Published in final edited form as:

Stat Biopharm Res. 2011 ; 3(3): 488–496. doi:10.1198/sbr.2011.10020.

Confidence intervals for the difference of median failure times applied to censored tumor growth delay data

Jianrong Wu

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

Abstract

In tumor xenograft experiments, tumor response to the anti-cancer agents is often assessed by tumor growth delay, the difference in median tumor quadrupling times between treatment and control. Tumor quadrupling time is often subjected to right censoring because death of the experimental mice or limitations of the follow-up period. In the literature, tumor growth delay is simply reported without a standard error or confidence interval being given, which ignores the noise in the experimental data. Here, we present seven confidence intervals for the difference of medians in a general right-censored failure time framework. Simulation studies are conducted to compare the coverage probability of these intervals for small samples. The proposed intervals are applied to tumor growth delay data from an actual tumor xenograft experiment.

Keywords

Bootstrap; Confidence interval; Density estimation; Right censoring; Tumor growth delay

1. Introduction

In cancer drug development, demonstrating the activity of anti-cancer drugs in preclinical animal models is important. For mouse models, human cancer cells are engrafted into mice to produce xenografts. Tumor-bearing mice are then randomized into control and treatment groups, and the maximum tolerated dose of the drug is administered to the treatment group. The volume of each tumor is measured at the initiation of the study and periodically throughout the study. Mice are euthanized when the tumor volume exceeds a predetermined threshold, thus resulting in incomplete longitudinal tumor volume data. In the literature, the maximum treatment-to-control ratio (T/C) is used to quantify the antitumor activity of the drug, where the T and C represent the mean (or median) of tumor volumes of treatment and control mice, respectively (Houghton, et al., 1997). However the maximum T/C ratio is often not available due to fast tumor growth in the control mice. Therefore, we can not assess antitumor activity using such an endpoint. An alternative endpoint is the time from initial treatment to an event defined as a tumor quadrupling. The difference in median tumor quadrupling times between the treatment and control groups, called tumor growth delay (TGD), quantifies the treatment effect. Teicher (2006) pointed out that TGD is a critically important measure of antitumor effectiveness because it most closely mimics clinical endpoints that require observation of the mice through the time of disease progression. TGD data are often subjected to right censoring because of the death of the experimental mice or

limitations of the follow-up period. Stuschke et al. (1990) demonstrated that survival analysis should be used to analyze TGD data. In the literature, however, TGD is simply reported without a standard error being given (Corbett, et al., 2003). Clearly, this practice ignores the noise in experimental data. A formal statistical approach requires obtaining a confidence interval of the TGD.

Recently several confidence intervals for the difference in median failure times between two groups have been developed. The earliest of these, proposed by Wang & Hettmansperger (1990), requires either the two-sample shift model assumption or estimation of the survival density function. Even though the density estimation can be accomplished with a window estimate using a kernel function, it is often difficult and requires large sample sizes (Wang & Hettmansperger, 1990). To avoid density estimation, Su & Wei (1993) developed a minimum dispersion statistic to construct such a confidence interval, but their method does not have an explicit solution of the confidence limits and tends to be conservative for small samples, as shown by their simulation study.

Here, seven nonparametric intervals for the difference in medians are discussed. These intervals were developed based on a minimum dispersion statistic proposed by Su & Wei (1993), a kernel density estimation proposed by Kosorok (1999), a histogram-type density estimation proposed by Collett (2003), a resampling method proposed by Keaney & Wei (1994) and three bootstrap intervals. The small sample performances of these intervals have not been compared in the literature. Therefore, simulation results for the coverage probabilities of these intervals will add great practical value.

2. Median failure time

Suppose a random censorship model consists of n pairs $(t_1, \delta_1), \dots, (t_n, \delta_n)$, where $t_j = t_j^0 \wedge c_j$ is either a true failure time, t_j^0 , or a censoring time, c_j , and $\delta_j = I(t_j = t_j^0)$ is a censoring indicator. Further assume that the true failure time t_j^0 has a population distribution function $F(t) = P(t_j^0 \leq t)$, survival function $S(t) = 1 - F(t)$, density function $f(t) = dF(t)/dt$, and median $m = S^{-1}(0.5)$. The survival function S can be estimated by the Kaplan-Meier method (Cox and Oakes, 1984), which is given by

$$\hat{S}(t) = \prod_{t_{(j)} \leq t} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}},$$

with $\hat{S}(t) = 0$ for $t > t_{(n)}$, where $\delta_{(j)}$ is the censoring indicator corresponding to the j^{th} ordered observation $t_{(j)}$. The variance of $\hat{S}(t)$ can be estimated by Greenwood's formula

$$\widehat{\text{var}}\{\hat{S}(t)\} = \hat{S}^2(t) \left\{ \sum_{t_{(j)} \leq t} \frac{\delta_{(j)}}{(n-j)(n-j+1)} \right\}. \quad (1)$$

