Biometrics & Biostatistics

Wu, J Biomet Biostat 2014, 5:4
http://dx.doi.org/10.472/2155-6180.1000210

Research Article Open Access

A New One-Sample Log-Rank Test

Jianrong Wu*

Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

Abstract

The one-sample log-rank test has been frequently used by epidemiologists to compare the survival of a sample to that of a demographically matched standard population. Recently, several researchers have shown that the one-sample log-rank test is conservative. In this article, a modified one-sample log-rank test is proposed and a sample size formula is derived based on its exact variance. Simulation results showed that the proposed test preserves the type I error well and is more efficient than the original one-sample log-rank test.

Keywords: Epidemiology; One-sample log-rank test; Time-to-event; Sample size; Standard population

Introduction

Two-sample log-rank tests are frequently used to design and make inferences for randomized phase III survival trials with two treatment arms. The primary aim of such a study is to compare the survival distributions between two treatment groups. In some cases, it is also interested in comparing the survival distribution of a single sample to that of a standard population. Such comparison arises naturally in epidemiologic studies and clinical trials. For example, in an epidemiologic study, in which the survival data of patients with a life-threatening disease have been prospectively collected, it may be of interest to know if the study sample experiences better survival than the demographically matched standard population. It is not appropriate to use the two-sample log-rank test to make this comparison because the variance could be overestimated; thus, the p-value from the two-sample log-rank test is invalid. However, an analog test statistic called the one-sample log-rank test [1] can be used for such study design and comparison.

There is relatively little literature available to design and make inferences for comparing the survival of a sample to a standard population. The one- sample log-rank test was first introduced by Breslow [2]. Its asymptotic property has been studied by Hyde [3], Anderson et al. [4], and Gill and Ware [5], and applications can be found in Finkelstein et al. [1], Berry [6], Woolson [7], and Anderson et al. [4]. Study designs using the one-sample log-rank test were considered by Finkelstein et al. [1]. Kwak and Jung [8], Jung [9], and Sun et al. [10] applied it to single-arm phase II clinical trial designs.

If a study is planned to determine whether the survival of the new study participants better than that of a standard population, then the study must be carefully designed to ensure sufficient power to detect a specific difference of the survival distributions. For the study design, a sample size formula of the one-sample log-rank test is given by Finkelstein et al. [1]. Kwak and Jung [8] proposed another sample size formula for single-arm phase II clinical trial design using the one-sample log-rank test. Wu [11] recently derived a new sample size formula based on its exact variance. However, simulation results done by Kwak and Jung [8], Sun et al. [10] and Wu [11] have shown that the one-sample log-rank test is conservative, even when the sample size is relatively large. Thus, it is necessary to develop a new test statistic that preserves the type I error rate and keeps the power as high as possible. Sun et al. [10] derived two corrections of the one-sample log-rank test statistics based on its Edgeworth expansion. However, a major drawback of their corrected tests is that they are more complicated test statistics involving higher-order moment estimations, which makes it difficult to derive their distributions under the alternative. Thus, they can't be used for the study design.

Here we propose a new and simple one-sample log-rank test to correct the conservativeness of the original one-sample log-rank test. A sample size formula is also derived for the new test for the purpose of the study design. The rest of the article is organized as follows. In Section 2, a new one-sample log-rank test is proposed. A sample size formula is derived in Section 3. In Section 4, simulation studies are conducted to compare the empirical type I error and power among four test statistics. An example is given in Section 5. Concluding remarks are given in Section 6.

One-Sample Log-Rank Tests

The one-sample log-rank test was first introduced by Breslow [2], and it has been used frequently by epidemiologists [3]. To introduce the one-sample log- rank test, let $\Lambda_0(x)$ and $S_0(x)$ be the known cumulative hazard and survival functions for the standard population, and let $\Lambda(x)$ and $S(x)$ be the unknown cumulative hazard and survival functions for the new study. Then the study may consider the following hypothesis of interest:

$$H_0: S(x) \le S_0(x) \text{ vs. } S(x) > S_0(x),$$

or an equivalent to the hypothesis, in terms of cumulative hazard function

$$H_0: \Lambda(x) \le \Lambda_0(x) \text{ vs. } \Lambda(x) > \Lambda_0(x).$$

Suppose during the accrual phase of the trial n subjects are enrolled in the study. Let T_i and C_i denote, respectively, the failure time and censoring time of the i^{th} subject. We assume that the failure time T_i and censoring time C_i are independent and $\{T_i, C_i, i=1,...,n\}$ are independent and identically distributed. Then the observed failure time and failure indicator are $X_i = T_i \wedge C_i$ and $\Delta_i = I(T_i \le C_i)$, respectively, for i^{th} subject. On the basis of the observed data $\{X_i, \Delta_i, i = 1, \cdots, n\}$, we define

