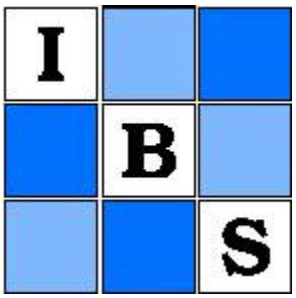

BIOMETRICS 38, 29–41
March 1982

A Confidence Interval for the Median Survival Time

Ron Brookmeyer[1] and John Crowley[2]
Departments of Statistics and Human Oncology, University of Wisconsin–Madison,
1120 West Johnson Street, Madison, Wisconsin 53715, U.S.A.

SUMMARY

A nonparametric asymptotic confidence interval for the median survival time is developed for the case where data are subject to arbitrary right censoring. This is accomplished by inverting a generalization of the sign test for censored data. A simulation study shows that this nonparametric confidence interval performs well for a variety of underlying survival functions. The procedure is applied to data from a clinical trial that compared four dosage regimens of 5-fluorouracil.

1. Introduction

It has become a common practice in the medical literature to give a point estimate for the median survival time. Peto *et al.* (1977) observed that this can be very misleading: if the survival curve is relatively flat in the neighborhood of 50% survival, there can be a great deal of variability in the estimated median. It would be more appropriate to cite a confidence interval for the median.

Confidence intervals can be derived with the assumption of a parametric form for the underlying survival distribution. The results of Bartholomew (1957), for example, lead to an interval estimate for the median when the survival curve is exponential. However, a distribution-free confidence interval is often desirable.

In the one-sample problem without censoring, a nonparametric confidence interval for the median is developed by inversion of the sign test. Similarly, for the problem with censoring, a nonparametric confidence interval for the median can be obtained by generalization of the sign test to enable censored data to be handled. This generalized sign test is given in §3, and the resulting confidence interval is presented in §4. Some Monte Carlo results of the proposed confidence interval are given in §5, followed by an example using data from a clinical trial of treatments for colorectal cancer.

2. The Survival Function and its Quantiles

Let $X_1^0, X_2^0, \ldots, X_n^0$ be the true survival times of a sample of size n. They are assumed to be independent, identically distributed random variables with survival function $S^0(t) = \text{pr}(X_j^0 > t)$ and cumulative distribution function $F^0(t) = 1 - S^0(t)$. When observations are subject to arbitrary right censoring, the period of followup for the ith individual is restricted to an amount T_i. Then the observed survival time for the ith individual is $X_i = \min(X_i^0, T_i)$. One also observes δ_i, which will indicate if X_i is censored or not. Thus, if $X_i < X_i^0$, the observation is said to be censored and we set $\delta_i = 0$. On the other hand, if $X_i = X_i^0$, the death is observed and we set $\delta_i = 1$.

---

[1] Present address: Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, 615 N. Wolfe Street, Baltimore, Maryland 21205, U.S.A.
[2] Present address: Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104, U.S.A.
Key words: Censoring; Confidence intervals; Medians; Sign test; Survival analysis.

Kaplan and Meier (1958) introduced the product-limit estimator $\hat{S}^0(t)$ of a survival function. It is defined as

$$1-\hat{F}^0(t)=\hat{S}^0(t)=\prod_{\substack{\text{distinct}\\ X_i\leqslant t}}\left(\frac{n_i-d_i}{n_i}\right).$$

The product is taken over all distinct observed survival times $\{X_i\}$; n_i is the number of observations greater than or equal to X_i (that is, n_i is the number of individuals at risk at time X_i), and d_i is the number of observed deaths that occurred at time X_i. In case of a tie between a censored observation and an observed death, the convention is adopted that the observed death is ranked before the censored observation. If the largest observation, say $X_{(n)}$, is censored, then $\hat{S}^0(t)$ is defined to be 0 for $t\geqslant X_{(n)}$.

The estimator $\hat{S}^0(t)$ is a right-continuous step function which jumps at the observed death points. It has some desirable properties. Kaplan and Meier showed that it is the distribution function which maximizes the likelihood of the observations within the class of all cumulative distribution functions which have probability mass at each observed death time. Efron (1967) showed that the Kaplan–Meier estimate $\hat{S}^0(t)$ of a survival distribution $S^0(t)$ satisfies the self-consistency relation

$$n\hat{S}^0(t)=N_x(t)+\sum_{X_i\leqslant t}(1-\delta_i)\frac{\hat{S}^0(t)}{\hat{S}^0(X_i)}, \tag{1}$$

where $N_x(t)$ is the number of X_i greater than t; it is assumed that there are no ties among the observed survival times; note that Efron originally worked with left-continuous survival functions.