The median failure time can be estimated as

$$\hat{m} = \min\{t_{(j)}, \hat{S}(t_{(j)}) < 0.5\}. \quad (2)$$

When the estimated survival function $\hat{S}(t)$ is exactly equal to 0.5 on an interval $[t_{(j)}, t_{(j+1)})$, the median is taken to be $\hat{m} = (t_{(j)} + t_{(j+1)})/2$. It has been shown that the asymptotic variance of $\hat{m}$ can be estimated by (Reid, 1981)

$$\widehat{\text{var}}(\hat{m}) = \frac{1}{\hat{f}^2(\hat{m})} \widehat{\text{var}}\{\hat{S}(\hat{m})\}, \quad (3)$$

where the $\hat{f}$ is a density estimate of f and $\widehat{\text{var}}\{\hat{S}(\hat{m})\}$ is given by Greenwood's formula (1) at $t = \hat{m}$. To use this asymptotic variance formula (3), we have to estimate the density function f . A common type of density estimation is a window estimate using a kernel function. For example, Kosorok (1999) proposed an optimal window estimate based on the kernel function (Kosorok, 1999)

$$\hat{f}(\hat{m}) = \int n^{1/5} \hat{Q}^{-1} K\left(\frac{\hat{m} - x}{n^{-1/5} \hat{Q}}\right) d\hat{F}(x),$$

where $\hat{F} = 1 - \hat{S}$, $\hat{Q}$ is twice the estimated interquartile range of F , and the kernel $K(\cdot)$ is triangular function on $[-1, 1]$.

$$K(x) = \begin{cases} x+1 & -1 \leq x \leq 0 \\ 1-x & 0 < x \leq 1 \\ 0 & |x| > 1. \end{cases}$$

Collett (2003) proposed a simple histogram-type estimate of the density function f given by

$$\hat{f}(\hat{m}) = \frac{\hat{S}(\hat{u}) - \hat{S}(\hat{l})}{\hat{l} - \hat{u}},$$

where

$$\hat{u} = \max_{1 \leq j \leq k} \{t_{(j)} | \hat{S}(t_{(j)}) \geq 0.5 + \varepsilon\}$$

and

$$\hat{l} = \min_{1 \leq j \leq k} \{t_{(j)} | \hat{S}(t_{(j)}) \leq 0.5 - \varepsilon\}$$

for a small value of ε . To estimate the density well often requires a large sample size (Wang and Hettmansperger, 1990). Therefore, two computationally intensive methods have been proposed to estimate the variance of $\hat{m}$ directly without density estimation.

Keaney & Wei (1994) proposed a resampling method to estimate the variance of $\hat{m}$. Let Z be a normal random variable with mean 0 and variance $\widehat{\text{var}}\{\hat{S}(\hat{m})\}$. Generate a large number of random samples $\{z_j, j = 1, \dots, M\}$ from Z . Then for each sample z_j , we obtain a solution ρ_j by solving the equation

$$\hat{S}(\rho_j) = 0.5 + z_j.$$

Then a reasonable estimate for $\text{var}(\hat{m})$ is

$$\widehat{\text{var}}(\hat{m}) = \frac{1}{M} \sum_{j=1}^M (\rho_j - \hat{m})^2.$$

Another computationally intensive method, a bootstrap procedure (Efron, 1981), can also be used to estimate the variance of $\hat{m}$, which will be discussed in next section.

3. Confidence interval for the difference in medians

Now consider a two-sample problem. The observations consist of $n_i (i = 1, 2)$ pairs $(t_{i1}, \delta_{i1}), \dots, (t_{in_i}, \delta_{in_i})$, where $t_{ij} = t_{ij}^0 \wedge c_{ij}$ is either a true failure time, t_{ij}^0 or a censoring time, c_{ij} , and $\delta_{ij} = I(t_{ij} = t_{ij}^0)$ is a censoring indicator. We assume that $t_{i1}^0, \dots, t_{in_i}^0$ are independent and identically distributed with survival function S_i , density f_i and median m_i , $i = 1, 2$. Let $\hat{S}_i(t)$ be the Kaplan-Meier survival function estimate and $\hat{m}_i$ be the median estimate defined by (2). We are interested in difference in medians $\tau = m_2 - m_1$. An estimate of τ is $\hat{\tau} = \hat{m}_2 - \hat{m}_1$, and a $100(1 - \alpha)\%$ large sample asymptotic confidence interval of τ can be obtained by

$$\hat{\tau} \pm z_{1-\alpha/2} \widehat{\text{se}}(\hat{\tau}),$$

where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ th percentile of the standard normal distribution and

$$\widehat{\text{se}}(\hat{\tau}) = \{\widehat{\text{var}}(\hat{m}_1) + \widehat{\text{var}}(\hat{m}_2)\}^{1/2}.$$

Here the variances $\widehat{\text{var}}(\hat{m}_1)$ and $\widehat{\text{var}}(\hat{m}_2)$ can be estimated by resampling or the bootstrap method or by a kernel density or histogram-type density estimation discussed in the previous section.