***Corresponding author:** Jianrong Wu, Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA, Tel: 901-595-2850; Fax: 901-595-8843; E-mail: jianrong.wu@stjude.org

Received July 30, 2014; **Accepted** August 15, 2014; **Published** August 20, 2014

Citation: Wu J (2014) A New One-Sample Log-Rank Test. J Biomet Biostat 5: 210. doi:10.4172/2155-6180.1000210

Citation: Wu J (2014) A New One-Sample Log-Rank Test. J Biomet Biostat 5: 210. doi:10.4172/2155-6180.1000210

$O=\sum_{i=1}^{n}\Delta_i$ as the observed number of events, and $E=\sum_{i=1}^{n}\Lambda_0(X_i)$ as the expected number of events (asymptotically), then the one-sample log-rank test is defined by

$$L_1=\frac{O-E}{\sqrt{E}}. \quad (1)$$

To study the asymptotic distribution of the one-sample log-rank test statistic, we formulate it using counting-process notations [12].

Specifically, let $N_i(x)=\Delta_i I\{X_i\le x\}$ and $Y_i(x)=I\{X_i\ge x\}$ be the failure and at-risk processes, respectively, then

$$O=\sum_{i=1}^{n}\int_0^\infty dN_i(x),\ E=\sum_{i=1}^{n}\int_0^\infty Y_i(x)d\Lambda_0(x).$$

Thus, the counting-process formulation of the one-sample log-rank test is given by

$$L_1=W/\hat{\sigma},$$

where

$$W=n^{-1/2}\sum_{i=1}^{n}\int_0^\infty \{dN_i(x)-Y_i(x)d\Lambda_0(x)\},$$

and

$$\hat{\sigma}^2=n^{-1}\sum_{i=1}^{n}\int_0^\infty Y_i(x)d\Lambda_0(x)$$

Under the null hypothesis $H_0: n^{-1}\sum_{i=1}^{n}Y_i(x)\to G(x)S_0(x)$, where $G(x)$ is the survival distribution of censoring time C. Thus, $\hat{\sigma}^2$ converges to $\upsilon^2=\int_0^\infty G(x)S_0(x)d\Lambda_0(x)$, which is the exact variance of W under the null hypothesis. As showed in the Appendix, the exact mean of W under the null is $E_{H_0}(W)=0$. Therefore, by counting process central limit theorem [12], under the null hypothesis, L_1 is asymptotically standard normal distribution. Hence, we reject the null hypothesis H_0 with one-sided type I error α if $L_1=W/\hat{\sigma}<-z_{1-\alpha}$, where $z_{1-\alpha}$ is the 100 (1 − α) percentile of the standard normal distribution.

Simulation results showed, however, that the one-sample log-rank test L_1 is conservative, even when the sample size is relatively large [8-11]. For example, the empirical type I error of L_1 could be as low as 0.036 for a one-sided type I error rate of 0.05 (Table 1). To preserve the type I error, Sun et al. [10] derived two corrections based on Edgeworth expansion which are given below. Let $\hat{\gamma}_0=n^{-1}\sum_{i=1}^{n}\Lambda_0(X_i)$, $\hat{\gamma}_1=(2n)^{-1}\sum_{i=1}^{n}\Lambda_0^2(X_i)$, $\hat{k}_{11}=\hat{\gamma}_1\hat{\gamma}_0^{-3/2}$, and $\hat{k}_{12}=\hat{\gamma}_0^{-1/2}$. Two corrected one-sample log-rank tests are given by

$$L_2=K_n-\frac{1}{\sqrt{n}}\left\{\frac{1}{2}\hat{k}_{11}+\frac{1}{6}\hat{k}_{12}(K_n^2-1)\right\}$$

