Lifetime Data Analysis, 3, 225–249 (1997)

An Asymptotic Analysis of the Logrank Test

ROBERT L. STRAWDERMAN strawder@umich.edu
Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029

Revised February 24, 1997

Abstract. Asymptotic expansions for the null distribution of the logrank statistic and its distribution under local proportional hazards alternatives are developed in the case of iid observations. The results, which are derived from the work of Gu (1992) and Taniguchi (1992), are easy to interpret, and provide some theoretical justification for many behavioral characteristics of the logrank test that have been previously observed in simulation studies. We focus primarily upon (i) the inadequacy of the usual normal approximation under treatment group imbalance; and, (ii) the effects of treatment group imbalance on power and sample size calculations. A simple transformation of the logrank statistic is also derived based on results in Konishi (1991) and is found to substantially improve the standard normal approximation to its distribution under the null hypothesis of no survival difference when there is treatment group imbalance.

Keywords: Asymptotic U-statistic, bias-corrected adjustment, clinical trials, Cox regression, Edgeworth expansion, LeCam's Third Lemma, martingale, sample size calculation, unbalanced versus randomized trials

1. Introduction

In clinical trials, the logrank test is often used as a way to compare two populations in terms of their survival experience. Tests of significance typically employ first-order normal theory approximations, and are often done within the framework of the Cox proportional hazards regression model (Cox, 1972) since the logrank test statistic can be derived as a score test of no regression effect (cf. Kalbfleisch and Prentice, 1980). In many instances, the logrank test is applied in settings which involve small sample sizes or imbalanced treatment group assignments. For example, in a Phase II clinical trial with 20–30 patients and approximately 25–30% censoring, the effective sample size is 15–20 patients. Such trial results are often used in the design of larger Phase III studies, and hence calculation of sample sizes, confidence limits for treatment differences, etc... may be adversely affected by tacit use of large sample normal theory approximations. A review of the use of the logrank test in the design and analysis of clinical trial data is given in Peto *et al*. (1977).

When patients are not randomized equally to treatment groups, the logrank test can be either conservative or anti-conservative. This was pointed out in Prentice and Marek (1979) and demonstrated (for uncensored data) in the simulation studies of Kellerer and Chmelevsky (1983, Fig. 2); see also Latta (1981). It has also been observed by many authors (e.g., Hsieh, 1987; Hsieh, 1992; Lakatos and Lan, 1992; Sposto and Krailo, 1987) that both power and sample size calculations are sensitive to treatment group balance. Such calculations are based on the first-order approximation to the power function derived under local proportional hazards alternatives (cf. Schoenfeld, 1981) and studies of their accuracy are generally done via simulation. In fact, as pointed out in Andersen, Borgan, Gill, and Keiding (1993, p. 398), nearly all small sample results for censored data linear rank statistics exist in the form of simulation studies that point out where problems are likely to occur. It is therefore worthwhile to determine if this wealth of simulation-based results can be explained theoretically, and if so, whether such theoretical results suggest improvements that are easily implemented. Both, it turns out, are possible in the case of the logrank test.

Gu (1992, Thm. 2.1) established an Edgeworth expansion to $o(n^{-1/2})$ for the distribution of the studentized score function derived under Cox's proportional hazard regression model for the case of a single covariate and for independent and identically distributed data; hereafter, let this studentized score function be denoted $\mathcal{S}_n$. These results can be used to study the behavior of the logrank statistic under the hypothesis of no survival difference as the latter is equivalent to the former when the regression parameter, say β, is zero. Of course, Gu (1992) points out this connection; however, no in-depth investigation of the logrank statistic is undertaken there. Careful examination of his results in this case shows that the distribution of the logrank statistic is predictably skewed and biased whenever p, the treatment group allocation probability, deviates from 1/2. These results can immediately be used to explain the conservatism (or lack thereof) of the logrank statistic in such cases. In addition, it is easy to establish exactly how the value of p effects the adequacy of the usual normal approximation.

In this paper, the results of Gu (1992) are both complemented and extended in a few ways. In Section 2.1, we introduce some notation, and briefly review the Cox regression model and its relationship to the logrank statistic from a counting process perspective. In Section 2.2, the Edgeworth expansion for the distribution of $\mathcal{S}_n$ is extended to $O(n^{-1})$. This minor extension allows direct application of some transformation theory results in Konishi (1991), leading to a simple transformation of $\mathcal{S}_n$ that is standard normal to $o(n^{-1/2})$ instead of $O(n^{-1/2})$. A second-order expansion is also obtained for the distribution of $\mathcal{S}_n$ under local proportional hazards alternatives. This is accomplished by adapting results of Taniguchi (1992), who devised an extension of LeCam's Third Lemma useful for investigating higher-order local power properties. In Section 3.1, these results for $\mathcal{S}_n$ are made specific to the logrank test, where a more in-depth investigation of each result is separately undertaken. Section 3.1 provides the Edgeworth expansion for the distribution of the logrank statistic and the second-order local power function for the logrank statistic is obtained in Section 3.2. From these results some theoretical insights into the behavioral characteristics of the logrank test are easily obtained. The accuracy of Schoenfeld's sample size formula and also the accuracy of the local power function as an approximation to the exact power function under fixed alternatives is investigated for a model often used in designing clinical trials. In Section 4, a normalizing transformation of the logrank statistic is proposed, and a method for consistently estimating this transformation under the null hypothesis is given. Through simulation, the estimated transformation is shown to produce a more normally distributed test statistic under imbalanced treatment group assignment. We close the paper in Section 5 with some discussion, including remarks on extensions to weighted logrank statistics and to more general settings beyond independent and identically distributed data.

2. Methods

2.1. Notation

Let (T_i, U_i, Z_i), $i = 1 \ldots n$ be independent and identically distributed copies of (T, U, Z), where T is a failure time, U is a censoring time, Z is a covariate taking values on a bounded subinterval of $\mathbb{R}$, and T and U are conditionally independent given Z. Let $X_i = \min(T_i, U_i)$, $D_i = I(X_i = T_i)$, $N_i(t) = I(X_i \le t, D_i = 1)$, and $Y_i(t) = I(X_i \ge t)$. Suppose further that the conditional distribution of T given Z follows the Cox proportional hazards regression model (Cox, 1972); that is, assume that $P\{T \ge t|Z\} = \exp\{-\Lambda(t|Z)\}$, where $\Lambda(t|Z) = \Lambda_0(t)\exp(\beta_0 Z)$, $\beta_0 \in \mathcal{B} \subset \mathbb{R}$ for an open interval $\mathcal{B}$, and $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ for an unspecified positive continuous function $\lambda_0(\cdot)$. Throughout, we let $\pi(t) = E[Y(t)]$.

Define $S^{(j)}(\beta, t) = n^{-1}\sum_{i=1}^n Z_i^j \exp(\beta Z_i) Y_i(t)$ for $j = 0, 1, 2$; then, it follows under the above assumptions that the partial log-likelihood process for β at any time t is given by

$$l_n(\beta, t) = \sum_{i=1}^n \int_0^t \left[\beta Z_i - \log\{S^{(0)}(\beta, u)\}\right] dN_i(u),$$

and associated with this are the score and observed information processes

$$U_n(\beta, t) = \sum_{i=1}^n \int_0^t \left\{ Z_i - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right\} dN_i(u)$$

and

$$I_n(\beta, t) = \sum_{i=1}^n \int_0^t \left[\frac{S^{(2)}(\beta, u)}{S^{(0)}(\beta, u)} - \left\{ \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right\}^2 \right] dN_i(u).$$

Let τ be such that $\pi(\tau) > 0$, define $U_n(\beta) = U_n(\beta, \tau)$ and $I_n(\beta) = I_n(\beta, \tau)$, and let

$$\nu^2 = \int_0^\tau \left[\frac{\alpha_2(u)}{\alpha_0(u)} - \left\{ \frac{\alpha_1(u)}{\alpha_0(u)} \right\}^2 \right] \alpha_0(u)\lambda_0(u)du \qquad (1)$$

be the limit in probability of $n^{-1}I_n(\beta_0)$, where $\alpha_k(t) = E[Z^k \exp\{\beta_0 Z\} Y(t)]$ is the limit in probability of $S^{(k)}(\beta_0, t)$ for all k and $t > 0$. Assume that $\nu^2 > 0$; then, $I_n^{-1/2}(\beta_0)U_n(\beta_0)$ is asymptotically standard normal (Fleming and Harrington, 1991 Thm. 8.4.1).

For future reference, let

$$\Delta = -\int_0^\tau \left\{ \theta_3^{(0)}(u) - \theta_2^{(0)}(u)\frac{\alpha_1(u)}{\alpha_0(u)} \right\} \alpha_0(u)\lambda_0(u)du \qquad (2)$$

and

$$\zeta = -\int_0^\tau \left[\theta_3^{(1)}(u)\Lambda_0(u) - \theta_2^{(1)}(u) \left\{ \int_0^u \frac{\alpha_1(s)}{\alpha_0(s)}\lambda_0(s)ds \right\} \right] \alpha_0(u)\lambda_0(u)du, \qquad (3)$$

where $\theta_k^{(i)}(t) = \frac{\alpha_k^{(i)}(t)}{\alpha_0(t)} - 2\alpha_{k-1}^{(i)}(t)\frac{\alpha_1(t)}{\alpha_0^2(t)} + \alpha_{k-2}^{(i)}(t)\frac{\alpha_1^2(t)}{\alpha_0^3(t)}$, $\alpha_k^{(i)}(t) = E[Z^k \exp\{(i+1)\beta_0 Z\} Y(t)]$, and $\alpha_k(t) = \alpha_k^{(0)}(t)$. These terms will be used to define various skewness and bias correction terms in the Edgeworth expansions which follow. The assumptions given earlier ensure that $\alpha_k^{(i)}(t)$ are bounded functions for $t \in [0, \tau]$.

Suppose $\beta = 0$ and Z is a 0-1 binary variable denoting treatment group membership; the terms "treatment group" and "control group" shall respectively refer to $Z = 1$ and $Z = 0$ throughout the remainder of this paper. Then, $S^{(2)}(0,t) = S^{(1)}(0,t)$, $S^{(1)}(0,t) = \sum_k Z_k Y_k(t)$, and $S^{(0)}(0,t) = Y(t)$; consequently, $I_n^{-1/2}(0)U_n(0)$ reduces to the studentized version of the logrank test statistic. The terms ν^2, Δ, and ζ also simplify greatly (cf. Gu, 1992, Example 2.1); in particular, setting $p = P\{Z = 1\}$ and $\pi(t) = E[Y(t)]$, straightforward calculations show that $\nu^2 = p(1-p)R_0$, $\Delta = -p(1-p)(1-2p)R_0$, and $\zeta = -p(1-p)(1-2p)R_1$, where $R_i = \int_0^\tau \Lambda_0^i(t)\pi(t)\lambda_0(t)dt$ for $i = 0, 1$. The implication of the iid assumption in this particular setting is that the censoring times are all drawn from the same distribution; that is, the case of unequal censoring is precluded from consideration. This is restrictive, but unavoidable since the expansion results of the next section rely heavily on the iid assumption; generalization is possible but not nearly as straightforward.