The pth quantile of the distribution F^0 is defined to be $(F^0)^{-1}(p)=\inf\{t:F^0(t)\geqslant p\}$. J. Sander, in Technical Report No. 5, Biostatistics, Stanford University (1975), studied the large sample properties of $(\hat{F}^0)^{-1}(p)$ under the model of random censorship, and showed under suitable regularity conditions that $n^{\frac{1}{2}}\{(\hat{F}^0)^{-1}(p)-(F^0)^{-1}(p)\}$ converges weakly to a zero-mean Gaussian process. As in the uncensored case, the expression for the variance of the pth sample quantile depends explicitly on $f^0\{(F^0)^{-1}(p)\}$, where f^0 is the continuous density corresponding to F^0. The problem with the use of this variance to derive a confidence interval is the difficulty in estimating the density at the median.

3. A Sign Test for Censored Data

As discussed in §2, estimates of a survival distribution, and thus its quantiles, can be readily obtained. A point estimate should be accompanied by a confidence interval. This will be accomplished by inversion of a generalized sign test for censored data.

First, we extend the sign test to the censored-data problem. Given a sample of size n, we are interested in the hypothesis-testing problem with H_0: median of survival function $S^0=M$; H_1: median of survival function $S^0\neq M$. Consider the scoring function $\hat{Q}(X_i,\delta_i)=\text{pr}(X_j^0>M\mid X_i,\delta_i,\hat{S}^0)$. This conditional probability is to be interpreted, in the spirit of Efron (1967), as if X_i^0 actually had the distribution $\hat{S}^0$. Then,

$$\hat{Q}(X_i,\delta_i)=\begin{cases}1, & X_i>M,\\ \hat{S}^0(M)/\hat{S}^0(X_i), & X_i\leqslant M,\quad \delta_i=0,\\ 0, & X_i\leqslant M,\quad \delta_i=1.\end{cases}$$

Define the test statistic $U = (1/n)\sum_1^n \hat{Q}(X_i, \delta_i)$. Then we have the following result:

Theorem. The test statistic U (used to test H_0: median $= M$) is equal to $\hat{S}^0(M)$, the Kaplan–Meier estimate evaluated at M.

The proof of the theorem, which uses the property of self-consistency, is contained in Appendix 1.

The asymptotic distribution of the Kaplan–Meier estimate is well known. Breslow and Crowley (1974) showed that

$$n^{\frac{1}{2}} \frac{\hat{S}^0(M) - S^0(M)}{\sigma} \to \mathrm{N}(0, 1),$$

where $\sigma^2 = \{S^0(M)\}^2 \int_0^M (1-F)^{-2}\, d\tilde{F}$. Here F is the cumulative distribution of observed lifetimes, i.e. $1 - F(t) = \mathrm{pr}(X_i > t)$, and $\tilde{F}(t)$ is the distribution of true observed lifetimes, i.e. $\tilde{F}(t) = \mathrm{pr}(X_i \leqslant t,\ \delta_i = 1)$. A consistent estimate of σ^2 (Greenwood, 1926) is given by

$$\hat{\sigma}^2 = \{\hat{S}^0(M)\}^2 \sum_{\substack{\text{distinct} \\ X_i \leqslant M}} \frac{nd_i}{N_x(X_i)\{N_x(X_i) + d_i\}}, \tag{2}$$

where d_i is the number of observed deaths at X_i and $N_x(X_i)$ is the number of patients with observed survival times larger than X_i. This estimate can be used when there are ties in the data, as well as in the no-ties situation.

Then

$$\{\hat{S}^0(M) - S^0(M)\}^2 / \{\hat{S}^0(M)\}^2 \sum_{\substack{\text{distinct} \\ X_i \leqslant M}} \frac{d_i}{N_x(X_i)\{N_x(X_i) + d_i\}}$$

is approximately $\chi^2(1)$.

Under the null hypothesis $S^0(M) = \frac{1}{2}$, and an approximate α-level test is not to reject H_0 when

$$\{\hat{S}^0(M) - \tfrac{1}{2}\}^2 \leqslant c_\alpha \{\hat{S}^0(M)\}^2 \sum_{\substack{\text{distinct} \\ X_i \leqslant M}} \frac{d_i}{N_x(X_i)\{N_x(X_i) + d_i\}},$$

where c_α is such that $\mathrm{pr}\{\chi^2(1) > c_\alpha\} = \alpha$.