and

$$L_2=\frac{1}{\hat{\xi}}\left\{\exp(\hat{\xi}K_n)-1\right\}+\frac{1}{\sqrt{n}}\left\{\frac{1}{2}\hat{k}_{11}+\frac{1}{6}\hat{k}_{12}\right\},$$

w.here $K_n=L_1$ and $\hat{\xi}=-\hat{k}_{12}/(3\sqrt{n})$. Note that Sun et al. [10] defined $K_n=-L_1$, whereas our simulation results showed that it should be $K_n=L_1$. A major drawback of the two corrected tests is that they are more complicated test statistics involving higher-order moment estimations, which makes it difficult to derive their distributions under the alternative. Thus, they cannot be used for the study design.

Since $n^{-1}\to E_{H_0}(\Lambda(X))$ and $n^{-1}O\to E_{H_0}(\Delta)$, and $E_{H_0}(\Lambda(X))=E_{H_0}(\Delta)=\mathrm{Var}_{H_0}(W)$ as shown in the Appendix, thus, to correct the conservativeness of the original one-sample log-rank test L_1, we propose a new one-sample log-rank test which is defined as

$$L_4=\frac{O-E}{\sqrt{(O+E)/2}}. \quad (2)$$

In counting-process formulation, it is given by

$$L_4=W/\hat{\upsilon},$$

where

$$W=n^{-1/2}\sum_{i=1}^{n}\int_0^\infty \{dN_i(x)-Y_i(x)d\Lambda_0(x)\}$$

and

$$\hat{\upsilon}^2=n^{-1/2}\sum_{i=1}^{n}\int_0^\infty \{dN_i(x)+Y_i(x)d\Lambda_0(x)\}/2.$$

As shown in the Appendix, under the null hypothesis,

$$\hat{\upsilon}^2\to\upsilon^2=\{E_{H_0}(\Delta)+E_{H_0}(\Lambda(X))\}/2=\mathrm{Var}_{H_0}(W).$$

Therefore, again by counting-process central limit theorem under the null hypothesis, L_4 is asymptotically standard normal distribution. Hence, we reject the null hypothesis H_0 if $L_4=W/\hat{\upsilon}<-z_{1-\alpha}$.

Simulation studies are conducted in Section 4 to compare the empirical type I error and power of the original one-sample log-rank test L_1 to that of the two corrections L_2 and L_3, and the new test L_4.

Sample Size Calculation

To design the study, sample size must be calculated to detect a specified survival difference at the alternative $\Lambda(t)=\Lambda_1(t)(<\Lambda_0(t))$, given the type I error α and power 1−β. For the sample size calculation, the exact variance of W has been derived by Wu [11]. Let the exact mean and variance of W at the alternative be $E_{H_1}(W)=\sqrt{n}\omega$ and $\mathrm{Var}_{H_1}(W)=\sigma^2$, respectively, where ω and σ² are given in the Appendix. By central limit theorem, $(W-\sqrt{n}\omega)/\sigma$ is approximately standard normal distribution under H_1. Under the alternative hypothesis,

$$\hat{\sigma}^2=n^{-1}\sum_{i=1}^{n}\int_0^\infty Y_i(x)d\Lambda_0(x)\to\sigma_0^2=\int_0^\infty G(x)S_1(x)d\Lambda_0(x),$$

and the power of the one-sample log-rank test $L_1=W/\hat{\sigma}$ should satisfy the following equations:

$$1-\beta=P(L_1<-z_{1-\alpha})\simeq P\left(\frac{W-\sqrt{n}\omega}{\sigma}<-\frac{\sigma_0}{\sigma}z_{1-\alpha}-\frac{\sqrt{n}\omega}{\sigma}\,|\,H_1\right)$$
$$\simeq\Phi\left(\frac{\sigma_0}{\sigma}z_{1-\alpha}-\frac{\sqrt{n}\omega}{\sigma}\right).$$

Therefore, the required sample size for the test statistic L_1 is given by

$$n=\frac{(\sigma_0 z_{1-\alpha}+\sigma z_{1-\beta})^2}{\omega^2},$$

where $\omega=\sigma_1^2-\sigma_0^2$ and $\sigma^2=p_1-p_1^2+2p_{00}-p_0^2-2p_{01}+2p_0p_1$, with $\sigma_0^2,\sigma_1^2,p_0p_1,p_{00}$ and p_{01} given in the Appendix.