Finally, let $\phi(\cdot)$ and $\Phi(\cdot)$ respectively denote the standard normal probability density and cumulative distribution functions, and define z_α to be the point such that $\Phi(z_\alpha) = \alpha$, $\alpha \in [0, 1]$.

2.2. *Small Sample Asymptotic Results for Cox Regression*

Theorem 1, given below, deals with expansions for the distribution of the studentized partial likelihood score function and provides the basis for much of the results throughout this paper.

THEOREM 1 *Let $\mathcal{S}_n = I_n^{-1/2}(\beta_0)U_n(\beta_0)$ for $\beta_0 \in \mathcal{B}$, $\rho_3 = -\frac{\Delta}{\nu^3}$, $\rho_2 = -\frac{\zeta}{2\nu^3}$, $\kappa_3 = -\frac{\Delta - 3\zeta}{\nu^3}$, and $\mu_n = -n^{-1/2}\rho_3/3$. Then,*

$$Pr\{\mathcal{S}_n \le z\} = \Phi(z) - n^{-1/2}\phi(z)\left(\frac{\rho_3}{6}(z^2-1) + \rho_2\right) + O(n^{-1}) \tag{4}$$

uniformly in z. In addition, if $\widehat{\rho_j} = \rho_j + O_p(n^{-1/2})$, then for $j = 1, 2$,

$$Pr\left\{T_{n,j}(\mathcal{S}_n) \le z\right\} = \Phi(z) + o(n^{-1/2}) \tag{5}$$

where

$$T_{n,1}(\mathcal{S}_n) = \widehat{\mu}_n^{-1}\left(\exp\{\widehat{\mu}_n\mathcal{S}_n\} - 1\right) - \frac{\widehat{\kappa_3}}{6n^{1/2}}$$

and

$$T_{n,2}(\mathcal{S}_n) = \mathcal{S}_n - n^{-1/2}\left\{\frac{\widehat{\rho_3}}{6}(\mathcal{S}_n^2 - 1) + \widehat{\rho_2}\right\}.$$

The proof of this result is given in the Appendix. Result (4) constitutes a minor extension of Gu (1992, Thm. 2.1), and simply sharpens the error term to $O(n^{-1})$ from $o(n^{-1/2})$. This is useful since (5) then follows directly from results in Konishi (1991); see the Appendix for further details. The dependence of the terms of the expansion on β_0 enter through ν^2, Δ, and ζ; while this dependence has been suppressed, it should be understood throughout. We note that the transformation $T_{n,1}(\cdot)$ given in the theorem is closely related to the transformation underlying the construction of the BC_a bootstrap interval of Efron (1987); see DiCiccio and Tibshirani (1987) and Konishi (1991) for further discussion.

Remark A. The terms κ_3 and ρ_3 are respectively first-order approximations to the standardized skewness of the normalized and studentized versions of the Cox partial likelihood score statistic when β_0 is the true parameter value. The quantity ρ_2 is a first-order correction for the bias induced by studentizing, rather than normalizing, the score function. It can be shown (cf. Gu, 1992, Thm. 2.1) that (4) is identical to the one-term Edgeworth expansion for the distribution of $I_n^{1/2}(\widehat{\beta})(\widehat{\beta} - \beta_0)$, where $\widehat{\beta}$ is the usual MPLE; hence, (5) also provides a way to construct second-order correct confidence intervals for the regression parameter β_0 without resorting to bootstrapping.

A limitation of the above result is the mathematically convenient assumption that $\pi(\tau) > 0$. Gu (1992) argues that this assumption has some practical value from the point of view of robust data analysis; see also Tsiatis, 1981. This is certainly true in some problems, but it does not come entirely without penalty. Consider, for example, a clinical trial of fixed length with accrual to t_0 and ending at time $\tau_0 > t_0$; then, requiring $\pi(\tau) > 0$ is equivalent to setting $\tau = \tau_0 - \delta$ for some $\delta > 0$ chosen *in advance* of seeing the data (cf. Fleming and Harrington, 1991, §8.4). In other words, any observations that become available in $(\tau, \tau_0]$ must be artificially censored at τ, resulting in a loss of efficiency.

Weak convergence of $\mathcal{S}_n(t) = I_n^{-1/2}(\beta_0, t)U_n(\beta_0, t)$ to a Gaussian process on the "maximal interval" $[0, \tau^*]$ for $\tau^* = \sup\{t : \pi(t) > 0\}$ follows if the condition $\pi(\tau) > 0$ is replaced by the assumptions that $P\{Y(t) = 1 \text{ for all } t \leq \tau\} > 0$ for all $\tau < \tau^*$ and that $\nu^2 > 0$ at τ^* (cf. Fleming and Harrington , 1991, Thm. 8.4.4). An extension of Theorem 1 to this case would be useful since the above-described problem no longer arises; however, this remains an open problem.

Taniguchi (1992) obtains a general result on second-order approximations to the distribution of a certain class of test statistics under contiguous alternatives. This is done by proving a second-order extension of LeCam's Third Lemma (cf. Andersen *et al.* , 1993, Thm. 8.1.1). By adapting his results, we can now establish the second-order local power properties of the studentized score function in the case of Cox regression. Theorem 2, which is of some independent interest, shall prove useful in studying the power properties of the logrank statistic. The proof is sketched in the Appendix.

THEOREM 2 *Consider testing $H_0 : \beta = \beta_0$ versus the local sequence of alternatives $H_{an} : \beta_n = \beta_0 + n^{-1/2}\varepsilon$, where $\varepsilon > 0$. Then, under the assumptions of* §*2.1,*

$$P_{\beta_n}\{\mathcal{S}_n \leq x\} = \Phi(x - \nu\varepsilon) - n^{-1/2}\phi(x - \nu\varepsilon)p_1(x - \nu\varepsilon) + o(n^{-1/2}),$$

where $p_1(x) = A_1(x^2 - 1) + A_2 x + 3A_1 + A_3$,

$$A_1 = -\frac{\Delta}{6\nu^3}, \quad A_2 = \frac{\varepsilon(\zeta - \Delta)}{2\nu^2}, \quad A_3 = \frac{\zeta - \Delta}{2\nu^3}\left(\varepsilon^2\nu^2 - 1\right),$$

and $\Delta = \Delta(\beta_0)$, $\zeta = \zeta(\beta_0)$, and $\nu^2 = \nu^2(\beta_0)$ are as defined in (1)–(3).

The left-hand side is the cdf of $\mathcal{S}_n$ under H_{an}; for critical regions of the form (c, ∞) the power is $1 - P_{\beta_n}\{\mathcal{S}_n \leq c\}$. Throughout, reference to power is meant to be interpreted in this way. The leading term provides (as expected) the usual first-order approximation to the power function under local alternatives; in addition, it is easy to check that this expansion reduces to (4) as $\varepsilon \to 0$.

3. Second-Order Asymptotics for the Logrank Test

3.1. An Edgeworth Expansion and Its Implications

The two-sample logrank statistic, defined as $\mathcal{L}_n = I_n^{-1/2}(0)U_n(0)$, is used to test for differences in survival between two populations. A test of no survival difference (equivalently, $\beta = 0$) having Type I error α may be formulated as

$$\text{Reject } H_0 \text{ if} \quad |\mathcal{L}_n| > c_{1-\alpha/2}$$

where $c_{1-\alpha/2}$ satisfies $Pr\{|\mathcal{L}_n| \geq c_{1-\alpha/2}\} = \alpha/2$. It is common practice to set $c_{1-\alpha/2} = z_{1-\alpha/2}$; that is, the distribution of $\mathcal{L}_n$ is assumed to be standard normal under $H_0 : \beta = 0$. The following corollary specializes the results of Theorem 1 to the logrank test and characterizes the dependence of the quality of this standard normal approximation on the group allocation probability p.

COROLLARY 1 *Let Z be a 0-1 binary random variable with $p = P(Z = 1)$. Suppose $\beta_0 = 0$, and define $\Lambda(t)$ to be the cumulative hazard function common to both groups. Then, the conclusions of Theorem 1 hold with $\mathcal{S}_n = \mathcal{L}_n$, $\nu^2 = p(1-p)R_0$, $\Delta = -p(1-p)(1-2p)R_0$, $\zeta = -p(1-p)(1-2p)R_1$, and $R_i = \int_0^\tau \Lambda^i(t)\pi(t)d\Lambda(t) > 0,\ i = 0, 1$. Furthermore, $0 \leq R_1 \leq R_0 \leq 1$.*

The proof, save the last statement, is a direct result of Theorem 1; for completeness, a brief proof that $0 \leq R_1 \leq R_0 \leq 1$ is given in the Appendix. It is interesting to note here that R_0 is exactly the probability of an event on the interval $[0, \tau]$; R_1 has no such simple interpretation. However, some easy calculus shows that as τ approaches $\tau^* = \sup\{t : \pi(t) > 0\}$, $R_k \to (k+1)^{-1}E[\Lambda^k(X)]$. Thus, $R_k,\ k = 0, 1$ are directly linked to the moments of the censored unit exponential random variable $\Lambda(X)$, which should not be terribly surprising since this merely reflects the fact that the logrank test (i) is invariant under monotone transformations of the data; and, (ii) reduces to the exponential ordered scores test of Savage (1956) in the case of no censoring.

It is helpful to recall (see Remark A) that $-\Delta/\nu^3$ and $-\zeta/(2\nu^3)$ represent the standardized skewness and bias of $\mathcal{L}_n$. Hence, when the treatment groups are balanced (i.e., $p = 1/2$), the bias and skewness of $\mathcal{L}_n$ are negligible since $\Delta = \zeta = 0$. On the other hand, for $p > 1/2$ ($p < 1/2$), both the skewness and bias of $\mathcal{L}_n$ increase in the positive (negative) direction, and can adversely impact the accuracy of significance levels. For example, if $p = 1/4$,

$$P\{\mathcal{L}_n \leq z_{1-\alpha/2}\} - \Phi(z_{1-\alpha/2}) = \frac{\phi(z_{1-\alpha/2})}{n^{-1/2}}\left(\frac{0.0361}{R_0^{1/2}}(z_{1-\alpha/2}^2 - 1) + 0.1082\,\frac{R_1}{R_0^{3/2}}\right) + O(n^{-1});$$

since the right-hand side is an even function of $z_{1-\alpha/2}$ and $R_j > 0$, it follows that $\Phi(z_{1-\alpha/2})$ tends to be too small in each tail. That is, the normal approximation is too conservative in the lower tail and too liberal in the upper tail relative to the specified level. For $|z_{1-\alpha/2}| > 1$ (i.e., in the tails), it is a simple matter to verify that this is in fact true for any $p \in (0, 1/2)$, and that the opposite is true for $p \in (1/2, 1)$. Remarkably, the direction of the bias and skewness of the logrank statistic do not depend upon the failure and censoring time distributions to terms which are $o(n^{-1/2})$. Of course, the magnitude of the correction does, but only through the values of $R_k,\ k = 0, 1$. It is important to remember that these are indeed asymptotic results and that the underlying distributions are likely to have some effect in practice, particularly in the case of small sample sizes.