4. A Confidence Interval for the Median

An asymptotic $1 - \alpha$ confidence region R_α for the median is immediately obtained as the set of all parameter values not rejected by the sign test at level α. That is,

$$R_\alpha = \left\{ m \mid \{\hat{S}^0(m) - \tfrac{1}{2}\}^2 \leqslant c_\alpha \{\hat{S}^0(m)\}^2 \sum_{\substack{\text{distinct} \\ X_i \leqslant M}} \frac{d_i}{N_x(X_i)\{N_x(X_i) + d_i\}} \right\}.$$

To find the region R_α, it is necessary only to check if observed death times are in the region. This is because the Kaplan–Meier estimate and its estimated variance jump only at observed death times. Thus, if t_1 and t_2 are two consecutive observed death times with $t_1 < t_2$ and $t_1 \in R_\alpha$, then the interval $[t_1, t_2)$ is contained in R_α. One would expect the confidence region R_α to be an interval which includes the estimated median. This is generally true, as the next results show.

We shall assume that the n observed survival times $X_1, \ldots, X_n$ are distinct, i.e. that no ties occur. Let t_1 and t_2 be consecutive observed death times with $t_1 < t_2$ and associated Kaplan–Meier estimates $\hat{S}_1^0$ and $\hat{S}_2^0$, respectively. The following results are proved in Appendix 2:

Property A. If $\hat{S}_2^0 > .5$, then $t_1 \in R_\alpha$ implies $t_2 \in R_\alpha$.

Property B. If $\hat{S}_1^0 < .5$ and $t_1 \notin R_\alpha$, then $t_2 \notin R_\alpha$ if $c_\alpha < 1$.

In practice, however, we are usually interested in values of $c_\alpha > 3$. Unfortunately, Property B does not hold for arbitrary c_α, as seen from the following example. Let n_i be the number of patients at risk at observed death times $t(n_i)$. Consider the following sequence $\{n_i\}$:

$$\{500, 499, 498, \ldots, 201, 200, 6, 5, 4, \ldots, 1\}.$$

A large number of patients were censored at death time $t(200)$. A calculation with $c_\alpha = 8$ shows that $\hat{S}_{200}^0 = .398$, $t(200) \notin R_\alpha$ but $t(6) \in R_\alpha$. However, such examples are rare and are characterized by an extreme censoring pattern.

Simulation results suggest that the confidence region is almost always an interval (see §5). Thus, it is reasonable to consider only the interval part of the confidence region R_α. Let $\{t_i\}$ be the ordered observed distinct death times. Define the confidence interval $I_\alpha = [t_i, t_j)$, where t_i is the smallest observed death time in R_α with $\hat{S}^0(t_i) > .5$, and t_j is the smallest observed death time not in R_α with $\hat{S}^0(t_j) < .5$.

Occasionally it happens that an upper confidence limit cannot be obtained. If the last observed death time is in I_α, then I_α becomes a one-sided confidence interval of the form $[t_i, \infty)$. Furthermore, if the Kaplan–Meier survival curve does not reach the median because of extensive censoring, only a lower confidence limit (or perhaps an empty interval) can be obtained.

An alternative confidence region R_α^* can be derived by replacing $\hat{S}^0(M)$ in the estimate $\hat{\sigma}^2$ by $\frac{1}{2}$, the value of $S^0(M)$ under H_0, but this often leads to a one-sided confidence interval of the form $[t_i, \infty)$. This can be verified analytically. Simulation results also supported the conclusion that the region R_α^* was far too frequently one-sided. In addition, we are not guaranteed an interval, as the example given above still applies.

5. Simulation Results

In order to investigate the coverage probability and length of the nonparametric confidence interval I_α, a computer simulation was performed. The nonparametric interval I_α was compared to two parametric intervals which assume an exponential survival distribution.

The results of Bartholomew (1957) can be used to derive the maximum likelihood estimate of the median and its asymptotic variance. For an exponential(λ) survival distribution given by $S^0(t) = \exp(-\lambda t)$, $\lambda > 0$, the maximum likelihood estimate of the median M is

$$\hat{M} = \frac{\ln 2 \sum (\text{observed survival times})}{d},$$

where d is the number of observed deaths. The estimated variance of $\hat{M}$ is $\hat{M}/\sum P_i$, where

$$P_i = 1 - \exp\left(\frac{-T_i \ln 2}{\hat{M}}\right),$$

T_i being the censoring time of the ith patient. Here, we have used the expected Fisher information to obtain the estimate of the variance. Thus an approximate $(1-\alpha)$-level confidence interval, called here the 'Bartholomew confidence interval', is given by

$$\left\{\hat{M}-\frac{Z_{\frac{1}{2}\alpha}\hat{M}}{(\sum P_i)^{\frac{1}{2}}}, \hat{M}+\frac{Z_{\frac{1}{2}\alpha}\hat{M}}{(\sum P_i)^{\frac{1}{2}}}\right\}, \tag{3}$$

where $Z_{\frac{1}{2}\alpha}$ is the $100(1-\frac{1}{2}\alpha)$ percentage point of an N(0, 1) distribution. In order to calculate the Bartholomew confidence interval given by (3), the censoring times $\{T_i\}$ of all patients must be known. This is often not the case; in particular, the censoring times are usually not known for those patients who actually died.