Similarly, under the alternative, $\hat{\upsilon}^2\to\bar{\sigma}^2=(\sigma_1^2+\sigma_0^2)/2$ (see Appendix); thus, the power of the new one-sample log-rank test $L_4=W/\hat{\upsilon}$ should satisfy the following equations:

Citation: Wu J (2014) A New One-Sample Log-Rank Test. J Biomet Biostat 5: 210. doi:10.4172/2155-6180.1000210

$$1-\beta = P(L_4 < -z_{1-\alpha}) \simeq P\left(\frac{W-\sqrt{n}\omega}{\sigma} < -\frac{\bar{\sigma}}{\sigma} z_{1-\alpha} - \frac{\sqrt{n}\omega}{\sigma} \mid H_1\right)$$

$$\simeq \Phi\left(\frac{\bar{\sigma}}{\sigma} z_{1-\alpha} - \frac{\sqrt{n}\omega}{\sigma}\right).$$

Therefore, the required sample size for test statistic L_4 is given by

$$n = \frac{(\bar{\sigma} z_{1-\alpha} + \bar{\sigma} z_{1-\beta})^2}{\omega^2}$$

where $\bar{\sigma}^2, \sigma^2$, and ω are the same as given above.

Simulation Studies

To study the performance of the two one-sample log-rank tests and their sample size formulas, we conducted simulation studies to compare the empirical power and type I error under different scenarios. In simulation studies, the survival distribution of the standard population was taken as the Weibull distribution $S_0(x) = e^{-\log(2)(x/m_0)^{\kappa}}$, or cumulative hazard function $\Lambda_0(x) = \log(2)(x/m_0)^{\kappa}$, with a known shape parameter κ and median survival time m0 under the null. Assume that the cumulative hazard function at the alternative is $\Lambda_1(x) = \log(2)(x/m_1)^{\kappa}$ with a common shape parameter κ, where the median survival time under the alternative $m_1 > m_0$. Therefore, the underlying Weibull model is a proportional hazards model with hazard ratio $\delta = (m_1/m_0)^{\kappa}$. The parameter settings for the simulation studies were set to κ=0.1, 0.25, 1, 2, and 5 to reflect cases of decreasing (κ<1), constant (κ=1) and increasing (κ>1) hazard functions. The hazard ratio δ under the alternative hypothesis was set to 1.2–2.0, with other parameters fixed as follows: $m_0=1$, accrual period $t_a=3$, and follow-up time $t_f=1$.

We assumed that subjects were recruited with a uniform distribution over the accrual period t_a and followed for t_f. We further assumed that no subject was lost to follow-up or drop-out during the study. Then the censoring time is uniformly distributed on the interval $[t_f, t_a+t_f]$. Thus, under the Weibull model, quantities p_0, p_1, p_{00}, and p_{01}, hence $\sigma_0^2, \sigma_1^2, \omega, \sigma^2$ can be calculated by numerical integrations. Given the nominal significance level of 0.05 and power of 90%, the required sample sizes for each design scenario were calculated for test statistics L_1 and L_4 (Table 1). The empirical type I error and power for the corresponding design were also simulated based on 100,000 samples generated from the Weibull distribution (Table 1). To compare the four test statistics, we also simulated the empirical type I error and power of the four test statistics L_1–L_4 given the same sample size n=30, 50, 100, and 200 (Table 2).

The sample size calculation (Table 1) showed that the original one-sample log-rank test L_1 required a larger sample size than that of the new test L_4. The simulated empirical type I errors for the corresponding sample size showed that the type I error of L_1 was always less than the nominal level. Thus, the original one-sample log-rank test L_1 was conservative. The empirical type I errors of the new test L_4 were close to the nominal level in most scenarios and were slightly liberal when the sample size was small. The simulation results in Table 2 with the same sample size further confirmed that the test L_1 was conservative and that L_4 preserved the type I error well and had a higher power than that of the L_1. It is consistent with the results from sample size calculations that L_4 had a smaller sample size than did L_1. Simulations were also done for the two corrected tests L_2 and L_3. The results showed that L_2 preserved the type I error well and had a higher power than L_1 and L_2, and L_3 was slightly conservative when sample size was small. Furthermore, the empirical type I error and power of test L_4 were also comparable to the two corrections L_2 and L_3.