Remark B. For calculating two-sided significance levels, a similar analysis shows that the first-order normal approximation is valid to $O(n^{-1})$. Thus, even though a transformation of $\mathcal{L}_n$ (see Theorem 1) based on a one-term Edgeworth expansion can theoretically improve *one-sided* significance levels, two-sided significance levels computed based on either $\mathcal{L}_n$ or $T_{n,j}(\mathcal{L}_n)$ are (theoretically) equivalent to the same order of approximation. However, as will be shown in Section 4, use of $T_{n,j}(\mathcal{L}_n)$ still provides a measurable benefit in terms of maintaining nominal significance levels when $p \neq 1/2$. It is possible to improve the order of approximation of the two-sided test to $O(n^{-3/2})$. By establishing a valid Edgeworth expansion to $o(n^{-1})$, a transformation of $\mathcal{L}_n$ depending upon its first four cumulants can then be found which is distributed as χ_1^2 to $O(n^{-3/2})$ (e.g., Rao and Mukerjee, 1995). With significant effort the relevant Edgeworth expansion can be determined via the results of Strawderman and Wells (1997, Appendix A) and Gu and Zheng (1993).

Remark C. Suppose $P\{U \geq t\} = S^{\gamma}(t)$ for $\gamma \geq 0$; then, for $k = 0, 1$,

$$R_k = \frac{1}{(\gamma+1)^{k+1}} \left(1 - \exp\{-\Lambda(\tau)(1+\gamma)\}(1 + \Lambda^k(\tau)(1+\gamma)^k)\right).$$

This formula covers both the "fixed interval" ($\gamma = 0$) and Koziol-Green, or proportional hazards, ($\gamma > 0$) censoring models. The fixed-interval censoring case is actually the case of no censoring, but where observation is truncated at τ. In this case, it is easy to see that both R_0 and R_1 converge to 1 as the truncation time $\tau \to \infty$. Suppose the analogous expansion to (4) for the *normalized* Cox score function, obtained by replacing ρ_3 with κ_3 and ρ_2 with 0 (cf. Gu, 1992, Thm. 2.1), remains valid as $\tau \to \infty$; then, these results suggest that

$$\Phi(z) + \frac{1-2p}{3(np(1-p))^{1/2}}\phi(z)(z^2-1) + O(n^{-1})$$

is a valid Edgeworth approximation to the null distribution of the Savage test (cf. Lawless, 1982, Example 8.2.1). This is in fact correct; the above result may be obtained directly from the results of Does (1983, eqn. 5.1) on Edgeworth expansions for linear rank statistics by using the logrank scores (given in his notation) $J(u) = -\log(1-u) - 1$ (cf. Prentice, 1978, p. 172).

3.2. Power under Local Alternatives

Consider testing $H_0 : \lambda_1(t) = \lambda_0(t)$ for all t versus $H_{an} : \lambda_{1n}(t) = \lambda_0(t)\exp\{n^{-1/2}\varepsilon\}$, where $\lambda_j(t)$ is the hazard function corresponding to $Z = j,\ j = 0, 1$ and $\varepsilon > 0$. Schoenfeld (1981) showed that $\Phi(z_{1-\alpha} - \nu\varepsilon)$ is the first-order approximation to the power function of the logrank statistic for detecting such local alternatives at a fixed significance level α and with power $1 - \omega$. Setting $\varepsilon = \sqrt{n}\psi$, he also showed that

$$n_{1-\omega} = \frac{(z_{1-\alpha} + z_{1-\omega})^2}{\nu^2\psi^2} \tag{6}$$

approximates the sample size required for detecting a log-hazard ratio difference of ψ at a fixed significance level α with power $1-\omega$ based on a one-sided test. Note that this formula is obtained by solving $\Phi(z_{1-\alpha} - \nu n^{1/2}\psi) = \omega$ for n.

For equal censoring distributions and balanced completely randomized designs, Corollary 1 implies that $\mathcal{L}_n$ is distributed as standard normal to $O(n^{-1})$ instead of the usual $O(n^{-1/2})$. The accuracy of this approximation reverts back to $O(n^{-1/2})$ once $p \neq 1/2$. If it is assumed that the respective group hazards are proportional rather than equal, then this is also the case for the partial likelihood score test of $H_0 : \beta = \beta_0$ when $\beta_0 \neq 0$ regardless of whether the treatment groups are balanced. Since the distribution of $\mathcal{S}_n$ (the Cox partial likelihood score function) behaves smoothly in β_0, then together these results imply that (6) (and hence the corresponding first-order approximation to the power function) should be reasonably accurate when the design is close to balanced and the proportional hazards differences to be detected are small. Similarly, one should therefore expect the quality of these approximations to degrade as imbalance increases and/or the proportional hazards alternatives to be detected become large. Evidence of this in practice can be found in Hsieh (1987, Table I; 1992, Table V); see also Lakatos and Lan (1992).

It is thus of interest to quantify the effects of treatment group imbalance and the alternative under consideration on power and sample size calculations. The following corollary to Theorem 2 gives the second-order power function of the logrank test under local proportional hazards alternatives.

COROLLARY 2 *Consider testing* $H_0 : \lambda_1(t) = \lambda_0(t)$ *for all* t *versus* $H_{an} : \lambda_{1n}(t) = \lambda_0(t)\exp\{n^{-1/2}\varepsilon\}$ *for* $\varepsilon > 0$. *Then, under the same assumptions as Theorem 2,*

$$P_{\lambda_{1n}}\{\mathcal{L}_n \leq x\} = \Phi(x - \nu\varepsilon) - n^{-1/2}\phi(x - \nu\varepsilon)p_1(x - \nu\varepsilon) + o(n^{-1/2}), \tag{7}$$

where $p_1(x) = A_1(x^2 - 1) + A_2 x + 3A_1 + A_3$, $\nu^2 = p(1-p)R_0$,

$$A_1 = \frac{1-2p}{6\nu},\ A_2 = \frac{\varepsilon(1-2p)}{2}\left(1 - \frac{R_1}{R_0}\right),\ \textit{and}\ A_3 = \frac{(1-2p)}{2\nu}\left(1 - \frac{R_1}{R_0}\right)(\varepsilon^2\nu^2 - 1).$$

This result shows that when $p = 1/2$, each of $A_1 \ldots A_3$ is zero, implying that the first-order normal approximation to the power function for detecting local proportional hazards alternatives (i.e., $\Phi(x - \nu\varepsilon)$) remains second-order accurate. Thus, as argued earlier, these results imply that (6) is typically accurate when the design is balanced and the assumptions made in deriving the sample size are met. In general (i.e., when $p \neq 1/2$), the factors determining the quality of the first-order approximation are primarily p and ε, and to a lesser extent $R_k, k = 0, 1$, which represent the influence of the failure and censoring distributions. Since $0 \le R_1 \le R_0 \le 1$ (see Corollary 1), then by fixing p and ε, the signs of $A_j,\ j = 1 \ldots 3$ are determined. Hence, it is possible to predict when $\Phi(x - \nu\varepsilon)$ over- or underestimates the power for fixed n, x, ε, and p.

3.2.1. *Evaluating the Accuracy of Schoenfeld's Formula*

By setting $\varepsilon = \sqrt{n}\psi$ in Corollary 2, one can assess the accuracy of (6) for various combinations of p and ψ by evaluating (7) at $n_{1-\omega}$. If the first-order normal approximation to the power function of the logrank test is accurate, then the power calculation based on (7) for $n_{1-\omega}$ should approximately be equal to $1 - \omega$. Correspondingly, lower than specified power indicates that sample sizes based on (6) are too small, while higher than specified power indicates the opposite.

To make this discussion more concrete, assume that failure times are exponentially distributed and that we wish to detect the difference $\lambda_1 = \lambda_0 e^{\psi}$; in the context of Corollary 2, this corresponds to setting $\varepsilon = n^{1/2}\psi$. Suppose that censoring occurs according to the following clinical trial model (cf. Kalish and Harrington, 1988): patients are accrued uniformly from study start for L years, and then followed for an additional B years. If U denotes the censoring variable, then

$$G(u) = \begin{cases} 1 & 0 < u \le B \\ (L + B - u)/L & B < u \le L + B \\ 0 & u > L + B \end{cases}$$

for $G(u) = P\{U \ge u\}$. This model can be extended to account for loss-to-followup by assuming that dropout occurs independently of treatment group and according to some specified distribution; however, this will not be done here. Treatment-dependent dropout is precluded here due to the assumption that the data are iid. In the above set-up, the local power function depends upon the failure and censoring distributions only through λ_0, ψ, L, and B. In fact, except for ψ, this dependence is only through the values of R_k, which in turn only need to be calculated under the hypothesis of no survival difference. Specifically,

$$R_k = \lambda_0^{k+1} \int_0^{L+B} t^k e^{-\lambda_0 t} G(t)\,dt,$$

which is easily evaluated for given λ_0, L, and B. Technically, in view of the fact that the theory requires there being a positive probability of being at risk when the analysis is done, one should only integrate to some time less than $L + B$. However, this shall be ignored since such a point can be chosen to be arbitrarily close to $L + B$, making virtually no difference in numerical calculations.

To study the accuracy of (6), consider designing a hypothetical clinical trial with L years of accrual lasting a total of 10 years (i.e., $B = 10 - L$ years of additional followup). The possible values for L under consideration are 1, 5, and 9 years. Survival is assumed to be exponential, and is respectively determined for the control and treatment groups by specifying the proportion surviving at the end of 10 years ($P\{T > 10\} = 0.8$ or 0.2) and the hazard ratio e^{ψ} (1.5, 1.35, 1.20, and 1.05). For given L, these parameters determine the values of R_0 and R_1. Figures 1-3 contain plots of the approximate power based on (7) for a two-sided logrank test at a significance level of 0.05 and for detecting a log-hazard ratio of ψ with power 0.90. Each figure corresponds to one value of L, and contains two panels: one for $P\{T > 10\} = 0.8$, the other for $P\{T > 10\} = 0.2$. Each curve in a given panel is plotted as a function of p, and there is a different curve for each of the 4 hazard ratios (HR=e^{ψ}) under consideration. The sample size used in the calculation of each curve at each combination of p and ψ is obtained from (6), and thus any two points on a given curve correspond to different sample sizes; this is discussed in more detail below. The values of R_0 and R_1 resulting from the assumptions specified in the plot heading are given for reference; it may be useful to recall that R_0 is the probability of death in the control group (i.e. $Z = 0$) and (asymptotically) in the treatment group. Deviations of each curve from 0.90 therefore reflect inaccuracy due to the use of the first-order normal approximation, at least in theory.

Assuming the second-order approximation to be accurate, the plots in Figures 1–3 indicate that the sample size formula (6) is most accurate when e^{ψ} is close to 1.0 (i.e., ψ close to 0) and the treatment group allocation probability p is close to 1/2, as suspected. Of course, the curves are by construction required to pass through 0.90 at $p = 1/2$; consequently, it is the steepness of the curve as one moves away from this point which tells the important story. The approximation (6) gets progressively worse the farther one moves away from the "best-case" values of $\psi = 0$ and $p = 1/2$. However, the deviations from 0.9 are not particularly extreme, suggesting that Schoenfeld's formula will still lead to a reasonable choice of sample size for moderately imbalanced designs.