The 'variance-stabilized confidence interval' is derived for an exponential survival distribution by transformation of the confidence interval for $\log\lambda$. The asymptotic variance of the maximum likelihood estimate $\hat{\lambda}$ is estimated by $\hat{\lambda}^2/d$, so $1/d$ estimates the variance of $\log\hat{\lambda}$. Here, we have used the observed Fisher information to obtain the estimate of variance. The confidence limits for $\log\lambda$ are transformed to provide a confidence interval for the median. The variance-stabilized $1-\alpha$ confidence interval for M is then

$$\left\{\frac{\ln 2}{\exp(\ln\hat{\lambda}+d^{-\frac{1}{2}}Z_{\frac{1}{2}\alpha})}, \frac{\ln 2}{\exp(\ln\hat{\lambda}-d^{-\frac{1}{2}}Z_{\frac{1}{2}\alpha})}\right\},$$

where $\hat{\lambda} = d/\sum$ (observed survival times).

Three experiments were performed, each with a different underlying survival distribution (exponential, Weibull and Rayleigh). Figure 1 gives a plot of the three hazard functions corresponding to the three distributions. For each experiment, 400 simulations were completed. For every simulation, survival and censoring times were generated for 50 patients. The censoring distribution was assumed to be uniform on $[0, T]$. For each simulation a $1-\alpha$ confidence interval for the median survival time was computed ($\alpha = .01, .05, .10, .20$ and $.25$). Then the observed coverage probability was recorded, i.e.

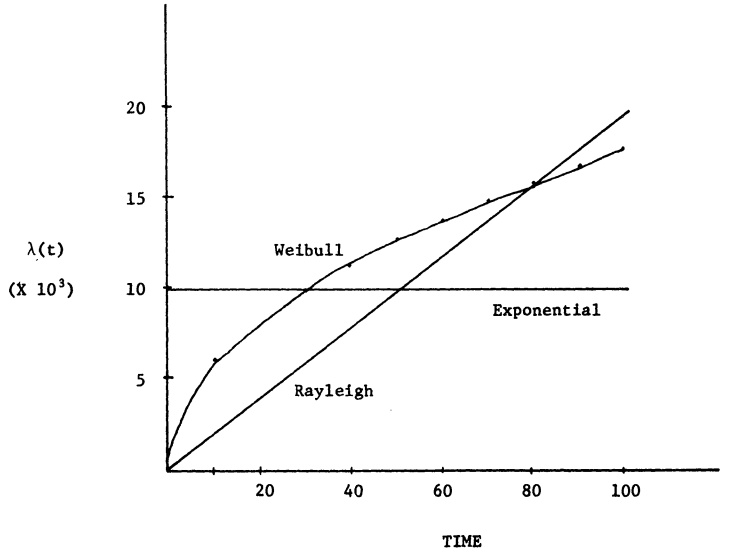

Figure 1. Hazards for three survival distributions. Exponential: $\lambda(t) = .01$. Weibull: $\lambda(t) = .0018t^{\frac{1}{2}}$. Rayleigh: $\lambda(t) = .0002t$.

the fraction of the 400 computed confidence intervals which contained the true median. If the interval estimates were performing adequately, the observed coverage probability would be close to $1-\alpha$. Several values of T were used ($T=650, 500, 250$ and 150), and for each the percentage of patients with censored survival times was recorded.

In order to compare interval lengths for the three types of confidence intervals (nonparametric, Bartholomew and variance-stabilized), average interval lengths were computed. If the censoring was extensive (roughly greater than 40%), the nonparametric method occasionally yielded only one-sided intervals. In these instances, average lengths could not be compared fairly. In addition, when average interval lengths were computed, so were average values of the Lehmann loss function. For an interval for M of the form $[t_i, t_j)$, this is defined (Lehmann, 1959, p. 82) to be

$$L(M, t_i, t_j)=\begin{cases} t_j - t_i & \text{if} \quad t_i \leq M \leq t_j, \\ t_j - M & \text{if} \quad M < t_i, \\ M - t_i & \text{if} \quad t_j > M. \end{cases}$$

A check was performed to see if the confidence regions were actually intervals. This was the case in all of the simulated samples.

Survival times for Experiment 1 were generated from an exponential (.01) distribution. The observed coverage probabilities are given in Table 1. For the three types of confidence intervals considered, the observed coverage probabilities were approximately the same. However, the average lengths of the confidence intervals were somewhat smaller for the parametric intervals than for the nonparametric intervals (see Table 2). Average values of the Lehmann loss function were similar to average interval lengths.