		δ=1.2			δ=1.3			δ=1.4		
κ	Test	n	α	1−β	n	α	1−β	n	α	1−β
0.1	L_1	534	.048	.903	269	.046	.906	169	.044	.907
	L_4	508	.051	.897	250	.051	.896	155	.053	.893
0.5	L_1	432	.047	.905	217	.046	.907	137	.046	.909
	L_4	411	.051	.899	203	.052	.901	125	.053	.897
1.0	L_1	356	.047	.907	178	.045	.909	112	.044	.912
	L_4	339	.050	.904	167	.050	.903	103	.049	.905
2.0	L_1	306	.046	.910	153	.043	.915	97	.042	.922
	L_4	292	.049	.907	144	.049	.910	89	.048	.913
5.0	L_1	288	.046	.912	144	.044	.917	91	.042	.925
	L4	275	.050	.909	135	.049	.912	84	.049	.916
		δ=1.5			δ=1.6			δ=1.7		
κ	Test	n	α	1−β	n	α	1−β	n	α	1−β
0.1	L_1	121	.045	.908	93	.044	.909	75	.043	.911
	L_4	109	.053	.897	82	.052	.893	66	.052	.894
0.5	L_1	97	.044	.912	75	.042	.913	60	.043	.910
	L_4	88	.053	.900	66	.053	.898	53	.053	.900
1.0	L_1	80	.043	.916	61	.042	.916	49	.041	.919
	L_4	72	.051	.904	55	.050	.907	44	.051	.908
2.0	L_1	69	.042	.927	53	.040	.929	43	.040	.934
	L_4	63	.049	.918	47	.050	.916	38	.049	.921
5.0	L_1	65	.040	.930	50	.039	.935	40	.040	.937
	L_4	59	.049	.919	45	.049	.924	36	.048	.928
		δ=1.8			δ=1.9			δ=2.0		
	Test	n	α	1−β	n	α	1−β	n	α	1−β
0.1	L_1	63	.041	.911	54	.042	.911	47	.041	.909
	L_4	54	.055	.893	46	.056	.891	40	.055	.892
0.5	L_1	50	.041	.912	43	.041	.913	38	.041	.915
	L_4	44	.055	.902	37	.053	.897	32	.054	.894
1.0	L_1	41	.040	.921	35	.040	.921	31	.040	.925
	L_4	36	.051	.908	31	.052	.911	27	.052	.912
2.0	L_1	36	.038	.938	31	.038	.940	27	.038	.942
	L_4	31	.048	.920	27	.050	.925	23	.049	.922
5.0	L_1	34	.040	.943	29	.038	.945	25	.036	.943
	L_4	30	.048	.930	25	.048	.929	22	.048	.932

Table 1: Sample size, simulated empirical type I error (α), and power (1−β) of test statistics L_1 and L_4 based on 100,000 simulation runs from the Weibull distribution with nominal type I error of 0.05 and power of 90% (one-sided test).

To compare the null distribution functions of the four test statistics to the standard normal for small sample sizes, we conducted 100,000 simulation runs to simulate the empirical distribution functions of L_1–L_4 under the null with sample size n=30 to 200 (Table 3). The simulation results showed that the distribution of L_1 had a light left tail, while L_4 had a slightly heavier left tail than a standard normal distribution function. The results explained the observations from previous simulations that the test L_1 was conservative and L_4 was slightly liberal when the sample size was small. The distribution of L_2 was almost the same as the standard normal distribution function, and the distribution of L_3 had a slightly lighter left tail when sample size was small. Overall, L_4 preserved type I error well and had power higher than that of L_1–L_3. The distribution function of L_4 was also close to the standard normal and comparable to that of L_2 and L_3. The major advantage of L_4 is its simplicity and ease with which it derives the asymptotic distribution under the alternative. Therefore, the proposed new one-sample log-rank test L_4 is preferred for the study design and data analysis of a study comparing the survival of a sample to that of the standard population.