It is important to understand that Figures 1–3 do not depict typical power curves. Within each panel, the sample size used in calculating the power via (7) at each combination of p and ψ is obtained from (6); the latter, once we have fixed the nominal power (i.e., 0.9), size, and alternative, is in fact a function of p only (i.e., (6) = $n(p)$). Hence, as p varies, so does $n = n(p)$; that is, each point along a particular curve thus corresponds to the power at a *different* sample size. This is important since the curves depicted here are all seen to be decreasing in p. For a fixed sample size n, the power curve for detecting a particular ψ is usually "skew-parabolic" as a function of p. This is the reason for using "power trace" in the figure titles; the term "power curve" is hereafter reserved for fixed values of n.

One should be careful not to overinterpret these results. The approximations used in generating Figures 1–3 are for alternatives that are $O(1)$ distance away from the null hypothesis, not $O(n^{-1/2})$; this is because the approximation, say, for detecting a fixed alternative correspoding to a hazard ratio of 1.5 is obtained by substituting $\varepsilon = \sqrt{n}\log 1.5$ in (7). Consequently, a significant extrapolation is being made in order to approximate the power at a given sample size. It has been shown by various authors that such extrapolation can produce inaccurate results; see, for example, Pfaff and Pfanzagl (1985) or Nelson and Savin (1990). Hence, it is worthwhile to investigate whether (7) is useful for approximating the exact power for detecting fixed alternatives.

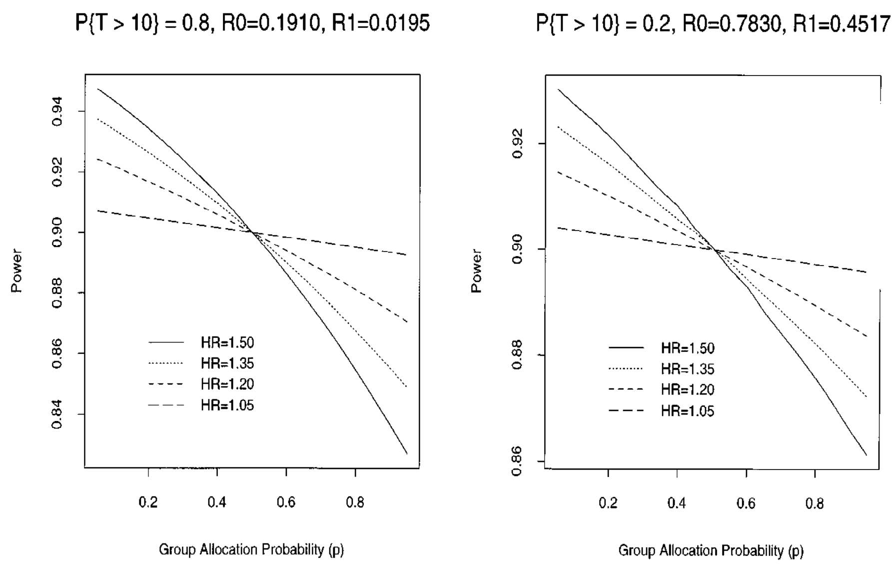

Figure 1. Power traces under Schoenfeld's sample size formula (2-sided, size = 0.05, power = 0.90, Accrual(L) = 1, Followup(B) = 9).

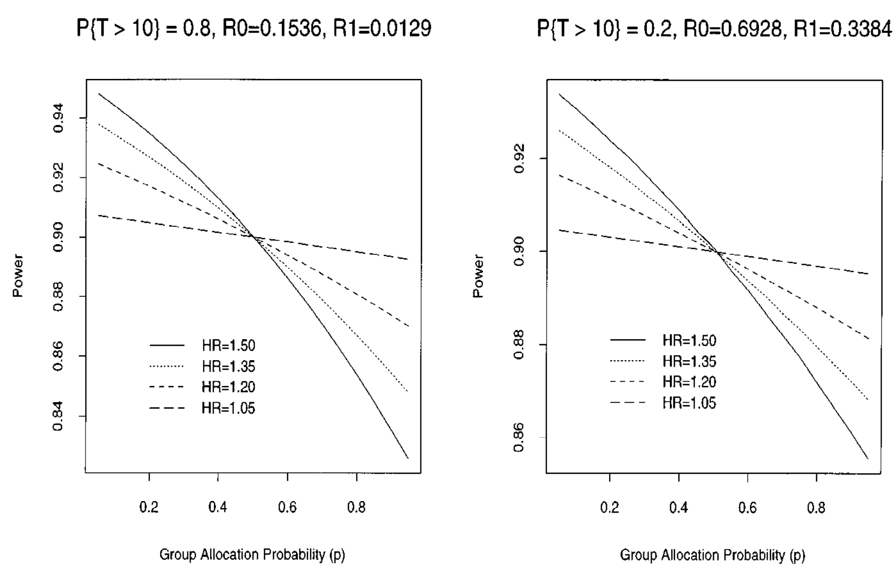

Figure 2. Power traces under Schoenfeld's sample size formula (2-sided, size = 0.05, power = 0.90, Accrual(L) = 5, Followup(B) = 5).

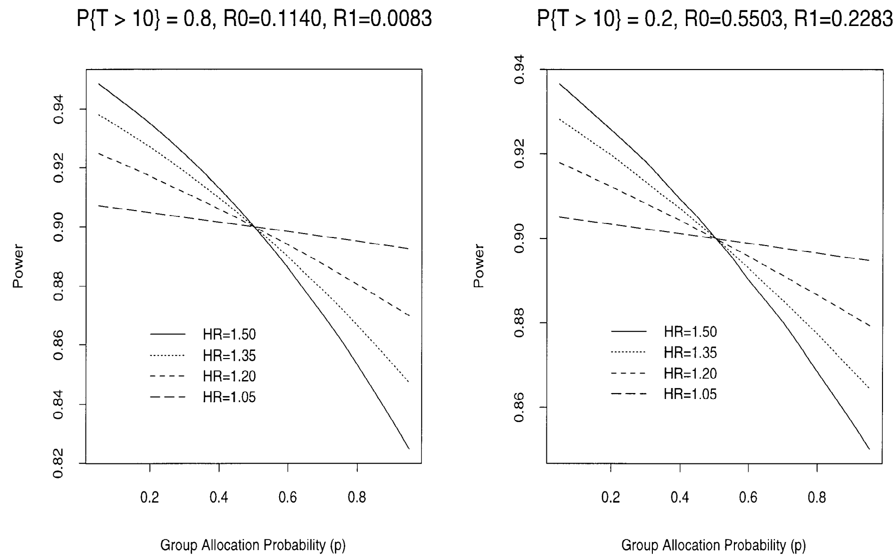

Figure 3. Power traces under Schoenfeld's sample size formula (2-sided, size $= 0.05$, power $= 0.90$, Accrual(L) $=$ 9, Followup(B) $=$ 1).

3.2.2. On the Accuracy of (7) under Fixed Alternatives

To assess whether (7) provides an adequate approximation for nonlocal alternatives, a simulation study for the settings corresponding to the right-hand panels of Figures 1–3 was done. Two sample sizes (n=50, 100) are considered, and power is calculated for values of the group allocation probability p from 0.1 to 0.9 in increments of 0.1. The exact power in each case is estimated via Monte Carlo (5000 replicates); the results are summarized in Tables 1–3. It should be noted here that the values of R_0 and R_1 reflect only the control group; recall that (7) is essentially derived by assuming the treatment and control groups to be from the same population asymptotically. The value of R_0 thus represents the probability of death within the control group; under a fixed alternative, the actual probability of death (and hence total censoring levels) in a finite sample depends additionally upon the fixed alternative e^{ψ} and the group allocation probability p. This has interesting implications for the accuracy of (7) as a function of p which will be discussed later.

The columns labeled "Normal" and Edg_1 correspond to the first- and second-order approximations to the power function. The approximation Edg_2 is discussed later in this Section. Tables 1–3 show that for alternatives close to the null hypothesis (e.g. $e^{\psi} = 1.05, 1.2$) and uniformly in p, Edg_1 is generally much more accurate as an approximation to the exact power than is the first-order normal approximation. Furthermore, when $p < 1/2$, Edg_1 is essentially uniformly more accurate than the first-order approximation for each ψ; however, both usually underestimate the exact power. An interesting change occurs for $p > 1/2$; the

Table 1. Simulation results for approximating $P\{\mathcal{L}_n > 1.96 | \lambda_1(t) = \lambda_0(t)e^{\psi}\}$

$L = 1$, $B = 9$, $P\{T > 10\} = 0.2$ (equiv. $\lambda_0 = -\log(0.2)/10$).

Hazard Ratio (e^{ψ})	$P\{Z = 1\}$	n = 50				n = 100			
		Normal	Edg_1†	Edg_2‡	Exact	Normal	Edg_1†	Edg_2‡	Exact
	0.1	0.060	0.094	0.094	0.121	0.065	0.089	0.089	0.091
	0.2	0.064	0.082	0.082	0.086	0.070	0.084	0.084	0.086
	0.3	0.066	0.077	0.077	0.075	0.074	0.082	0.082	0.079
	0.4	0.067	0.072	0.072	0.074	0.076	0.079	0.080	0.083
1.05	0.5	0.068	0.068	0.068	0.075	0.076	0.076	0.077	0.079
	0.6	0.067	0.063	0.063	0.070	0.076	0.072	0.073	0.076
	0.7	0.066	0.057	0.057	0.060	0.074	0.067	0.067	0.068
	0.8	0.064	0.049	0.049	0.051	0.070	0.059	0.059	0.062
	0.9	0.060	0.037	0.037	0.042	0.065	0.046	0.046	0.047
	0.1	0.096	0.140	0.140	0.162	0.123	0.157	0.158	0.157
	0.2	0.117	0.143	0.144	0.149	0.159	0.181	0.182	0.187
	0.3	0.131	0.147	0.148	0.150	0.183	0.196	0.198	0.212
	0.4	0.139	0.146	0.148	0.148	0.196	0.203	0.206	0.213
1.2	0.5	0.141	0.141	0.144	0.136	0.201	0.201	0.205	0.202
	0.6	0.139	0.131	0.134	0.135	0.196	0.190	0.195	0.190
	0.7	0.131	0.116	0.119	0.123	0.183	0.169	0.175	0.168
	0.8	0.117	0.095	0.097	0.108	0.159	0.139	0.143	0.147
	0.9	0.096	0.064	0.065	0.071	0.123	0.094	0.097	0.104
	0.1	0.140	0.194	0.195	0.215	0.198	0.244	0.245	0.253
	0.2	0.186	0.221	0.223	0.223	0.280	0.312	0.316	0.314
	0.3	0.216	0.238	0.242	0.249	0.334	0.355	0.362	0.366
	0.4	0.234	0.245	0.251	0.258	0.365	0.376	0.386	0.388
1.35	0.5	0.240	0.240	0.248	0.259	0.376	0.376	0.390	0.399
	0.6	0.234	0.224	0.234	0.246	0.365	0.355	0.373	0.362
	0.7	0.216	0.196	0.207	0.209	0.334	0.314	0.333	0.334
	0.8	0.186	0.154	0.165	0.162	0.280	0.250	0.268	0.272
	0.9	0.140	0.097	0.103	0.114	0.198	0.158	0.171	0.169
	0.1	0.188	0.253	0.254	0.274	0.285	0.344	0.345	0.361
	0.2	0.264	0.308	0.312	0.308	0.417	0.459	0.465	0.469
	0.3	0.315	0.343	0.351	0.339	0.500	0.527	0.539	0.511
	0.4	0.344	0.358	0.370	0.371	0.545	0.558	0.577	0.572
1.5	0.5	0.353	0.353	0.371	0.358	0.559	0.559	0.585	0.576
	0.6	0.344	0.330	0.351	0.346	0.545	0.531	0.563	0.556
	0.7	0.315	0.287	0.311	0.307	0.500	0.472	0.510	0.503
	0.8	0.264	0.224	0.247	0.247	0.417	0.376	0.415	0.410
	0.9	0.188	0.135	0.151	0.150	0.285	0.231	0.261	0.259

† $R_0 = 0.7830$, $R_1 = 0.4517$
‡ Uses $\widetilde{R}_k,\ k = 0, 1$; see Section 3.2.2

normal approximation often overestimates the power while Edg_1 usually underestimates it. A clear pattern here is that the accuracy of Edg_1 degrades as $p \to 1$ and $\psi \to \infty$.