Survival times for Experiment 2 were generated from a Weibull (.0012, 1.5) distribution [$S^0(t)=\exp(-\lambda t^\gamma)$, with $\lambda=.0012$, $\gamma=1.5$]. Coverage probabilities are given in Table 3. The observed coverage probabilities for the nonparametric interval were consistently good. However, when there was little censoring the observed coverage probabilities for the parametric intervals were poor. The average confidence interval lengths were approximately the same for the three types of intervals (Table 4), as were average values of the Lehmann loss function.

Table 1

$1-\alpha$ *confidence intervals: observed coverage probabilities. Exponential* (.01) *survival distribution*

Confidence interval	% censored	α				
		.01	.05	.10	.20	.25
Nonparametric	15	.015	.065	.100	.230	.270
	20	.012	.060	.120	.220	.255
	37	.010	.057	.120	.227	.272
	52	.012	.055	.107	.220	.270
Bartholomew	15	.022	.055	.102	.202	.237
	20	.022	.065	.102	.235	.275
	37	.025	.063	.115	.220	.260
	52	.022	.063	.095	.217	.297
Variance-stabilized	15	.010	.060	.112	.200	.262
	20	.010	.060	.122	.212	.267
	37	.012	.067	.125	.217	.265
	52	.007	.060	.107	.235	.313

Table 2
Average confidence interval length. Exponential (.01) *survival distribution*

	15% censored		20% censored	
α:	.01	.10	.01	.10
Nonparametric	74.4	48.0	75.9	48.6
Bartholomew	55.5	35.5	57.1	36.5
Variance-stabilized	57.1	35.9	58.8	37.0

Table 3
$1-\alpha$ *confidence intervals: observed coverage probabilities. Weibull* (0.0012, 1.5) *survival distribution*

Confidence interval	% censored	α				
		.01	.05	.10	.20	.25
Nonparametric	12	.012	.057	.110	.227	.262
	16	.010	.055	.102	.215	.252
	32	.012	.057	.117	.235	.287
	51	.005	.063	.102	.217	.272
Bartholomew	12	.102	.245	.365	.510	.587
	16	.090	.200	.297	.465	.522
	32	.032	.095	.160	.245	.300
	51	.005	.030	.060	.135	.195
Variance-stabilized	12	.040	.155	.297	.457	.520
	16	.032	.125	.227	.405	.472
	32	.012	.057	.055	.165	.272
	51	.005	.027	.055	.165	.217

Table 4
Average confidence interval length. Weibull (.0012, 1.5) *survival distribution*

	12% censored		16% censored		32% censored	
α:	.01	.10	.01	.10	.01	.10
Nonparametric	48.4	32.0	49.1	32.5	53.7	34.8
Bartholomew	45.3	28.9	47.0	30.0	57.4	36.7
Variance-stabilized	46.4	29.2	48.1	30.3	58.6	36.8

Table 5
$1-\alpha$ confidence intervals: observed coverage probabilities. Rayleigh (.0001) survival distribution

Confidence interval	% censored	α				
		.01	.05	.10	.20	.25
Nonparametric	14	.010	.060	.105	.220	.265
	18	.007	.057	.120	.215	.280
	36	.007	.060	.132	.232	.280
	57	.015	.055	.117	.205	.240
Bartholomew	14	.167	.422	.570	.775	.840
	18	.117	.335	.477	.670	.737
	36	.020	.080	.155	.247	.290
	57	.002	.007	.020	.085	.130
Variance-stabilized	14	.057	.257	.477	.712	.785
	18	.032	.192	.372	.600	.688
	36	.005	.035	.097	.217	.265
	57	.002	.020	.057	.167	.212

Survival times for Experiment 3 were generated from a Rayleigh (.0001) distribution [$S^0(t) = \exp(-\lambda t^2)$, with $\lambda = .0001$]. Results are given in Tables 5 and 6. The nonparametric confidence interval performed well for all degrees of censoring. However, observed coverage probabilities for the parametric confidence intervals were very poor when the censoring was less than 20%. As the censoring increased, the observed coverage probabilities for the parametric intervals improved, but the interval lengths increased dramatically. Average parametric interval lengths were larger than the nonparametric ones. Similar results were obtained for the Lehmann loss function.

The simulation results indicate that the parametric methods perform well when the data arise from an exponential distribution. When the data are not generated from an exponential model, the observed coverage probabilities for the parametric intervals fluctuate wildly with the degree of censoring. The nonparametric method performed well for the three survival distributions investigated. In addition, the nonparametric interval length and the Lehmann loss function were reasonable when compared against the parametric competitors. Although the nonparametric method performed well in moderate sample sizes, the normal approximation to the distribution of $\hat{S}^0$ may be inadequate in small samples. Suitable transformations of $\hat{S}^0$ may improve the approximation; however they have not been considered here.