To avoid density estimation, Su & Wei (1993) consider following quantity

$$W(\tau, m_1) = \frac{(\hat{S}_1(m_1) - 0.5)^2}{\hat{\sigma}_1^2} + \frac{(\hat{S}_2(\tau + m_1) - 0.5)^2}{\hat{\sigma}_2^2}$$

where $\hat{\sigma}_i^2$ is the Greenwood's estimate for the variance of $\hat{S}_i(t)$ evaluated at $\hat{m}_i$, $i = 1, 2$. Here, m_1 is a nuisance parameter. A natural way to eliminate this parameter is to minimize $W(\tau, m_1)$ with respect to m_1 for a given τ , resulting a minimum dispersion pivotal quantity

$$G(\tau) = \min_{m_1} W(\tau, m_1).$$

Su & Wei showed that $G(\tau)$ is asymptotically chi-square distributed with 1 degree of freedom. Therefore, a $100(1 - \alpha)\%$ confidence interval for τ can be constructed as

$$\{\tau; G(\tau) < \chi_{1, 1-\alpha}^2\},$$

where $\chi_{1, 1-\alpha}^2$ is the $100(1 - \alpha)$ percentage point of χ_1^2 . Su & Wei claimed that for a fixed δ , $W(\tau, m_1)$ is a convex function of m_1 . Frick (1997) pointed out that this statement is not correct. However, $W(\tau, m_1)$ takes at most $(n_1 + 1)(n_2 + 1)$ different values. Therefore, the interval limits can be solved numerically.

Finally, confidence intervals of τ can also be obtained by using the following bootstrap procedures for the right-censored data (Efron, 1981):

1. Independently draw a large number of bootstrap samples, $\{(t_{i1}^{*b}, \delta_{i1}^{*b}), \dots, (t_{in_i}^{*b}, \delta_{in_i}^{*b}), b=1, \dots, B\}$, $i = 1, 2$, with each bootstrap sample obtained by sampling with replacement from $\{(t_{i1}, \delta_{i1}), \dots, (t_{in_i}, \delta_{in_i})\}$, $i = 1, 2$.
2. Calculate median estimates of Kaplan-Meier curves for bootstrap samples obtained in step 1), say $\hat{m}_i^{*b}$, $i = 1, 2$, $b = 1, \dots, B$.
3. Calculate $\hat{\tau}^{*b} = \hat{m}_2^{*b} - \hat{m}_1^{*b}$, $b = 1, \dots, B$.
4. A large sample bootstrap variance interval of τ is given by

$$\hat{\tau} \pm z_{1-\alpha/2} \{\widehat{\text{var}}^*(\hat{\tau})\}^{1/2},$$

where

$$\widehat{\text{var}}^*(\hat{\tau}) = \frac{1}{B-1} \sum_{i=1}^B (\hat{\tau}^{*b} - \hat{\tau}^*)^2$$

$$\text{with } \hat{\tau}^* = \sum_{i=1}^B \hat{\tau}^{*b} / B.$$

5. Let $F_*(s) = \#\{\hat{\tau}^{*b} < s\}/B$ be the bootstrap distribution of $\{\hat{\tau}^{*b}, b = 1, \dots, B\}$, then a $100(1 - \alpha)\%$ bootstrap percentile interval is given by

$$[F_*^{-1}(\alpha/2), F_*^{-1}(1 - \alpha/2)]. \quad (4)$$

6. Let $\hat{z} = \Phi^{-1}\{F_*(\hat{\tau})\}$, where Φ is the standard normal distribution function, then a $100(1 - \alpha)\%$ bias-corrected bootstrap percentile interval is given by

$$[F_*^{-1}\{\Phi(2\hat{z} - z_{1-\alpha/2})\}, F_*^{-1}\{\Phi(2\hat{z} + z_{1-\alpha/2})\}]. \quad (5)$$

Remark

In a case with more heavily censored data, there is a significant number of bootstrap samples whose Kaplan-Meier curves do not reach 0.5 survival probability. For each of those samples, Keaney & Wei (1994) set the bootstrap median estimate to be the largest failure time. However, such a bootstrap procedure tends to be conservative. In our bootstrap procedure, those bootstrap samples whose Kaplan-Meier curves do not reach 0.5 survival probability are simply excluded.

