a_2/(1 + \exp(2a_2a_1(u^* - u)))$. The result for $\rho = 1$ follows by letting $a_1^* = 2a_2a_1$ and $a_2^* = 2a_2$. On the other hand, with $\rho = 0$, $L_{a,\rho}(v) = \log[a/(a_2 - v)]$, or $L_{a,\rho}^{-1}(u) = a - a \exp(-u)$. Hence, $\tilde{H}(u) = a_2 \exp(-a_1(u^* - u))$ and the result for the case follows by letting $\tilde{a}_1 = a_1$ and $\tilde{a}_2 = a_2 \exp(-a_1 u^*)$.

REFERENCES

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

- β -Blocker Heart Attack Trial Research Group (1982). A randomized trial of Propranolol in patients with acute myocardial infarction. *Journal of the American Medical Association*, **247**, 1707–1714.
- Chow, Y. S., and Teicher, H. (1988). *Probability Theory*. 2nd ed. New York: Springer-Verlag.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187–220.
- Fleming, T. R. and Harrington, D. P. (1981). A class of hypothesis tests for one and two samples censored survival data. *Comm. Statist. A* **10** 763–794.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: John Wiley.
- Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika*. **52** 203–23.
- Gill, R.D. (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts **124**. Amsterdam: Mathematisch Centrum.
- Gohagan, J.K., Prorok, P.C., Kramer, B.S. and Cornett, J.E. (1994). Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute. *J. Urology* **152** 1905–1909.
- Hanks, G.E. and Scardino, P.T. (1996). Does screening for prostate cancer make sense? *Scientific American* **275**, 114–115.
- Harrington, D. P. and Fleming, T. R. (1982) A class of rank test procedures for censored survival data. *Biometrika* **69**, 133–143.
- Hájek, J. and Šíďák, Z. (1967). *Theory of Rank Tests*. New York: Academic Press.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* **50** 163–170.
- Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* **22** 719–748.
- Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *J. Roy. Statist. Soc. A* **135** 185–206.
- Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, **65**, 167–79.
- Randles, R.H. and Wolfe, D.A. (1979). *Introduction to the Theory of Nonparametric Statistics*. New York: Wiley.
- Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete* **51** 269–286.
- Shorack, G. R. & Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.
- Tarone, R. E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika* **64** 156–160.
- Wang, J. G. (1987). A note on the uniform consistency of the Kaplan-Meier estimator. *Ann. Statist.* **15** 1313–1316.
- Kosorok, M. R. and Yin, C. L. (1999). The versatility of function-indexed weighted log-rank statistics. *J. Amer. Statist. Assoc.* **94** 320–332.

STATISTICS DEPARTMENT, HILL CENTER, BUSCH CAMPUS, RUTGERS UNIVERSITY, PISCAT-
AWAY, NJ 08854

E-mail address: buyske@stat.rutgers.edu

BIOMETRY BRANCH, DIVISION OF CANCER PREVENTION, NATIONAL CANCER INSTITUTE,
EPN 344, 6130 EXECUTIVE BOULEVARD, BETHESDA, MD 20892-7354

STATISTICS DEPARTMENT, HILL CENTER, BUSCH CAMPUS, RUTGERS UNIVERSITY, PISCAT-
AWAY, NJ 08854

E-mail address: zyng@stat.rutgers.edu