Emerson (1982) related the generalized sign test discussed here to a binomial distribution for small samples. Following a similar simulation design to compare the two studies,

Table 6
Average confidence interval length. Rayleigh (.0001) survival distribution

	14% censored		18% censored		36% censored	
α:	.01	.10	.01	.10	.01	.10
Nonparametric	43.8	29.0	44.4	29.2	49.1	32.1
Bartholomew	51.5	33.0	54.1	34.6	70.0	44.8
Variance-stabilized	52.7	33.2	55.1	34.7	70.9	44.4

he obtained results which appear quite good. Efron (1981) developed a confidence interval by the method of bootstrapping censored samples; that is, the distribution of the sample median is estimated by Monte Carlo simulation of samples generated from $\hat{S}^0$. These techniques are summarized and reviewed by Reid (1981).

6. An Example

The confidence interval procedure discussed earlier was applied to data from a Phase III colorectal cancer clinical trial (Ansfield *et al.*, 1977). [The data analyzed were an updated version reported in the Central Oncology Group Final Report (COG 7030), Spring 1977.] Four dosage regimens of 5-fluorouracil were compared. Table 7 summarizes the data for each of the four groups. The 95% confidence intervals for each of the four medians are also shown. Treatment 1, an intravenous loading-course schedule, had the largest median survival time of 61 weeks. Although the medians 'appear' to be different, all four confidence intervals overlap. In fact, the Gehan (Breslow) test shows no significant difference among the groups ($P = .22$). Since there were ties in the data, the Kaplan–Meier variance was estimated by (2). Table 8 details the calculations for Treatment 1. A numerical check shows that all the confidence regions are indeed proper intervals ($R_\alpha \equiv I_\alpha$). Figure 2 shows the Kaplan–Meier estimate for Treatment 1 and several $1-\alpha$ confidence intervals ($\alpha = .01, .05, .10, .20$ and $.25$).

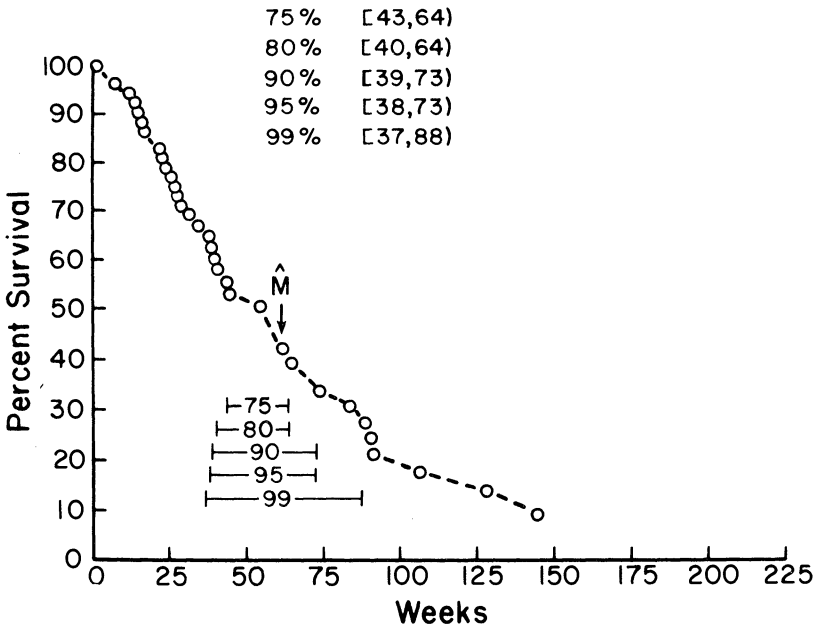

Figure 2. Confidence intervals for Treatment 1.

Table 7
Sample sizes, medians and 95% confidence intervals for four treatments

	Treatment			
	1	2	3	4
Censored	16	8	14	7
Observed deaths	37	48	44	45
Sample size	53	56	58	52
Median survival time (weeks)	61	41	47	29
95% nonparametric confidence interval	[38, 73)	[31, 51)	[28, 60)	[25, 46)
95% variance-stabilized confidence interval	(38, 73)	(28, 50)	(30, 55)	(27, 48)