So far, we have discussed seven intervals for the difference in medians of two groups: two density estimation intervals, a minimum dispersion interval, a resampling interval, and three bootstrap intervals. Small sample coverage probabilities of these intervals are studied through Monte Carlo simulations in the next section.

3. Simulation Studies

The small sample performance of the intervals discussed in the previous section has not been studied and compared in the literature. Simulations were carried out to study the 95% coverage probabilities of these intervals under small samples of $n=10, 20$, and 30 per group. The failure times were generated from various survival distributions S_1 and S_2 in the following 4 scenarios,

1. $S_1(t) = e^{-t}$ and $S_2(t) = e^{-t}$.
2. $S_1(t) = e^{-t}$ and $S_2(t) = e^{-(1.13t)^{1.5}}$.
3. $S_1(t) = e^{-t}$ and $S_2(t) = 1 - \Phi(\log(1.44t))$.
4. $S_1(t) = e^{-e^{-24}t^{8.62}}$ and $S_2(t) = e^{-e^{-21}t^{5.78}}$.

where scenarios 1–3 are taken from Su & Wei's work (1993) with equal medians of the two groups. We added one more scenario, scenario 4, which is similar to real tumor growth delay data, with a difference in the medians of two groups $\tau = 20.1$.

The censoring times are generated from the uniform distribution $U(0, c_i)$ with $c_i, i = 1, 2$, which are determined by some prespecified censoring proportions at 10%, 20% and 30% of two groups. For each above scenario, we used 5,000 independent Monte Carlo samples to ensure that our Monte Carlo error for estimating a 95% coverage probability was about

$0.003 = (0.05 \times 0.95/5000)^{1/2}$. For each Monte Carlo sample, an additional 2,000 resampling samples and 2,000 bootstrap samples were generated for the corresponding method. Monte Carlo samples whose Kaplan-Meier curves did not reach 0.5 survival probability were replaced by new Monte Carlo samples from simulation runs. Monte Carlo samples whose Kaplan-Meier curves did not reach 0.25 survival probability were replaced from the kernel density method. Resampling samples whose value of $\{0.5 + z_j\}$ less than 0 or greater than 1 were excluded. Bootstrap samples whose Kaplan-Meier curves did not reach 0.5 survival probability were also excluded. The empirical coverage probability of each interval is the frequency of the number of intervals including the true τ . The simulation results are summarized in Tables 1–3. The results show that, overall, the interval based on Collett's histogram-type density estimation is liberal in terms of short interval length and low coverage probability; in contrast, intervals based on the Keaney & Wei's resampling method and Su & Wei's minimum dispersion statistic are conservative. Therefore, they are not suitable for small sample data. Kosorok's kernel density interval is liberal when $n = 10$ but it has an excellent coverage probability when $n = 20, 30$; the bootstrap variance interval could be either slightly conservative or liberal; the bootstrap percentile interval performs well for scenario 4 but is slightly conservative for scenarios 1–3; the bias-corrected bootstrap percentile interval does not correct the coverage probability toward the nominal level when $n = 10, 20$ but do correct the coverage probability toward the nominal level when $n = 30$.

4. Analysis of tumor growth delay data

An actual tumor xenograft experiment conducted by the Pediatric Preclinical Testing Program is used to illustrate tumor growth delay data analysis using the proposed methods. In this tumor xenograft model, glioblastoma tumor line D456 was tested for sensitivity to a single cytotoxic agent, cisplatin (Houghton, et al., 2007). Twenty mice were equally randomized to control and treatment groups. Tumor volumes were measured on a weekly schedule (Table 4). Therefore, the exact time of tumor quadrupling was not observed. A naive approach using the first week past tumor quadrupling is biased and inefficient (Thompson, et al., 1999). An interpolation formula can be used to calculate the approximate quadrupling time of a tumor (Wu & Houghton, 2009),

$$t_e = t_1 + (t_2 - t_1) \log(4V_0/V_{t_1}) / \log(V_{t_2}/V_{t_1}), \quad (6)$$

where t_e is the interpolated quadrupling time, t_1 and t_2 are the lower and upper observation times bracketing the quadrupling tumor volume and V_0 is the initial tumor volume. For the D456-cisplatin model, the tumor quadrupling times were calculated using the interpolation formula (5) (Table 4). Tumor volumes quadrupled in all of the control mice before day 21; therefore, the mice were euthanized before the end of the study, and no tumor volumes were recorded on or after day 21. Seven mice in the treatment group had quadrupled tumor volumes before day 35. Two mice had no measurable tumor at the end of study, and one mouse died on day 21; therefore, the quadrupling times of these three mice were censored on days 42, 42, and 21, respectively. The median tumor quadrupling times were $\hat{m}_1 = (8.2 + 9.2)/2 = 8.7$ days and $\hat{m}_2 = 24.9$ days for the control and treatment groups, respectively. The estimated TGD was $\hat{m}_2 - \hat{m}_1 = 16.2$ days. Seven confidence intervals of TGD are given