It may initially seem curious that the accuracy of (7) should so depend heavily on whether $p < 1/2$ versus $p > 1/2$. There is a rather simple explanation for this phenomenon. Data realized under a fixed nonlocal alternative are generated from a mixture of two populations.

Table 2. Simulation results for approximating $P\{\mathcal{L}_n > 1.96|\lambda_1(t) = \lambda_0(t)e^{\psi}\}$

$L = 5$, $B = 5$, $P\{T > 10\} = 0.2$ (equiv. $\lambda_0 = -\log(0.2)/10$).

Hazard Ratio (e^{ψ})	$P\{Z = 1\}$	n = 50 Normal	n = 50 Edg$_1$†	n = 50 Edg$_2$‡	n = 50 Exact	n = 100 Normal	n = 100 Edg$_1$†	n = 100 Edg$_2$‡	n = 100 Exact
1.05	0.1	0.060	0.093	0.093	0.110	0.064	0.087	0.087	0.083
	0.2	0.063	0.081	0.081	0.089	0.069	0.082	0.082	0.085
	0.3	0.065	0.075	0.075	0.077	0.072	0.080	0.080	0.080
	0.4	0.066	0.071	0.071	0.068	0.074	0.078	0.078	0.074
	0.5	0.067	0.067	0.067	0.070	0.075	0.075	0.075	0.078
	0.6	0.066	0.062	0.062	0.064	0.074	0.071	0.071	0.073
	0.7	0.065	0.056	0.056	0.058	0.072	0.065	0.065	0.066
	0.8	0.063	0.048	0.048	0.045	0.069	0.058	0.058	0.056
	0.9	0.060	0.036	0.036	0.038	0.064	0.046	0.046	0.043
1.2	0.1	0.093	0.137	0.137	0.145	0.117	0.152	0.152	0.162
	0.2	0.112	0.138	0.139	0.149	0.150	0.172	0.173	0.175
	0.3	0.124	0.140	0.142	0.140	0.171	0.185	0.188	0.184
	0.4	0.132	0.139	0.141	0.146	0.184	0.191	0.194	0.197
	0.5	0.134	0.134	0.136	0.138	0.188	0.188	0.192	0.191
	0.6	0.132	0.124	0.127	0.129	0.184	0.177	0.182	0.179
	0.7	0.124	0.110	0.113	0.111	0.171	0.158	0.163	0.159
	0.8	0.112	0.089	0.092	0.093	0.150	0.129	0.134	0.133
	0.9	0.093	0.060	0.062	0.064	0.117	0.088	0.092	0.100
1.35	0.1	0.132	0.188	0.189	0.203	0.185	0.234	0.235	0.237
	0.2	0.174	0.210	0.213	0.216	0.259	0.293	0.298	0.299
	0.3	0.202	0.225	0.229	0.229	0.308	0.331	0.339	0.345
	0.4	0.218	0.229	0.236	0.243	0.337	0.348	0.361	0.353
	0.5	0.223	0.223	0.233	0.229	0.346	0.346	0.363	0.365
	0.6	0.218	0.207	0.218	0.214	0.337	0.326	0.346	0.351
	0.7	0.202	0.180	0.192	0.191	0.308	0.287	0.308	0.299
	0.8	0.174	0.142	0.153	0.147	0.259	0.227	0.248	0.244
	0.9	0.132	0.089	0.097	0.099	0.185	0.144	0.158	0.154
1.5	0.1	0.176	0.244	0.246	0.253	0.264	0.326	0.329	0.333
	0.2	0.245	0.292	0.297	0.306	0.384	0.430	0.438	0.428
	0.3	0.291	0.321	0.331	0.332	0.461	0.491	0.507	0.504
	0.4	0.317	0.332	0.347	0.344	0.503	0.518	0.542	0.530
	0.5	0.326	0.326	0.346	0.344	0.517	0.517	0.549	0.541
	0.6	0.317	0.302	0.327	0.317	0.503	0.488	0.527	0.528
	0.7	0.291	0.262	0.289	0.283	0.461	0.431	0.475	0.458
	0.8	0.245	0.202	0.228	0.225	0.384	0.339	0.384	0.380
	0.9	0.176	0.122	0.139	0.139	0.264	0.207	0.240	0.231

† $R_0 = 0.6928$, $R_1 = 0.3384$
‡ Uses $\widetilde{R}_k,\ k = 0, 1$; see Section 3.2.2

The shape and location of the exact power function thus depends on the corresponding mean, variance, and skewness of $\mathcal{L}_n$ for that particular setting. However, (7) is derived by exploiting the behavior of the test statistic under the null hypothesis. That is, the assumption that all data are asymptotically generated from a distribution governed by $\lambda_0(t)$, the hazard corresponding to the control group, plays a central role in its development. It can be

Table 3. Simulation results for approximating $P\{\mathcal{L}_n > 1.96|\lambda_1(t) = \lambda_0(t)e^{\psi}\}$

$L = 1$, $B = 9$, $P\{T > 10\} = 0.2$ (equiv. $\lambda_0 = -\log(0.2)/10$).

Hazard Ratio (e^{ψ})	$P\{Z = 1\}$	n = 50 Normal	n = 50 Edg$_1$†	n = 50 Edg$_2$‡	n = 50 Exact	n = 100 Normal	n = 100 Edg$_1$†	n = 100 Edg$_2$‡	n = 100 Exact
	0.1	0.058	0.093	0.093	0.113	0.062	0.087	0.087	0.088
	0.2	0.061	0.080	0.080	0.084	0.067	0.080	0.081	0.085
	0.3	0.063	0.074	0.074	0.069	0.070	0.078	0.078	0.082
	0.4	0.064	0.069	0.069	0.069	0.071	0.075	0.075	0.072
1.05	0.5	0.065	0.065	0.065	0.070	0.072	0.072	0.072	0.071
	0.6	0.064	0.060	0.060	0.061	0.071	0.068	0.068	0.062
	0.7	0.063	0.054	0.054	0.055	0.070	0.062	0.062	0.065
	0.8	0.061	0.046	0.046	0.048	0.067	0.055	0.055	0.049
	0.9	0.058	0.035	0.035	0.037	0.062	0.044	0.044	0.048
	0.1	0.087	0.133	0.134	0.142	0.108	0.144	0.144	0.150
	0.2	0.103	0.131	0.132	0.132	0.135	0.158	0.159	0.161
	0.3	0.114	0.130	0.132	0.133	0.153	0.167	0.170	0.167
	0.4	0.120	0.128	0.130	0.132	0.163	0.170	0.174	0.178
1.2	0.5	0.122	0.122	0.124	0.113	0.166	0.166	0.171	0.165
	0.6	0.120	0.112	0.115	0.119	0.163	0.156	0.161	0.151
	0.7	0.114	0.099	0.102	0.104	0.153	0.139	0.144	0.140
	0.8	0.103	0.080	0.083	0.082	0.135	0.114	0.119	0.110
	0.9	0.087	0.054	0.056	0.053	0.108	0.078	0.082	0.078
	0.1	0.120	0.179	0.180	0.187	0.164	0.215	0.216	0.214
	0.2	0.155	0.193	0.195	0.189	0.225	0.261	0.265	0.261
	0.3	0.178	0.202	0.206	0.214	0.266	0.290	0.298	0.294
	0.4	0.191	0.203	0.210	0.209	0.290	0.301	0.314	0.307
1.35	0.5	0.195	0.195	0.205	0.208	0.297	0.297	0.314	0.320
	0.6	0.191	0.180	0.191	0.200	0.290	0.278	0.298	0.291
	0.7	0.178	0.156	0.168	0.161	0.266	0.244	0.265	0.249
	0.8	0.155	0.122	0.133	0.132	0.225	0.193	0.212	0.211
	0.9	0.120	0.077	0.084	0.086	0.164	0.122	0.136	0.133
	0.1	0.157	0.228	0.230	0.225	0.229	0.295	0.298	0.297
	0.2	0.214	0.263	0.268	0.262	0.329	0.378	0.387	0.377
	0.3	0.251	0.283	0.293	0.283	0.395	0.428	0.445	0.433
	0.4	0.273	0.289	0.305	0.304	0.432	0.448	0.474	0.460
1.5	0.5	0.280	0.280	0.301	0.281	0.444	0.444	0.478	0.479
	0.6	0.273	0.258	0.283	0.274	0.432	0.416	0.457	0.442
	0.7	0.251	0.221	0.248	0.240	0.395	0.363	0.408	0.385
	0.8	0.214	0.170	0.195	0.184	0.329	0.283	0.326	0.307
	0.9	0.157	0.102	0.119	0.112	0.229	0.172	0.203	0.194

† $R_0 = 0.5503$, $R_1 = 0.2283$

‡ Uses $\widetilde{R}_k$, $k = 0, 1$; see Section 3.2.2

seen that the influence of any specific alternative ψ only enters the correction term in (7) through $\varepsilon = \sqrt{n}\psi$; the terms $A_1 - A_3$ are otherwise completely determined by $\lambda_0(t)$. As $p \to 0$, the observed data primarily reflect the characteristics of the control group; the terms $A_1 - A_3$ are thus computed appropriately and the alternative under consideration plays a comparatively minor role. In contrast, as more individuals are assigned to the treatment group (i.e., as $p \to 1$), it is not difficult to see that the correction term becomes increasingly less able to capture the true distributional characteristics of $\mathcal{L}_n$ (i.e., mean, variance, and skewness). The quality of the approximation in this case also depends more heavily upon the distance between the alternative and null hypotheses than when $p < 1/2$. At the minimum, improvements to this approximation therefore must account for the fact that the composition of the sample changes with p. How to accomplish this in a theoretically appropriate manner is not clear.