Table 8
Calculations for 95% confidence interval for median of Treatment 1

Observed death time (weeks)	$\hat{S}^0$	$SE(\hat{S}^0)$	× if $(\hat{S}^0-\frac{1}{2})^2<3.84\{SE(\hat{S}^0)^2\}$
6	.962	.027	
11	.943	.032	
13	.925	.037	
14	.906	.040	
15	.887	.044	
16	.868	.047	
21	.829	.052	
22	.810	.055	
23	.791	.057	
25	.771	.058	
26	.752	.060	
27	.732	.062	
28	.713	.063	
31	.692	.064	
34	.671	.066	
37	.649	.067	
38	.627	.069	×
39	.604	.070	×
40	.582	.071	×
43	.559	.072	×
44	.534	.072	×
54	.508	.074	×
61	.423	.080	×
64	.395	.079	×
73	.338	.079	
83	.308	.077	
88	.277	.076	
90	.246	.073	
91	.215	.070	
106	.179	.067	
128	.144	.062	
144	.096	.057	

ACKNOWLEDGEMENTS

This research was supported in part by NIH grant CA-18332. We would like to thank Dr John Emerson for some helpful comments on an earlier manuscript, and Dr Fred Ansfield for permission to use his data.

RÉSUMÉ

Un intervalle de confiance asymptotique non paramétrique pour le temps de survie médian est développé dans le cas où les données sont arbitrairement tronquées à droite. Ceci est réalisé à l'aide d'une généralisation du test des signes pour les données tronquées. Une étude de simulation montre que cet intervalle non paramétrique convient bien pour une variété de fonctions de survie sous jacentes. La procédure est appliquée aux données d'une expérience clinique comparant des régimes suivant quatre dosages de 5-fluorouracil.

REFERENCES

Ansfield, F., Klotz, J. and the Central Oncology Group (1977). A Phase III study comparing the clinical utility of four dosage regimens of 5-fluorouracil. *Cancer* **39,** 34–40.

Bartholomew, D. J. (1957). A problem in life testing. *Journal of the American Statistical Association* **52,** 350–355.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Annals of Statistics* **2,** 437–453.

Efron, B. (1967). The two-sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol.* 4, L. LeCam and J. Neyman (eds), 831–854. Berkeley: University of California Press.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association* **76,** 312–319.

Emerson, J. (1982). Nonparametric confidence intervals for the median in the presence of right censoring. *Biometrics* **38,** 17–27.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52,** 203–223.

Greenwood, M. (1926). *The Natural Duration of Cancer.* Reports on Public Health and Medical Subjects, No. 33, 1–26. London: HMSO.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53,** 457–481.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* New York: Wiley.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *British Journal of Cancer* **35,** 1–39.

Reid, N. (1981). Estimating the median survival time. *Biometrika* **68,** 601–608.

Received July 1980; *revised January* 1981

APPENDIX 1

Proof of the Theorem

To prove that the test statistic U is the Kaplan–Meier estimate at M, write

$$\begin{aligned} U &= \frac{1}{n}\sum_{i=1}^{n} \hat{Q}(X_i, \delta_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{pr}(X_i^0 > M \mid X_i, \delta_i, \hat{S}^0) \\ &= \frac{1}{n}\left\{\sum_{\{\delta_i : \delta_i = 1\}} \mathrm{pr}(X_i^0 > M \mid X_i, \delta_i, \hat{S}^0) + \sum_{\{\delta_i : \delta_i = 0\}} \mathrm{pr}(X_i^0 > M \mid X_i, \delta_i, \hat{S}^0)\right\}. \end{aligned}$$

We have

$$\sum_{\{\delta_i : \delta_i = 1\}} \mathrm{pr}(X_i^0 > M \mid X_i, \delta_i, \hat{S}^0) = N_x^1(M),$$

where $N_x^1(M)$ = number of $X_i > M$ with $\delta_i = 1$. Also,

$$\sum_{\{\delta_i : \delta_i = 0\}} \text{pr}(X_i^0 > M \mid X_i, \delta_i, \hat{S}^0) = N_x^0(M) + \sum_{X_i \leq M} \frac{(1-\delta_i)\hat{S}^0(M)}{\hat{S}^0(X_i)},$$

where $N_x^0(M)$ = number of $X_i > M$ with $\delta_i = 0$. Thus,

$$\begin{aligned} U &= \frac{1}{n}\left\{N_x^1(M) + N_x^0(M) + \sum_{X_i \leq M} \frac{(1-\delta_i)\hat{S}^0(M)}{\hat{S}^0(X_i)}\right\} \\ &= \frac{1}{n}\left\{N_x(M) + \sum_{X_i \leq M} \frac{(1-\delta_i)\hat{S}^0(M)}{\hat{S}^0(X_i)}\right\}. \end{aligned}$$

Comparison of the term in brackets with the definition of selfconsistency given by (1) shows that $U = (1/n)\{n\hat{S}^0(M)\} = \hat{S}^0(M)$.