in Table 6. The interval obtained by Collett's histogram-type density method ($\epsilon = 0.025$) was the narrowest among the seven, which is consistent with the simulation results, liberal. The interval obtained by Su & Wei's minimum dispersion method was the longest, which is also consistent with the simulation results, conservative. Keaney & Wei's resampling interval had a wide interval which is also conservative. Kosorok's kernel density interval was unexpectedly wider than the bootstrap intervals in the example. Three bootstrap intervals were close. The bootstrap percentile interval had good coverage probabilities for sample sizes as small as 10 per group and therefore is recommended for small sample censored tumor growth delay data analysis.

5. Conclusion and discussion

Seven confidence intervals for the difference in medians have been discussed in a general right-censored failure time framework. Monte Carlo simulations were conducted to compare the coverage probabilities of the seven intervals. The simulation results showed that the proposed bootstrap percentile interval has good coverage probabilities for a sample size as small as 10 per group. The interval based on kernel density estimation also has good coverage probabilities for a sample size as small as 20 per group. The intervals obtained using the resampling method and the minimum dispersion method are conservative, and the interval obtained using histogram-type density estimation is liberal. The findings from our simulation study are valuable in choosing the right method for comparing the median failure times of two groups from a small clinical trial or preclinical study.

Acknowledgments

This work was supported in part by National Cancer Institute (NCI) support grants CA21765 and N01-CM-42216 and the American Lebanese Syrian Associated Charities (ALSAC).

References

- Collett, D. Modeling Survival data in medical research. 2. Chapman & Hall; London: 2003.
- Corbett TH, White K, et al. Discovery and Preclinical Antitumor Efficacy Evaluations of LY32262 and LY33169. *Investigational New Drugs*. 2003; 21:33–45. [PubMed: 12795528]
- Cox, DR., Oakes, D. Analysis of survival data. Chapman & Hall; London: 1984.
- Efron B. Censored data and the bootstrap. *Journal of American Statistical Association*. 1981; 76:312–319.
- Frick H. Nonparametric comparisons of medians of survival functions. *The Statistical Software Newsletter*. 1997; 27:357–361.
- Houghton PJ, Morton CL, Gorlick R, et al. The Pediatric Preclinical Testing Program: Description of Models and Early Testing Results. *Pediatr Blood Cancer*. 2007; 49:928–940. [PubMed: 17066459]
- Keaney KM, Wei JL. Interim analyses based on median survival times. *Biometrika*. 1994; 81:279–286.
- Kosorok M. Two-sample quantile tests under general conditions. *Biometrika*. 1999; 86:909–921.
- Reid N. Estimating the median survival time. *Biometrika*. 1981; 68:601–608.
- Stuschke M, Budach V, Bamberg M, Budach W. Methods for analysis of censored tumor growth delay data. *Radiation Research*. 1990; 122:172–180. [PubMed: 2336463]
- Su JQ, Wei LJ. Nonparametric estimation for the difference or ratio of median failure times. *Biometrics*. 1993; 49:603–607. [PubMed: 8369391]
- Teicher BA. Tumor models for efficacy determination. *Molecular Cancer Therapeutics*. 2006; 5:2435–2443. [PubMed: 17041086]

- Thompson J, George EO, Poquette CA, et al. Synergy of topotecan in combination with vincristine for treatment of pediatric solid tumor xenografts. *Clinical Cancer Research*. 1999; 5:3617–3631. [PubMed: 10589779]
- Wang JL, Hettmansperger TP. Two-sample inference for median survival times based on one-sample procedures for censored survival data. *Journal of American Statistical Association*. 1990; 85:529–536.
- Wu J, Houghton PJ. Assessing Cytotoxic Treatment Effects in Preclinical Tumor Xenograft Models. *Journal of Biopharmaceutical Statistics*. 2009; 19:755–762. [PubMed: 20183441]

Table 1
Empirical coverage probabilities of 95% confidence intervals for the difference of median failure times ($n=10$)