To propose an adhoc solution, note that the influence of the underlying failure and censoring time distributions occurs primarily though R_k. Hence, consider replacing the values of R_k with the weighted mixture

$$\widetilde{R}_k = pR_k(\Lambda_1) + (1-p)R_k(\Lambda_0),$$

where $R_k(\Lambda) = \int_0^\tau \Lambda^i(t)\pi(t)d\Lambda(t)$. It is easy to show that $\widetilde{R}_0$ is the probability of death in the combined sample; as before, $\widetilde{R}_1$ has no simple interpretation. The value of $\widetilde{R}_k$ depends both on p and ψ and hence adapts to the changing composition of the sample under a fixed alternative as p moves between zero and one. It is a well-known phenomenon that the error of an asymptotic expansion increases from the point at which the expansion is started (e.g., Pfaff and Pfanzagl, 1985). Hence, a heuristic justification for why this approach may behave reasonably well is that the resulting local approximation based on (7) should be located closer the distribution of interest, hence reducing extrapolation. This is similar in spirit to a suggestion made in Pfaff and Pfanzagl (1985) for improving the numerical accuracy of approximations like (7) to power under fixed alternatives. However, solid theoretical justification for why this proposal should necessarily work is lacking at this time. That being said, the approximation Edg_2 in Tables 1–3 is computed in exactly the same manner as Edg_2, but with $\widetilde{R}_k$ in place of R_k. The resulting accuracy of this approximation is rather astounding, particularly for large alternatives and p near one, where it uniformly improves upon Edg_1. As a bonus, it behaves very similarly to Edg_1 for $p < 1/2$; this is of course to be expected since $\widetilde{R}_k \to R_k$ as $p \to 0$.

To interpret the simulation results in the context of Figures 1-3, the above suggests that when $p < 1/2$, (7) provides a reasonable approximation to the exact power function under a fixed alternative, provided the alternative is "not too far away" from the null hypothesis. For $p > 1/2$, the actual power of the logrank test apparently lies somewhere in between the curve generated by (7) and 0.9. An interesting interpretation of the relationship between power, the sampling proportion p, and the sample size n can be gleaned from these results. Common wisdom dictates that a balanced design (i.e., $p = 1/2$) will provide maximum power for detecting proportional hazards alternatives. Aside from being intuitively appealing, the first order-approximation to the local power function always leads one to this conclusion. However, the results of the simulation study (and indeed the structure of the approximation (7)) show that a balanced design does not necessarily lead to the maximum power. The optimal allocation for a particular setting depends in a complicated manner on the sample size, the alternative of interest, and the failure/censoring time distributions. The prevailing trend in Tables 1–3 is that for alternatives closer to the null hypothesis of no survival difference the power is maximized for some $p < 1/2$; the closer the alternative is to the null, the closer p is to zero. This corresponds to allocating more patients to the group having the lower hazard rate, which tends to balance the number of *events* in each group more evenly than when $p = 1/2$. In contrast, allocating less patients to the group with the lower hazard (i.e., $p > 1/2$) causes an even greater imbalance in the number of events. Interestingly, the optimal allocation moves closer to $p = 1/2$ once the alternative to be detected is far enough away from the null (e.g., $e^{\psi} = 1.5$).

These results suggest that if prior knowledge of the superiority of one treatment over the other is strong, an unequal allocation design can provide equal power for detecting a specified difference at a smaller sample size and hence at less cost. It is not difficult to see from (7) that there are an infinite number of (n, p) combinations which lead to the same power. As seen in Tables 1–3, the maximal power does not necessarily occur at $p = 1/2$. Sposto and Krailo (1987) argued on rather heuristic grounds that unbalanced allocation could provide such benefits. In the examples considered, the difference in power between the optimal allocation and that for $p = 1/2$ is negligible, being less than 2% in all cases. This suggests that any statistical gains realized by unequally allocating patients to treatment are likely to be negligible. Further discussion of the potential benefits of unequal allocation beyond statistical considerations can be found in Sposto and Krailo (1987).

4. A Normalizing Transformation for $\mathcal{L}_n$

It has been shown that the normal approximation to the distribution of $\mathcal{L}_n$ under the hypothesis H_0 : *no survival difference* is only accurate to $O(n^{-1/2})$ when there is imbalance in the treatment groups, and that both bias and skewness increase in absolute magnitude as p moves away from 1/2. The results of Section 3 can be used to improve the normal approximation to the distribution of the logrank statistic in this setting.

Let $T_{n,j}(\cdot),\ j = 1, 2$ be one of the two transformations defined in Theorem 1. Then, if p, R_0, and R_1 can be consistently estimated (see Corollary 1) under H_0, the results of Corollary 1 imply that the distribution of $T_{n,j}(\mathcal{L}_n)$ will be standard normal to $o(n^{-1/2})$. For any such transformation $T_n(\cdot)$, we may then reformulate the logrank test as

$$\text{Reject } H_0 \text{ if } \quad |T_n(\mathcal{L}_n)| > z_{1-\alpha/2}.$$

Other possibilities are available here; for example, we could have simply inverted the Edgeworth expansion of Corollary 1 to obtain a Cornish-Fisher expansion for the quantiles of the null distribution of $\mathcal{L}_n$. One theoretical advantage of this transformation-based approach is that the (approximate) distribution function of $T_{n,j}(\mathcal{L}_n)$ (i.e., the standard normal cdf) is monotone on $\mathbb{R}$, and thus avoids monotonicity problems usually associated with Edgeworth and Cornish-Fisher approximations; Hall (1992, §3.8) provides further discussion. A further advantage is that hypothesis tests and significance level calculations require no specialized probability computations and consequently are straightforward to carry out in practice.

The proposed transformations depend on the cumulants of $\mathcal{L}_n$; from Corollary 1, it is evident that one needs to be able to estimate p, R_0, and R_1 consistently under the null hypothesis in order to use these results in practice. The obvious estimator for p is $\widehat{p}$, the empirical proportion of patients in the group corresponding to $Z = 1$. Finding estimates of R_0 and R_1 is equally straightforward. Let $\overline{Y}(u) = \sum_i Y_i(u)$ and $\overline{N}(u) = \sum_i N_i(u)$ be the number at risk and the number of events at time u in the pooled sample. Let

Table 4. Simulation results (size): n=30, $P\{Z = 1\} = 0.2$.

		Exponential			Weibull		
% Censored	α	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$
63.1	0.025	0.0224	0.0211	0.0088	0.0227	0.0211	0.0089
	0.975	0.9731	0.9792	0.9489	0.9754	0.9819	0.9498
43.2	0.025	0.0276	0.0266	0.0139	0.0293	0.0284	0.0146
	0.975	0.9699	0.9747	0.9472	0.9719	0.9755	0.9484
31.7	0.025	0.0341	0.0328	0.0165	0.0321	0.0314	0.0163
	0.975	0.9694	0.9734	0.9470	0.9712	0.9753	0.9478
24.5	0.025	0.0321	0.0312	0.0177	0.0320	0.0309	0.0175
	0.975	0.9709	0.9740	0.9485	0.9701	0.9728	0.9469
20.0	0.025	0.0353	0.0344	0.0182	0.0341	0.0331	0.0182
	0.975	0.9692	0.9724	0.9458	0.9681	0.9715	0.9461

$\widehat{\pi}(u) = n^{-1}\overline{Y}(u)$, and define $\widehat{\Lambda}(t) = \int_0^t (\overline{Y}(u))^{-1} d\overline{N}(u)$ to be the Nelson-Aalen estimator based on the *pooled* data. Recalling that $R_k = \int_0^\tau \Lambda^k(u)\pi(u)d\Lambda(u)$, an obvious estimator for R_k is obtained by substituting in $\widehat{\pi}(u)$ for $\pi(u)$ and $\widehat{\Lambda}(u)$ for $\Lambda(u)$. Straightforward calculations show that

$$\widehat{R}_k = \frac{1}{n}\int_0^\tau \widehat{\Lambda}^k(u)d\overline{N}(u),$$

from which we obtain $\widehat{R}_0 = n^{-1}\sum_i D_i$ and $\widehat{R}_1 = n^{-1}\sum_i D_i\widehat{\Lambda}(X_i)$. Both can be easily calculated using standard software, and under the null hypothesis, $\widehat{R}_k \xrightarrow{P} R_k$ since $\sup_{t\in[0,\tau]} |\widehat{\pi}(t) - \pi(t)|$ and $\sup_{t\in[0,\tau]} |\widehat{\Lambda}(t) - \Lambda(t)|$ both converge in probability to zero (cf. Andersen *et al.* , 1993, §IV.1.2). Furthermore, $\widehat{p} = p + O_p(n^{-1/2})$ and $\widehat{R}_k = R_k + O_p(n^{-1/2})$; hence, the results of Theorem 1 apply as stated to $T_{n,j}(\mathcal{L}_n)$, $j = 1, 2$ under the null hypothesis of no survival difference.

A small simulation study to assess the accuracy of the normal approximation to $\mathcal{L}_n$ and its transformed versions was done. Failure time data were simulated under both exponential and Weibull models under various levels of uniform censoring for two different total sample sizes ($n = 30$ and 60). Individuals were assigned a covariate value of $Z = 1$ with probability $p = 0.2$ independently of survival/censoring. The results are summarized in Tables 4 and 5 below. The statistics considered are $T_{n,1}$, $T_{n,2}$, and $T_{n,3} = \mathcal{L}_n$; the entries in Tables 4 and 5 are empirical estimates of $P\{T_{n,i} \le z_\alpha\}$, $i = 1\ldots3$ for $\alpha = 0.025$ and 0.975 based on 25,000 simulated datasets.

At $n = 30$, the transformed methods generally do very well; the approximation in the upper tail is excellent, while that in the lower tail is not quite as good. The normal approximation is generally much less accurate in both tails. Results $T_{n,1}$ and $T_{n,2}$ at $\alpha = 0.05$, 0.95 (not reported) are even more encouraging. For $n = 60$, the normal approximation to the distribution of the transformed statistics is excellent in each case; the normal approximation to the distribution of $T_{3,n}$ leads to noticeably inferior results. In every case, the normal approximation to the distribution of $T_{3,n}$ leads to probabilities which are smaller

Table 5. Simulation results (size): $n = 60$, $P\{Z = 1\} = 0.2$.

		Exponential			Weibull		
% Censored	α	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$
63.1	0.025	0.0265	0.0256	0.0146	0.0266	0.0258	0.0137
	0.975	0.9749	0.9763	0.9608	0.9755	0.9769	0.9621
43.2	0.025	0.0281	0.0274	0.0180	0.0275	0.0267	0.0171
	0.975	0.9723	0.9739	0.9588	0.9741	0.9750	0.9607
31.7	0.025	0.0289	0.0284	0.0190	0.0266	0.0261	0.0165
	0.975	0.9736	0.9742	0.9603	0.9735	0.9743	0.9588
24.5	0.025	0.0284	0.0281	0.0174	0.0288	0.0282	0.0183
	0.975	0.9714	0.9722	0.9569	0.9736	0.9743	0.9588
20.0	0.025	0.0282	0.0278	0.0181	0.0301	0.0297	0.0184
	0.975	0.9731	0.9739	0.9613	0.9735	0.9742	0.9607

than the specified levels. This is in exact correspondence with what the Edgeworth analysis of Section 3.1 predicts should happen when $p < 1/2$.