			δ									
κ	n	Test	1.0	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.5	30	L_1	.040	.169	.264	.369	.479	.577	.665	.737	.799	.846
		L_2	.049	.197	.299	.411	.523	.622	.704	.774	.829	.870
		L_3	.046	.190	.290	.400	.512	.612	.695	.765	.821	.864
		L_4	.055	.210	.317	.430	.539	.636	.719	.783	.839	.879
	50	L_1	.042	.241	.388	.544	.677	.784	.863	.912	.945	.968
		L_2	.051	.267	.422	.575	.708	.807	.878	.926	.955	.973
		L_3	.049	.260	.414	.567	.701	.801	.874	.923	.953	.972
		L_4	.054	.279	.435	.591	.718	.817	.887	.930	.957	.975
	100	L_1	.043	.399	.635	.812	.919	.967	.988	.996	.999	1
		L_2	.050	.420	.656	.831	.926	.972	.990	.996	.999	1
		L_3	.048	.414	.651	.827	.924	.971	.989	.996	.999	1
		L_4	.051	.431	.665	.833	.930	.973	.991	.997	.999	1
	200	L_1	.046	.635	.885	.976	.996	1	1	1	1	1
		L_2	.050	.651	.893	.979	.997	1	1	1	1	1
		L_3	.049	.647	.891	.978	.996	1	1	1	1	1
		L_4	.051	.656	.896	.979	.997	1	1	1	1	1
1	30	L_1	.039	.193	.316	.441	.569	.673	.760	.827	.879	.916
		L_2	.049	.226	.356	.487	.609	.715	.796	.856	.900	.932
		L_3	.043	.207	.331	.461	.583	.693	.778	.841	.889	.924
		L_4	.051	.232	.365	.492	.619	.718	.797	.858	.903	.933
	50	L_1	.041	.281	.460	.631	.768	.861	.924	.959	.979	.988
		L_2	.050	.308	.493	.663	.794	.879	.935	.966	.982	.991
		L_3	.045	.291	.473	.644	.780	.869	.929	.962	.980	.990
		L_4	.051	.317	.501	.669	.797	.882	.938	.967	.983	.991
	100	L_1	.044	.461	.718	.884	.959	.988	.997	.999	1	1
		L_2	.051	.487	.738	.894	.964	.990	.997	.999	1	1
		L_3	.047	.473	.726	.887	.962	.989	.997	.999	1	1
		L_4	.052	.490	.741	.897	.965	.990	.997	.999	1	1
	200	L_1	.046	.716	.935	.992	.999	1	1	1	1	1
		L_2	.051	.732	.941	.992	.999	1	1	1	1	1
		L_3	.048	.725	.939	.992	.999	1	1	1	1	1
		L_4	.051	.734	.942	.993	.999	1	1	1	1	1
2	30	L_1	.037	.220	.363	.514	.647	.758	.836	.894	.933	.959
		L_2	.051	.262	.413	.560	.694	.792	.867	.916	.948	.967
		L_3	.040	.225	.369	.516	.652	.760	.843	.898	.936	.959
		L_4	.048	.256	.407	.557	.687	.791	.862	.911	.945	.967
	50	L_1	.041	.317	.526	.709	.838	.916	.961	.982	.992	.997
		L_2	.050	.354	.564	.739	.859	.931	.969	.986	.994	.998
		L_3	.041	.322	.530	.711	.839	.919	.963	.982	.992	.997
		L_4	.050	.349	.561	.738	.858	.928	.968	.985	.993	.997
	100	L_1	.042	.519	.789	.929	.981	.996	.999	1	1	1
		L_2	.051	.551	.807	.937	.984	.996	.999	1	1	1
		L_3	.045	.527	.791	.930	.981	.996	.999	1	1	1
		L_4	.049	.546	.807	.937	.983	.996	.999	1	1	1
	200	L_1	.044	.781	.966[19]	.997	1	1	1	1	1	1
		L_2	.050	.796	.968	.997	1	1	1	1	1	1
		L_3	.046	.784	.965	.997	1	1	1	1	1	1
		L_4	.049	.795	.969	.998	1	1	1	1	1	1

Table 2: Simulation studies for empirical type I error (δ=1) and power (δ>1) of four test statistics, L_1−L_4, based on 100,000 simulation runs from the Weibull distribution with nominal type I error of 0.05 (one-sided test).