Scenario	Censoring	Bias	Method																NS	NB												
			Kosorok				Collett				Keaney & Wei				Su & Wei						B _v				B _p				B _{bc}			
			CP	L	CP	L	CP	L	CP	L	CP	L	CP	L	CP	L	CP	L			CP	L	CP	L	CP	L	CP	L	CP	L	CP	L
1	(1.,1)	-.002	.951	1.7	.890	1.8	.985	2.1	.992	2.6	.968	2.1	.969	2.1	.952	2.1	1	7.7														
	(1.,2)	-.013	.939	1.7	.891	1.8	.985	2.1	.992	2.5	.968	2.1	.969	2.1	.952	2.1	14	20.6														
	(1.,3)	-.058	.939	1.7	.875	1.7	.983	2.0	.993	2.4	.956	2.0	.967	1.9	.939	1.9	116	42.2														
	(2.,2)	.000	.928	1.7	.887	1.7	.984	2.0	.991	2.4	.957	2.0	.964	2.0	.932	1.9	31	33.7														
	(3.,3)	.008	.923	1.6	.875	1.6	.978	1.9	.988	2.1	.938	1.8	.954	1.7	.923	1.7	221	74.8														
2	(1.,1)	-.029	.942	1.5	.875	1.8	.979	1.8	.993	2.1	.956	1.8	.963	1.7	.941	1.7	1	6.0														
	(1.,2)	-.034	.935	1.5	.869	1.8	.982	1.8	.993	2.1	.958	1.8	.967	1.8	.941	1.7	16	13.6														
	(1.,3)	-.044	.937	1.5	.877	1.8	.982	1.8	.993	2.1	.954	1.7	.966	1.7	.941	1.7	80	31.3														
	(2.,2)	-.016	.926	1.5	.872	1.8	.977	1.8	.991	2.0	.945	1.7	.963	1.7	.931	1.6	31	25.5														
	(3.,3)	.008	.924	1.4	.863	1.7	.978	1.7	.988	1.8	.943	1.5	.963	1.5	.931	1.4	205	66.5														
3	(1.,1)	.017	.945	1.7	.889	1.7	.985	2.1	.991	2.7	.969	2.2	.969	2.2	.951	2.1	1	8.6														
	(1.,2)	.002	.938	1.7	.886	1.7	.983	2.0	.991	2.5	.964	2.1	.964	2.1	.944	2.0	21	21.4														
	(1.,3)	-.042	.936	1.6	.883	1.6	.984	1.9	.991	2.3	.966	1.9	.966	1.9	.944	1.9	107	41.7														
	(2.,2)	.015	.931	1.6	.875	1.7	.983	2.0	.993	2.4	.966	2.0	.966	2.0	.942	1.9	38	33.7														
	(3.,3)	.010	.933	1.5	.870	1.5	.980	1.8	.993	2.0	.944	1.7	.964	1.7	.927	1.6	206	75.1														
4	(1.,1)	-.032	.938	12.7	.839	12.7	.991	17.6	.977	14.9	.941	12.4	.954	12.1	.940	12.0	0	0.6														
	(1.,2)	.050	.927	13.0	.829	13.2	.992	19.3	.974	15.6	.936	12.9	.947	12.5	.926	12.5	0	2.2														
	(1.,3)	.016	.917	13.3	.832	14.0	.992	21.2	.974	16.5	.936	13.5	.940	13.0	.921	12.9	21	6.2														
	(2.,2)	-.006	.928	13.1	.833	13.2	.993	19.7	.976	15.9	.940	13.1	.952	12.6	.932	12.6	0	3.6														
	(3.,3)	-.076	.914	13.6	.829	13.9	.994	21.7	.977	17.0	.936	13.7	.944	13.2	.925	13.2	32	10.5														

Bias: average of the estimates minus the true of median difference in 5,000 simulation runs

CP: empirical coverage probability

L: average interval lengths

B_v: bootstrap variance interval

B_p: bootstrap percentile interval

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

B_{pc}: bias-corrected bootstrap percentile interval
NS: number of Monte Carlo samples whose Kaplan-Meier curves did not reach 0.5 survival probability
NB: average number of bootstrap samples whose Kaplan-Meier curves did not reach 0.5 survival probability

Table 2
Empirical coverage probabilities of 95% confidence intervals for the difference of median failure times ($n=20$)