We hasten to point out here that the cumulants being estimated are really first-order asymptotic approximations to the actual cumulants of $\mathcal{L}_n$. Use of alternative consistent estimators will not improve upon the asymptotic order of approximation; however, they may further improve the performance of $T_{n,j}(\mathcal{L}_n)$ in practice, particularly for smaller sample sizes. Using the fact that the logrank statistic can be expressed as a stochastic integral with respect to a martingale (cf. Fleming and Harrington , 1991, Ch. 7), one can devise alternative approximations via the Bartlett identities for martingales (Mykland, 1994). However, the resulting estimators are not nearly as simple to compute as the ones derived above (cf. Mykland and Ye, 1992).

5. Discussion

The first-order properties of the logrank statistic are well-understood and there exist numerous simulation studies describing the behavior of the test under various paradigms. This paper attempts to give a reasonably in-depth theoretical justification of this behavior through its second-order properties. It is, to the author's knowledge, the first paper to characterize the second-order local power properties of the logrank statistic under proportional hazards alternatives. The second-order asymptotic analyses of the preceding sections verify many behavioral characteristics observed by other authors, including the inadequacy of the standard normal approximation in the case of unbalanced data. Two transformations of $\mathcal{L}_n$ have also been proposed, both of which are extremely simple to compute and significantly improve the standard normal approximation to the distribution of the test statistic under the null hypothesis of no survival difference.

The second-order approximation to the power curve derived here is easy to compute and may prove useful in study design. For modest sample sizes the simulation results of

Section 3.2.2 show that the approximation (7) is reasonably accurate. Proper use of (7) in practice, such as computing the power for detecting a specific alternative based on the results of a clinical trial, is a more delicate matter. The assumptions leading to the derivation of the local power curve (7) dictate that $R_k,\ k = 0, 1$ are determined by the control group (i.e., $Z = 0$); all influence of the treatment group on power is then completely specified through the proportional hazards assumption. In view of this, a reasonable choice for estimating R_k (compare with Corollary 1) is $\widehat{R}_k = \int_0^\tau \widehat{\Lambda}_0^k(t)\widehat{\pi}(t)d\widehat{\Lambda}_0(t)$, where $\widehat{\pi}(\cdot) = \exp\{-\widehat{\Lambda}_0(t)\}\widehat{G}(t)$, $\widehat{\Lambda}_0(\cdot)$ is Breslow's estimator of the cumulative baseline hazard under the Cox proportional hazards regression model, and $\widehat{G}(\cdot)$ is the Kaplan-Meier estimator of the censoring distribution for the combined sample. The latter is a reasonable choice given our assumption that the censoring distribution is the same for both treatment groups. Further work on this problem and on improving the accuracy of (7) under fixed alternatives would be very valuable. Towards the latter, an adhoc approximation has been proposed here that has been shown to work very well in practice. The work of Pfanzagl and Pfaff (1985) may also prove useful here; see also Pfanzagl (1985, §6.3) for an interesting idea which couples local approximation techniques with the Hellinger distance between the distributions specified under the null and alternative hypotheses.

Theoretical justification for these results is confined to the case of iid data with bounded covariates; the latter certainly poses no restriction in the case of the logrank statistic. Generalizations of particular interest are to the case of unequal censoring and to the class of weighted logrank statistics (cf. Fleming and Harrington, 1991, Ch. 7). The general problem lies, of course, in the ability to establish the higher-order properties of these censored data statistics at all. One route to developing more general results, at least empirically, is to recognize that the majority of these results depend only on quantities which approximate the skewness and bias of the normalized and studentized test statistic under the hypothesis of no difference. Thus, devising approximations to these quantities yield an obvious method for potentially improving the small sample behavior of the class of weighted logrank tests under possibly unequal censoring and/or imbalance. Since weighted logrank statistics have martingale representations under the null hypothesis of no difference, the work of Mykland (1994) on Bartlett identities for martingales should prove useful here. In addition, the work of Mykland (1993, 1995) on asymptotic expansions for martingales provides some theoretical support for these conjectures of improved accuracy.

Appendix

Proof of Theorem 1: Let $\mathcal{S}_n = I^{-1/2}(\beta_0, \tau)U_n(\beta_0, \tau)$ where $\pi(\tau) > 0$. Then, under the assumptions of Section 2.1, Gu (1992) establishes (4) to $o(n^{-1/2})$ by first showing that $\mathcal{S}_n$ has a U-statistic representation of degree 2 plus error (cf. Gu, 1992, eqns. 4.6–4.11), and then directly applies the results of Bickel, Götze, and van Zwet (1986). Statistics having this representation were subsequently termed *asymptotic U-statistics* by Lai and Wang (1993), who derived Edgeworth expansions for such statistics under weaker conditions than those found in Bickel, Götze, and van Zwet (1986). Strawderman and Wells

(1997, Appendix A) extended the results of Lai and Wang (1993) to a general class of studentized statistics in which both the numerator and studentizing quantity have asymptotic U-statistic representations. Theorem A.1 of Strawderman and Wells (1997) establishes the U-statistic representation for the studentized statistic in the general case, and Corollary A.1 of Strawderman and Wells (1997) establishes the Edgeworth expansion for its distribution to $O(n^{-1})$. The asymptotic U-statistic representations for $n^{-1/2}U_n(\beta_0, \tau)$ and $I_n(\beta_0, \tau)$ required to apply the results of Strawderman and Wells (1997) are given in Lemmas 2.1 and 2.2 of Gu and Zheng (1993), who investigated Bartlett-type corrections for the partial likelihood ratio statistic. These representations allow direct application of Corollary A.1 of Strawderman and Wells (1997), yielding (4). Finally, given (4), (5) follows directly from equation (4.5) and Theorem 2 of Konishi (1991), which only depends upon the statistic of interest having an Edgeworth expansion of the form (4). ■

Proof of Theorem 2: We wish to establish an Edgeworth expansion for $\mathcal{S}_n$, the studentized score function evaluated at β_0, when $\beta_n = \beta_0 + n^{-1/2}\varepsilon$ is the true parameter value. This will be done using a higher-order version of LeCam's Third Lemma, and will allow one to evaluate the power of a score test for $H_0 : \beta = \beta_0$ based on the Cox partial likelihood function under the contiguous sequence of alternatives $H_{an} : \beta_n = \beta_0 + n^{-1/2}\varepsilon$ for $\varepsilon > 0$. In order for such a result to hold, it is certainly required that the analogous first-order result holds; that this is so under the assumptions of §2.1 follows from results found in Bickel, Klassen, Ritov, and Wellner (1993, pp. 330–335 and §A.9); see also Andersen, Borgan, Gill, and Keiding (1993, §VIII.4.3) and Slud (1987, §7E). The extension shall be established using results in Taniguchi (1992).

It assumed throughout that the assumptions of §2.1 are in effect, and that $\varepsilon = O(1) > 0$. Let $W = (X, \Delta, Z)$ (equivalently, $W = (N(t), Y(t), Z, t \in [0, \tau])$), and for any $\beta, \beta_n \in \mathcal{B}$ such that $\beta_n = \beta + n^{-1/2}\varepsilon$, respectively define $P_{n,\beta}$ and P_{n,β_n} to be the corresponding probability distributions of $(W_1, \ldots, W_n)$. Let $\Xi_n = \log dP_{n,\beta_n}/dP_{n,\beta}$ denote the corresponding loglikelihood ratio; note that this leads to the usual formula based on the ratio of partial loglikelihoods (cf. Andersen *et al.* , 1993, Thm. VII.2.1). Also, let $n^{-1/2}U_n(\beta)$ and $\mathcal{S}_n(\beta)$ respectively denote the score and the studentized score functions.

This being done, we now closely follow the sequence of arguments in §3 of Taniguchi (1992) to arrive at the main result. Without loss generality, fix $\beta_0 \in \mathcal{B}$ and let $\beta_n = \beta_0 + n^{-1/2}\varepsilon$. Define ν^2, Δ, and ζ as in (1)-(3), and let $Z_1 = n^{-1/2}U_n(\beta_0) \equiv \mathcal{S}_n$, $Z_2 = -n^{-1/2}\left(I_n(\beta_0) - n\nu^2\right)$, and $Z_3 = n^{-1/2}\left(l_n^{(3)}(\beta_0) - E_{\beta_0}[l_n^{(3)}(\beta_0)]\right)$, where $l_n^{(k)}(\beta)$ denotes the k^{th} derivative of $l_n(\beta)$ with respect to β and $E_{\beta_0}[\cdot]$ denotes expectation taken under P_{n,β_0}. Now, since $l^{(k)}(\beta)$ is continuously differentiable in β through $k = 5$, say, Assumption B1 of Taniguchi (1992) is easily met. In view of Lemma 2.4 of Gu and Zheng (1993), it is then easy to establish Assumption B3 of Taniguchi (1992). Combining the results of Gu (1992, Theorem 2.1), Theorem 1 above, and Gu and Zheng (1993, Lemmas 2.1–2.4), it follows $E_{\beta_0}[Z_1^2] = \nu^2 + O(n^{-1})$, $E_{\beta_0}[Z_1Z_2] = -\zeta + O(n^{-1})$, $E_{\beta_0}[Z_1^3] = -\Delta + 3\zeta$, $E_{\beta_0}[n^{-1}I_n(\beta_0)] = \nu^2 + O(n^{-1})$, and $E_{\beta_0}[n^{-1}l_n^{(3)}(\beta_0)] = \Delta + O(n^{-1})$, it being implicitly understood in the notation that ν^2, Δ, and ζ are evaluated at β_0. Furthermore,

$\text{cum}(Z_{i_1}, \ldots, Z_{i_J}) = O(n^{-(J+2)/2})$ for $J = 2, \ldots, 6$ and $i_k = 1$ or 2 for any k. To see this, it is helpful to note that this expression only involves sums of expectations of the form $E_{\beta_0}[Z_1^k Z_2^j]$ for $k + j = 2$, 3, and so on up to 6; the result then follows from standard properties of cumulants (e.g., Brillinger, 1975, §2.3 or McCullagh, 1987, Ch. 2). Hence, Assumption B2 of Taniguchi (1992) holds for the processes Z_j, $j = 1 \ldots 3$ defined earlier, and therefore by Lemma 2 of Taniguchi (1992), the loglikelihood ratio Ξ_n has the stochastic expansion $\widetilde{\Xi}_n + n^{-1}c_n$, where

$$\widetilde{\Xi}_n = -\frac{\varepsilon}{2}\nu^2 + \varepsilon Z_1 + \frac{\varepsilon^2}{2n^{1/2}} Z_2 - \frac{\varepsilon^3}{6n^{1/2}}\Delta$$

and c_n satisfies $P_{n,\beta_0}(|c_n| > k_n n^{1/2}) = O(n^{-1+\eta})$ for $0 < \eta < 1/2$ and k_n such that $k_n \to 0$ and $k_n n^{1/2} \to \infty$.