An Example

This example, Example V.1.5, is taken from Anderson et al. [4]. During the period 1962-1977, 205 patients with malignant melanoma had a radical operation performed at the Department of Plastic Surgery, University Hospital of Odense, Demark. A total of 57 patients died of malignant melanoma, 14 died of other causes; and the remaining 134 patients were alive as of January 1, 1978. If one is interested in studying deaths due to causes other than malignant melanoma and comparing those data to the standard life tables for the Danish population during 1971-1975, then using classical one-sample log-rank test, there are *O*=14 observed deaths versus *E*=21.244 expected deaths (see Anderson et al., page 338), yielding an observed value of the test statistic $L_1 = (O-E)/\sqrt{E} = -1.57$, which is not significant compared to $-z_{1-\alpha} = -1.645$ for the significance level α=0.05. However, the new

			x						
κ	n	Test	-3.0	-1.96	-0.67	0.0	0.67	1.96	3.0
0.5	30	L_1	.0003	.0169	.2428	.4949	.7352	.9632	.9959
		L_2	.0013	.0242	.2539	.4987	.7442	.9767	.9991
		L3	.0012	.0228	.2450	.4888	.7368	.9748	.9989
		L_4	.0021	.0285	.2504	.4949	.7440	.9783	.9993
	50	L_1	.0006	.0190	.2446	.4958	.7412	.9669	.9964
		L_2	.0013	.0251	.2524	.4997	.7498	.9753	.9991
		L_3	.0012	.0240	.2461	.4920	.7437	.9742	.9989
		L_4	.0021	.0283	.2506	.4958	.7477	.9771	.9991
	100	L_1	.0008	.0210	.2470	.4974	.7430	.9692	.9977
		L_2	.0012	.0254	.2527	.4995	.7481	.9756	.9989
		L_3	.0011	.0245	.2479	.4942	.7438	.9748	.9988
		L_4	.0019	.0280	.2512	.4974	.7475	.9770	.9989
	200	L_1	.0008	.0210	.2480	.4969	.7447	.9702	.9978
		L_2	.0012	.0252	.2527	.4999	.7492	.9754	.9988
		L_3	.0012	.0246	.2491	.4960	.7461	.9748	.9987
		L_4	.0016	.0259	.2512	.4969	.7479	.9758	.9988
1	30	L_1	.0005	.0167	.2374	.4870	.7334	.9628	.9961
		L_2	.0011	.0248	.2517	.4999	.7464	.9756	.9992
		L3	.0009	.0210	.2319	.4750	.7291	.9724	.9989
		L_4	.0019	.0266	.2440	.4870	.7412	.9771	.9994
	50	L_1	.0005	.0192	.2427	.4908	.7367	.9668	.9969
		L_2	.0012	.0251	.2532	.5001	.7458	.9754	.9989
		L_3	.0010	.0221	.2382	.4814	.7316	.9728	.9988
		L_4	.0018	.0271	.2480	.4908	.7430	.9770	.9991
	100	L_1	.0008	.0199	.2460	.4956	.7415	.9695	.9977
		L_2	.0013	.0250	.2514	.4995	.7466	.9748	.9988
		L_3	.0011	.0232	.2404	.4865	.7368	.9731	.9986
		L_4	.0020	.0256	.2499	.4956	.7456	.9767	.9990
	200	L_1	.0009	.0214	.2484	.4958	.7423	.9712	.9979
		L_2	.0013	.0246	.2526	.5008	.7483	.9748	.9984
		L_3	.0012	.0233	.2451	.4916	.7410	.9736	.9982
		L_4	.0016	.0251	.2513	.4958	.7453	.9760	.9988
2	30	L_1	.0005	.0167	.2308	.4789	.7256	.9626	.9960
		L_2	.0014	.0262	.2532	.5007	.7451	.9763	.9990
		L_3	.0007	.0194	.2201	.4630	.7179	.9718	.9986
		L_4	.0016	.0255	.2373	.4789	.7329	.9765	.9992
	50	L_1	.0006	.0180	.2344	.4834	.7297	.9656	.9970
		L_2	.0012	.0252	.2528	.4994	.7461	.9742	.9987
		L_3	.0010	.0201	.2273	.4689	.7236	.9704	.9984
		L_4	.0016	.0250	.2395	.4834	.7351	.9757	.9991
	100	L_1	.0008	.0192	.2398	.4899	.7374	.9694	.9977
		L_2	.0012	.0245	.2512	.4980	.7481	.9749	.9988
		L_3	.0009	.0211	.2331	.4760	.7307	.9718	.9986
		L_4	.0016	.0247	.2437	.4899	.7415	.9762	.9990
	200	L_1	.0008	.0206	.2445	.4947	.7415	.9713	.9979
		L_2	.0014	.0251	.2501	.4992	.7472	.9743	.9987
		L_3	.0012	.0225	. 371	.4838	.7351	.9722	.9985
		L_4	.0014	.0244	.2470	.4947	.7444	.9759	.9988
		$\Phi(x)$	.0013	.0250	.2514	.5000	.7486	.9750	.9987