Appendix 2

Proofs of Property A and Property B

Proof of Property A. Let

$$V_j = \sum_{X_i \leq t_j} \frac{\delta_i}{N_x(X_i)\{N_x(X_i)+1\}},$$

and let N be the number of patients at risk at t_2. Then $t_1 \in R_\alpha$ implies $(\hat{S}_1^0 - \frac{1}{2})^2 < c_\alpha(\hat{S}_1^0)^2 V_1$ and $\{N\hat{S}_2^0/(N-1) - \frac{1}{2}\}^2 < \{N/(N-1)\}^2(\hat{S}_2^0)^2 c_\alpha V_1$, since $\hat{S}_1^0 = N\hat{S}_2^0/(N-1)$. Simplifying,

$$(\hat{S}_2^0)^2 - \frac{N-1}{N}\hat{S}_2^0 + \frac{1}{4}\left(\frac{N-1}{N}\right)^2 < (\hat{S}_2^0)^2 c_\alpha V_1.$$

Adding $(\hat{S}_2^0)^2 c_\alpha/\{(N-1)N\}$ to both sides of the inequality gives

$$(\hat{S}_2^0)^2 - \frac{N-1}{N}\hat{S}_2^0 + \frac{1}{4}\left(\frac{N-1}{N}\right)^2 + \frac{(\hat{S}_2^0)^2 c_\alpha}{(N-1)N} < (\hat{S}_2^0)^2 c_\alpha V_2, \tag{4}$$

since $V_2 = V_1 + 1/\{(N-1)N\}$. Thus it suffices to show that $(\hat{S}_2^0 - \frac{1}{2})^2$ is less than the left-hand side of (4). We have

$$\begin{aligned} \frac{1}{4} &< \frac{1}{4} + \frac{1}{4N^2} + \frac{c_\alpha}{4N^2} = \frac{1}{2N} + \frac{1}{4}\left(1 - \frac{2}{N} + \frac{1}{N^2}\right) + \frac{c_\alpha}{4N^2} \\ &< \frac{\hat{S}_2^0}{N^2} + \frac{1}{4}\left(\frac{N-1}{N}\right)^2 + \frac{c_\alpha(\hat{S}_2^0)^2}{N^2} < \frac{\hat{S}_2^0}{N} + \frac{1}{4}\left(\frac{N-1}{N}\right)^2 + \frac{c_\alpha(\hat{S}_2^0)^2}{N(N-1)}, \end{aligned}$$

since $\hat{S}_2^0 \geq \frac{1}{2}$.
This implies that

$$(\hat{S}_2^0)^2 - \hat{S}_2^0 + \frac{1}{4} < (\hat{S}_2^0)^2 - \hat{S}_2^0 + \frac{\hat{S}_2^0}{N} + \frac{1}{4}\left(\frac{N-1}{N}\right)^2 + \frac{c_\alpha(\hat{S}_2^0)^2}{N(N-1)}.$$

Comparison with (4) gives the desired result:

$$(\hat{S}_2^0 - \tfrac{1}{2})^2 < (\hat{S}_2^0)^2 c_\alpha V_2.$$

Proof of Property B. Since $t_1 \notin R_\alpha$, we have (4) with the inequality sign reversed:

$$(\hat{S}_2^0)^2 - \frac{N-1}{N}\hat{S}_2^0 + \frac{1}{4}\left(\frac{N-1}{N}\right)^2 + \frac{(\hat{S}_2^0)^2 c_\alpha}{(N-1)N} > (\hat{S}_2^0)^2 c_\alpha V_2.$$

It suffices to show that $(\hat{S}_2^0-\frac{1}{2})^2$ is greater than the left-hand side of the above inequality. Since $c_\alpha<1$,

$$\begin{aligned}\frac{1}{4}&>\frac{1}{4}-\frac{1}{4N^2}+\frac{c_\alpha}{4N^2}=\frac{N-1}{2N^2}+\frac{1}{4}\left(\frac{N-1}{N}\right)^2+\frac{c_\alpha}{4N^2}\\&>\frac{\hat{S}_2^0}{N}+\frac{1}{4}\left(\frac{N-1}{N}\right)^2+\frac{c_\alpha}{4N^2}\\&>\frac{\hat{S}_2^0}{N}+\frac{1}{4}\left(\frac{N-1}{N}\right)^2+\frac{c_\alpha(\hat{S}_2^0)^2}{N(N-1)},\end{aligned}$$

since $\hat{S}_1^0<\frac{1}{2}$ which implies $\hat{S}_2^0<\frac{1}{2}(N-1)/N$. This gives

$$(\hat{S}_2^0-\tfrac{1}{2})^2=(\hat{S}_2^0)^2-\hat{S}_2^0+\tfrac{1}{4}>(\hat{S}_2^0)^2c_\alpha V_2.$$