Scenario	Censoring	Bias	Method												NS	NB		
			Kosorok		Collett		Keaney & Wei		Su & Wei		B _v		B _p				B _{bc}	
			CP	L	CP	L	CP	L	CP	L	CP	L	CP	L			CP	L
1	(1,1)	.002	.956	1.3	.891	1.4	.974	1.5	.988	1.7	.962	1.4	.972	1.4	.944	1.4	0	0.1
	(1,2)	.017	.951	1.3	.881	1.4	.970	1.5	.988	1.7	.958	1.5	.967	1.5	.940	1.5	0	1.3
	(1,3)	.013	.947	1.3	.886	1.4	.972	1.5	.989	1.7	.957	1.5	.968	1.5	.941	1.5	9	7.6
	(2,2)	.003	.951	1.3	.889	1.4	.976	1.5	.990	1.7	.965	1.5	.970	1.5	.944	1.5	1	2.5
	(3,3)	.003	.947	1.3	.888	1.4	.977	1.5	.992	1.7	.963	1.5	.970	1.5	.945	1.5	16	15.1
2	(1,1)	−.006	.954	1.1	.882	1.2	.972	1.3	.984	1.4	.956	1.2	.966	1.2	.936	1.2	0	0.2
	(1,2)	.003	.951	1.1	.880	1.2	.972	1.3	.984	1.4	.957	1.2	.967	1.2	.940	1.2	0	3.2
	(1,3)	−.016	.957	1.1	.884	1.2	.979	1.3	.986	1.4	.964	1.2	.973	1.2	.950	1.2	3	3.2
	(2,2)	−.014	.948	1.1	.886	1.2	.972	1.3	.986	1.4	.957	1.3	.967	1.3	.938	1.3	1	4.6
	(3,3)	−.003	.942	1.1	.871	1.2	.968	1.3	.991	1.4	.955	1.3	.966	1.3	.939	1.3	8	20.9
3	(1,1)	−.006	.957	1.2	.884	1.3	.973	1.4	.985	1.6	.960	1.4	.966	1.4	.939	1.4	0	0.1
	(1,2)	.019	.958	1.3	.888	1.3	.975	1.4	.987	1.6	.968	1.5	.970	1.4	.943	1.4	0	0.8
	(1,3)	.019	.958	1.2	.894	1.3	.978	1.4	.989	1.6	.967	1.4	.969	1.4	.944	1.4	1	6.7
	(2,2)	.008	.958	1.3	.888	1.3	.976	1.5	.987	1.7	.967	1.5	.970	1.5	.947	1.5	0	2.0
	(3,3)	.003	.952	1.2	.882	1.3	.979	1.4	.988	1.7	.967	1.5	.970	1.5	.940	1.5	4	12.9
4	(1,1)	−.013	.959	9.4	.823	8.8	.967	10.3	.973	9.8	.941	8.9	.953	8.7	.933	8.7	0	0.0
	(1,2)	.074	.955	9.7	.830	9.3	.970	11.1	.974	10.3	.943	9.4	.955	9.1	.937	9.2	0	0.0
	(1,3)	.050	.945	10.1	.828	9.7	.974	11.9	.974	10.9	.943	10.0	.957	9.6	.938	9.6	0	0.2
	(2,2)	.056	.955	9.8	.840	9.4	.970	11.2	.976	10.3	.944	9.5	.955	9.2	.939	9.2	0	0.0
	(3,3)	.061	.940	10.2	.828	10.0	.972	12.2	.978	11.1	.937	10.2	.955	9.8	.924	9.8	0	0.4

Bias: average of the estimates minus the true of median difference in 5,000 simulation runs

CP: empirical coverage probability

L: average interval lengths

B_v: bootstrap variance interval

B_p: bootstrap percentile interval

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

B_{bc}: bias-corrected bootstrap percentile interval
NS: number of Monte Carlo samples whose Kaplan-Meier curves did not reach 0.5 survival probability
NB: average number of bootstrap samples whose Kaplan-Meier curves did not reach 0.5 survival probability

Table 3
Empirical coverage probabilities of 95% confidence intervals for the difference of median failure times ($\mu=30$)