Table 3: Simulated distribution functions of L_1−L_4 compared to the standard normal distribution function based on 100,000 simulation runs from the Weibull distribution.

one-sample log-rank test $L_4 = (O - E)/\sqrt{O + E/2} = -1.726 < -1.645$, or a p-value of 0.042; thus, we can claim that the mortality from other causes among patients with melanoma is significantly lower than that of the Danish general population.

Conclusions

A simple one-sample log-rank test is proposed, and its sample size formula is derived. Simulation results showed that the new test L_4 preserves the type I error well and is comparable to the two corrections based on Edgeworth expansion [10]. The proposed new test L_4 had power higher than that of the original test L_1 and the two corrections L_2 and L_3. The sample size formula derived from the new test statistic L_4 provides adequate power for the study design. To use the one-sample log-rank test to design a study and make inferences, the underlying distribution or hazard function of the standard population has to be correctly specified, because both study design and inference depend on the validity of this assumption. In an epidemiologic study, the standard population is often well defined. Therefore, one can use the method proposed by Finkelstein et al. [1] to calculate the expected number of events and estimate the survival distribution of the standard population. In a phase II clinical trial, the survival function of the historical control can be estimated from meta-analysis or other sources [10]. Nevertheless, a simple one-sample log-rank test is proposed, and its sample size formula is derived to provide a study design that preserves the type I error and ensures sufficient power to detect the difference of survival distributions between a sample and a standard population.

Acknowledgment

This work was supported in part by the National Cancer Institute (NCI) support grant P30CA021765-35.

References

1. Finkelstein DM, Muzikansky A, Schoenfeld DA (2003) Comparing survival of a sample to that of a standard population. J Natl Cancer Inst 95: 1434-1439.
2. Breslow NE (1975) Analysis of survival data under the proportional hazards model. International statistical Review 43: 44-58.
3. Hyde J (1977) Testing survival under right censoring and left truncation. Biometrika 64: 225-230.
4. Anderson PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes, Springer, New York.
5. Gill MH, Ware JH (1979) Comparing observed life table data with a known survival curve in the presence of random censorship. Biometrics 35: 385-391.
6. Berry G (1983) The analysis of mortality by the subject-years method. See comment in PubMed Commons below Biometrics 39: 173-184.
7. Woolson RF (1981) Rank-tests and a one-sample log-rank test for comparing observed survival-data to a standard population. Biometrics 37: 687-696.
8. Kwak M, Jung SH (2014) Phase II clinical trials with time-to-event endpoints: optimal two-stage designs with one-sample log-rank test. Stat Med 33: 2004-2016.
9. Jung SH (2013) Randomized phase II cancer clinical trial. CRC Press: Chapman & Hall.
10. Sun X, Peng P, Tu D (2011) Phase II cancer clinical trials with a one-sample log-rank test and its corrections based on the Edgeworth expansion. See comment in PubMed Commons below Contemp Clin Trials 32: 108-113.
11. Wu J (2014) Sample size calculation for the one-sample log-rank test. Technique report, Department of Biostatistics, St. Jude Children's Research Hospital.
12. Fleming TR, Harrington DP (1991) Counting processes and survival analysis. John Wiley and Sons, New York.

Citation: Wu J (2014) A New One-Sample Log-Rank Test. J Biomet Biostat 5: 210. doi:10.4172/2155-6180.1000210