Now, expanding $I_n^{-1/2}(\beta_0)$ about $n\nu^2$, and multiplying the result by $n^{-1/2}U_n(\beta_0)$, we have that the *studentized* score function $\mathcal{S}_n$ has the stochastic expansion $\widetilde{Z}_n + n^{-1}c_n$, where

$$\widetilde{Z}_n = \frac{Z_1}{\nu} + n^{-1/2}\frac{Z_1 Z_2}{2\nu^3}$$

and c_n is as above. Thus, $\mathcal{S}_n$ is a member of the class of statistics $\mathcal{F}$ defined Taniguchi (1992, p. 215) with constants $a_1 = 0$ and $a_2 = (2\nu^3)^{-1}$. Since Assumptions B1-B3 of Taniguchi (1992) are met, the asymptotic cumulants of the 2×1 vector $(\widetilde{Z}_n, \widetilde{\Xi}_n)'$ satisfy the conditions specified in Proposition 3 of Taniguchi (1992) with (in his notation) $I = \nu^2$, $J = -\zeta$, and $K = -\Delta + 3\zeta$.

Thus far, it has been established that $\tilde{Z}_n$ and $\widetilde{\Xi}_n$ are respectively stochastic approximations to $\mathcal{S}_n(\beta_0)$ and $\log dP_{n,\beta_n}/dP_{n,\beta_0}q$ valid to the order discussed above. In addition, power series expansions for the cumulants of $(Z_n, \Xi_n)'$ valid to $o(n^{-1/2})$ have been established using $(\widetilde{Z}_n, \widetilde{\Xi}_n)'$. To complete the proof, we employ Theorem 4 of Taniguchi (1992), which is a direct application of Theorem 1 in the same paper. In view of the proof given in Taniguchi (1992, Theorem 1), we see that it is sufficient to verify that (A1) under P_{n,β_0}, $(\widetilde{Z}_n, \widetilde{\Xi}_n)'$ has an Edgeworth expansion which holds uniformly on $\mathbb{R}^2$ to $O(n^{-1+\eta})$ for $0 < \eta < 1/2$ and that (A2) $P_{n,\beta_n}\{|\Xi_n| > (1/2 + \delta)\log n^{1/2}\} = o(n^{-1/2})$ for $0 < \delta < \delta + \eta < 1/2$. The power series expansions for the cumulants discussed above, combined with the general results of Skovgaard (1986; see also Barndorff-Nielsen and Cox, 1989, Thm. 6.7) on multivariate Edgeworth expansions, can be used to verify (A1). We can establish (A2) by first noting that the definition of the loglikelihood ratio statistic allows us to rewrite this requirement as $E_{\beta_0}[\exp\{\Xi_n\}I\{|\Xi_n| > (1/2 + \delta)\log n^{1/2}] = o(n^{-1/2})$, and then use the representation for Ξ_n and the cumulant evaluations done earlier to obtain the final result. ■

Proof of Corollary 1: The main results, including the formulas of R_i, follow immediately from Theorem 1 in light of the simplifications of various expansion terms discussed at the end §2.1. To see that $0 \leq R_0 \leq R_1 \leq 1$, recall that $X = \min\{T, U\}$ for T, U independent. It is obvious that $R_i \geq 0$, $i = 0, 1$. Furthermore, letting $\mathcal{R}_k = \int_0^\infty \Lambda^k(t)P\{X \geq t\}\lambda(t)dt$, it can be shown that $\mathcal{R}_k = E[\Lambda^k(T)G(T)]$ for $G(u) = P\{U \geq u\}$. Noting that $\Lambda(T) \sim$ Exp(1), it is then easy to see that $\mathcal{R}_k \leq 1,\ k = 0, 1$. Thus, since R_k is simply $\mathcal{R}_k$ truncated to $[0, \tau]$, we see that $R_k \leq 1,\ k = 0, 1$. Finally, letting $H(u)$ denote the cdf of $\Lambda(U)$ and using the fact that we may write $R_k = \int_0^{\Lambda(\tau)} u^k e^{-u} P\{\Lambda(U) \geq u\} du$, it follows that

$$\begin{aligned}
R_0 - R_1 &= \int_0^{\Lambda(\tau)} (1-u)e^{-u} P\{\Lambda(U) \geq u\} du \\
&= \int_0^{\Lambda(\tau)} (1-u)e^{-u} \left(\int_u^{\infty} dH(s) \right) du \\
\text{(by Fubini's theorem)} &= \int_0^{\infty} \int_0^{\Lambda(\tau)} \left[(1-u)e^{-u} I\{s \geq u\} \right] du\, dH(s) \\
&= \int_0^{\infty} \int_0^{\min(\Lambda(\tau), s)} (1-u)e^{-u}\, du\, dH(s) \\
&= \int_0^{\infty} \min(\Lambda(\tau), s) e^{-\min(\Lambda(\tau), s)}\, dH(s)
\end{aligned}$$

which is nonnegative. ■

Acknowledgments

The partial support of NIH grant R01 DK49529 01 is gratefully acknowledged. The author would also like to thank Edsel Peña and Martin T. Wells for helpful discussions, and the LIDA editorial board and referees for a prompt review.

References

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

P. K. Andersen, and R. D. Gill, "Cox's regression model for counting processes: A large sample study," *Annals of Statistics* vol. 10 pp. 1100–1120, 1982.

O. E. Barndorff-Nielsen and D. R. Cox, *Asymptotic Techniques for Use in Statistics*, Chapman Hall: London, UK, 1989.

P. J. Bickel, C. Klassen, Y. Ritov, and J. Wellner, *Efficient and Adaptive Estimation in Semi-parametric Models*, Johns Hopkins University Press: Baltimore, MD, 1993.

P. J. Bickel, F. Götze, and W. R. van Zwet, "The Edgeworth expansion for U-statistics of degree two," *Annals of Statistics* vol. 14 pp. 1463–1484, 1986.

D. Brillinger, , *Time Series: Data Analysis and Theory*, Holden-Day: San Francisco, CA, 1981.

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

T. DiCiccio, and R. Tibshirani, "Bootstrap confidence intervals and bootstrap approximations," *Journal of the American Statistical Association* vol. 82 pp. 163–170, 1987.

R. J. M. M. Does, "An Edgeworth expansion for simple linear rank statistics under the null-hypothesis," *Annals of Statistics* vol. 11 pp. 607–624, 1983.

B. Efron, "Better bootstrap confidence intervals," *Journal of the American Statistical Association* vol. 82 pp. 171–200, 1987.

T. R. Fleming, and D. P. Harrington, *Counting Processes and Survival Analysis*, John Wiley and Sons: New York, NY, 1991.

M. Gu, “On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship,” *Canadian Journal of Statistics* vol. 20 pp. 399–414, 1992.

M. Gu, and Z. Zheng, “On the Bartlett adjustment for the partial likelihood ratio test in the Cox regression model,” *Statistica Sinica* vol. 3 pp. 543–555, 1992.

P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag: New York, NY, 1992.

F. Y. Hsieh, “A simple method of sample size calculation for unequal-sample-size designs that use the logrank or t-test,” *Statistics in Medicine* vol. 6 pp. 577–581, 1987.

F. Y. Hsieh, “Comparing sample size formulae for trials with unbalanced allocation using the logrank test,” *Statistics in Medicine* vol. 11 pp. 1091–1098, 1992.

J. Kalbfleisch, and R. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley and Sons: New York, NY, 1980.

L. Kalish, and D. Harrington, “Efficiency of balanced treatment allocation for survival analysis,” *Biometrics* vol. 44 pp. 815–821, 1988.

A. M. Kellerer, and D. Chmelevsky, “Small-sample properties of censored-data rank tests,” *Biometrics* vol. 39 pp. 675–682, 1983.

S. Konishi, “Normalizing transformations and bootstrap confidence intervals,” *Annals of Statistics* vol. 19 pp. 2209–2225, 1991.

T. L. Lai, and J. Q. Wang, “Edgeworth expansions for symmetric statistics with applications to bootstrap methods,” *Statistica Sinica* vol. 3 pp. 517–542, 1993.

E. Lakatos, and G. Lan, “A comparison of sample size methods for the logrank statistic,” vol. 11 pp. 179–191, 1992.

R. Latta, “A Monte Carlo study of some two-sample rank tests with censored data,” *Journal of the American Statistical Association* vol. 76 pp. 713–729, 1981.

J. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons: New York, NY, 1982.

P. McCullagh, *Tensor Methods in Statistics*, Chapman Hall: London, UK, 1987.

P. Mykland, “Asymptotic expansions for martingales,” *Annals of Probability* vol. 21 pp. 800–818, 1993.

P. Mykland, “Bartlett identities for martingales,” *Annals of Statistics* vol. 22 pp. 21–38, 1994.

P. Mykland, “Martingale expansions and second-order inference,” *Annals of Statistics* vol. 23 pp. 707–731, 1995.

P. Mykland, and J. Ye, “Cumulant and Bartlett identities in Cox regression,” Department of Statistics Technical Report No. 332, University of Chicago, Chicago IL, 1992.

F. D. Nelson, and N. E. Savin, “The danger of extrapolating asymptotic local power,” *Econometrica* vol. 58 pp. 977–981.

R. Peto, M. C. Pike, P. Armitage, N. E. Breslow, D. R. Cox, S. Howard, N. Mantel, K. McPherson, J. Peto, and P. Smith, “Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II: Analysis and examples,” vol. 35 pp. 1–39, 1977.

T. Pfaff, and J. Pfanzagl, “On the accuracy of asymptotic expansions for power functions,” *Journal of Statistical Computation and Simulation* vol. 22 pp. 1–25.

J. Pfanzagl, *Asymptotic Expansions for General Statistical Models (With the Assistance of W. Wefelmeyer)*, Lecture Notes in Statistics, Springer-Verlag: Berlin, 1985.

R. Prentice, “Linear rank tests with right censored data,” *Biometrika* vol. 65 pp. 167–179, 1978.

R. Prentice, and P. Marek, “A qualitative discrepancy between censored data rank tests,” *Biometrics* vol. 35 pp. 861–867, 1979.

C. R. Rao, and R. Mukerjee, “Comparison of Bartlett-type adjustments for the efficient score statistic,” *Journal of Statistical Planning and Inference* vol. 46 pp. 137–146, 1995.

I. R. Savage, “Contributions to the theory of rank order statistics—the two sample case,” *Annals of Mathematical Statistics* vol. 27 pp. 590–615.

D. Schoenfeld, “The asymptotic properties of nonparametric tests for comparing survival distributions,” *Biometrika* vol. 68 pp. 316–319, 1981.

I. M. Skovgaard, “On multivariate edgeworth expansions,” *International Statistical Review* vol. 54 pp. 169–186, 1986.

E. V. Slud, *Martingale Methods in Statistics*, manuscript preprint, University of Maryland, 1987.

R. Sposto, and M. Krailo, “Use of unequal allocation in survival trials,” *Statistics in Medicine* vol. 6 pp. 119–125, 1987.

R. L. Strawderman, and M. T. Wells, “ Accurate bootstrap confidence limits for the cumulative hazard and survivor functions under random censoring,” *Journal of the American Statistical Association* 1997, to appear.

M. Taniguchi, “An automatic formula for the second-order approximation of the distributions of test statistics under contiguous alternatives,” *International Statistical Review* vol. 60 pp. 211–225, 1992.

A. A. Tsiatis, “A large sample study of Cox’s regression model,” *Annals of Statistics* vol. 9 pp. 93–108, 1981.