Method																		
Scenario	Censoring	Bias	Kosorok		Collett		Keaney & Wei		Su & Wei		B _v		B _p		B _{hc}			
			CP	L	CP	L	CP	L	CP	L	CP	L	CP	L	CP	L	NS	NB
1	(1,1)	.000	.951	1.0	.897	1.1	.973	1.2	.986	1.3	.964	1.2	.972	1.1	.950	1.1	0	0
	(1,2)	.000	.953	1.0	.902	1.1	.971	1.2	.985	1.3	.965	1.2	.970	1.2	.951	1.2	0	0.1
	(1,3)	.000	.953	1.1	.904	1.1	.973	1.2	.986	1.4	.964	1.2	.970	1.2	.953	1.2	0	1.8
	(2,2)	.001	.956	1.1	.903	1.1	.976	1.2	.987	1.4	.968	1.2	.976	1.2	.955	1.2	0	0.2
	(3,3)	-.003	.949	1.1	.897	1.1	.975	1.2	.986	1.4	.966	1.3	.975	1.3	.952	1.3	0	0.2
2	(1,1)	-.008	.953	0.9	.892	0.9	.964	1.0	.983	1.1	.952	1.0	.963	1.0	.943	1.0	0	0.0
	(1,2)	-.008	.952	0.9	.890	0.9	.974	1.0	.982	1.1	.953	1.0	.963	1.0	.943	1.0	0	0.0
	(1,3)	-.011	.948	0.9	.889	0.9	.976	1.0	.982	1.1	.958	1.0	.969	1.0	.943	1.0	0	0.6
	(2,2)	-.011	.953	0.9	.897	0.9	.983	1.0	.986	1.1	.964	1.0	.970	1.0	.948	1.0	0	0.1
	(3,3)	-.014	.954	0.9	.888	0.9	.984	1.0	.984	1.2	.963	1.1	.970	1.2	.947	1.1	1	2.4
3	(1,1)	.005	.952	1.0	.899	1.0	.970	1.1	.981	1.2	.959	1.0	.967	1.1	.946	1.1	0	0
	(1,2)	.008	.954	1.0	.889	1.0	.967	1.1	.983	1.3	.963	1.1	.965	1.1	.944	1.2	0	0.1
	(1,3)	.005	.959	1.0	.896	1.0	.970	1.1	.985	1.3	.965	1.2	.970	1.1	.950	1.2	0	1.5
	(2,2)	.002	.956	1.0	.900	1.0	.971	1.1	.982	1.3	.966	1.2	.968	1.2	.945	1.2	0	0.3
	(3,3)	.004	.954	1.0	.897	1.1	.973	1.2	.986	1.3	.969	1.2	.976	1.2	.950	1.3	0	3.4
4	(1,1)	0.021	.958	7.6	.866	7.2	.963	8.1	.972	7.8	.943	7.3	.962	7.1	.944	7.1	0	0.0
	(1,2)	0.028	.956	8.0	.852	7.4	.959	8.5	.972	8.3	.940	7.7	.960	7.5	.942	7.5	0	0.0
	(1,3)	0.035	.952	8.4	.850	7.9	.960	9.1	.972	8.8	.941	8.2	.958	7.9	.942	7.9	0	0.0
	(2,2)	0.023	.956	8.0	.852	7.5	.961	8.6	.974	8.3	.939	7.7	.961	7.5	.943	7.5	0	0.0
	(3,3)	0.032	.951	8.4	.852	8.0	.962	9.3	.973	8.9	.941	8.3	.958	8.0	.943	8.0	0	0.0

Bias: average of the estimates minus the true of median difference in 5,000 simulation runs

CP: empirical coverage probability

L: average interval lengths

B_v: bootstrap variance interval

B_p: bootstrap percentile interval

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

B_{pc}: bias-corrected bootstrap percentile interval
NS: number of Monte Carlo samples whose Kaplan-Meier curves did not reach 0.5 survival probability
NB: average number of bootstrap samples whose Kaplan-Meier curves did not reach 0.5 survival probability

Table 4Tumor volumes (cm^3) measured in D456-cisplatin tumor xenograft model

Group	Days	Mouse									
		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Control	0	0.32	0.55	0.47	0.29	2.12	0.68	0.55	0.24	2.27	0.65
	7	0.55	2.03	1.55	0.39	6.51	2.57	2.00	0.64	8.03	2.58
	14	1.65	3.78	4.95	1.02	11.5	6.01	3.90	2.22	10.8	4.17
	21	.	.	2.37	.	.	.	.	.	.	.
Drug	0	0.19	0.45	0.30	0.44	1.32	0.15	0.15	0.18	0.27	0.49
	7	0.15	0.70	0.58	0.36	1.46	0.08	0.22	0.15	0.81	0.79
	14	0.24	0.87	1.21	0.71	1.84	0.06	0.22	0.40	1.03	0.97
	21	0.25	1.60	1.73	1.23	2.57	0.03	0.32	0.87	2.06	1.57
	28	0.49	1.69	.	2.32	•	0.04	0.82	.	.	2.88
	35	0.05	2.43	.	.	•	0.02	.	.	.	.
	42	0.00	.	.	.	•	0.00	.	.	.	.

.: indicates a missing value due to a mouse being euthanized after its tumor quadrupled

•: indicates that mouse died due to toxicity

Table 5
Tumor quadrupling times (days) calculated from interpolation formula for D456-cisplatin model

Group	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Control	12.4	7.8	8.2	15.1	10.3	7.5	8.0	9.2	9.8	7.2
Drug	42 *	29.1	14.0	24.9	21 *	42 *	25.8	19.1	14.6	23.5

* : indicates censoring

Table 6

95% confidence intervals of TGD for D456-cisplatin tumor xenograft model

Method	95% C.I.	Length
Kosorok's kernel	[7.7, 24.7]	17.0
Collett's histogram ($\epsilon = 0.025$)	[11.9, 20.5]	8.6
Keaney & Wei's resampling	[8.3, 24.1]	15.8
Su & Wei's min dispersion	[4.8, 24.9]	20.1
Bootstrap variance	[10.6, 21.8]	11.2
Bootstrap percentile	[9.0, 21.1]	12.1
Bias-corrected bootstrap percentile	[9.9, 21.2]	11.3