STATISTICAL INFERENCE FOR TUMOR GROWTH INHIBITION *T/C* RATIO

Jianrong Wu

Jianrong Wu, Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA;
Address correspondence to Jianrong Wu, Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678, USA; jianrong.wu@stjude.org

Copyright notice

Abstract

The tumor growth inhibition *T/C* ratio is commonly used to quantify treatment effects in drug screening tumor xenograft experiments. The *T/C* ratio is converted to an antitumor activity rating using an arbitrary cutoff point and often without any formal statistical inference. Here, we applied a nonparametric bootstrap method and a small sample likelihood ratio statistic to make a statistical inference of the *T/C* ratio, including both hypothesis testing and a confidence interval estimate. Furthermore, sample size and power are also discussed for statistical design of tumor xenograft experiments. Tumor xenograft data from an actual experiment were analyzed to illustrate the application.

Keywords: Bootstrap, Confidence interval, Hypothesis, Likelihood ratio, Log-normal, Sample size, Tumor growth inhibition, Xenografts

1. INTRODUCTION

In preclinical drug screening tumor xenograft experiments, human cancer cells are engrafted into mice to produce xenograft models. After tumors grow to a certain size, tumor-bearing mice are randomized into control (*C*) and treatment (*T*) groups, and the maximum tolerated dose of the drug is administered. The volume of each tumor (per mouse) is measured at the initiation of the study and periodically throughout the study. Mice are euthanized usually when the tumor volume reaches four times its initial volume, thus resulting in incomplete longitudinal tumor volume data. Statistical analysis of tumor xenograft data plays an important role in assessing the antitumor activity of the treatment. However, analysis of such data presents several statistical challenges, such as a limited number of available samples and missing data. To avoid missing data difficulties, one approach is to assess the treatment effect at the last time point with complete observations of both control and treatment groups. The treatment-to-control ratio (*T/C*) at the last time point with complete observations is then used to quantify the antitumor activity of the drug. The *T/C* ratio is converted to an antitumor activity rating using an arbitrary cutoff point (i.e., *T/C* ≤ 0.45) without any formal statistical inference (Corbett et al., 2003; Houghton et al., 2007). The disadvantages of this approach are obvious. It discards useful data and could result in a considerable loss of efficency. Furthermore, the degree of antitumor activity is

assessed on the basis of an arbitrary cutoff point, which ignores the noise in the data. Several new statistical methods have recently been developed to use the observed data in a more efficient way. Vardi et al. (2001) proposed two nonparametric two-sample *U*-tests. The proposed tests, however, assess the treatment effect by the cross-treatment difference instead of ratio and yields only a *p*-value. Tan et al. (2002) proposed a small-sample *t*-test. They deal with the missing data via the expectation maximization algorithm. The method, however, was derived on the basis of restrictive model assumption: a multivariate normal distribution for the repeated log tumor volumes with a Toeplitz covariance matrix. Furthermore, the method generates the *p*-value only and does not quantify the tumor growth inhibition. That is why the *T/C* ratio is still widely used in drug screening tumor xenograft data analysis (Atadja et al., 2004; Bissery et al., 1991; Corbett et al., 2003; Houghton et al., 2007). Here, we propose a valid statistical inference for the *T/C* ratio to assess the treatment effect, instead of using an arbitrary cutoff point.

Hothorn (2006) proposed an interval approach for the *T/C* ratio. Antitumor activity is assessed by the upper limit of the confidence interval of the *T/C* ratio. Hothorn's interval estimate of the *T/C* ratio is obtained on the basis of an assumed normal distribution of the tumor volume. Although Hothorn pointed out that a log-normal distribution could be used for inference of the ratio, there was no further discussion in his report. Control tumors often follow an exponential growth curve. Therefore, a log-normal distribution of tumor volume is a more reasonable assumption (Heitjan et al., 1993; Tan et al., 2002). For small-sample tumor xenograft data, however, the underlying distribution is sometime difficult to assess. Therefore, we propose a nonparametric bootstrap method and a small-sample likelihood ratio method to make a statistical inference of the *T/C* ratio. If the underlying distribution is difficult to assess, then the nonparametric bootstrap method can be used. If a log-normal distribution can be assumed, then the small-sample likelihood ratio statistic can be used. Furthermore, sample size and power calculation are also discussed for the purpose of statistical design of tumor xenograft experiments. Tumor xenograft data from an actual experiment were analyzed to illustrate the application.

2. INFERENCE FOR *T/C* RATIO

Calculating a *T/C* ratio from raw tumor volumes could result in a biased estimate of the drug effect because of heterogeneous initial tumor volumes. Therefore, the raw tumor volume is first divided by its initial tumor volume to yield the relative tumor volume. For notation convenience, let V_d be the relative tumor volume of group *d* at a given time with mean θ_d and variance τ_d^2, where *d* = *T* or *C* to represent the treatment or control group, respectively. The ratio of means

$$\gamma = \frac{\theta_T}{\theta_C}$$

represents the *T/C* ratio. Suppose {V_{Ti}, *i* = 1,…, *n*} and {V_{Cj}, *j* = 1, …, *m*} are the observed relative tumor volumes at a given time for the treatment and control groups, respectively. Then the *T/C* ratio of γ can be estimated by the ratio of sample means,

$$\widehat{\gamma} = \frac{\overline{V}_T}{\overline{V}_C}$$

which quantifies the treatment effect ($0 < \widehat{\gamma} \leq 1$). In general, a small value of $\widehat{\gamma}$ indicates a strong treatment effect. The standard error of $\widehat{\gamma}$ can be estimated by the Delta method,

$$\widehat{se}(\widehat{\gamma}) \simeq \frac{1}{\overline{V}_C} \left(\frac{\widehat{\sigma}_T^2}{n} + \frac{\overline{V}_T^2}{\overline{V}_C^2} \frac{\widehat{\sigma}_C^2}{m} \right)^{1/2}$$

where $\overline{V}_d$ and $\widehat{\sigma}_d^2$ are the sample mean and variance of the relative tumor volumes of group d.

A significant treatment effect can be assessed by testing the following hypothesis:

$$H_0{:}\gamma = 1 \mathrm{vs.} H_a{:}\gamma < 1 \tag{1}$$

However, confidence intervals are often preferred over hypothesis-testing procedures because they provide information regarding the size of the effect and its uncertainty. Both hypothesis testing and interval estimate of the *T/C* ratio are discussed in this section under two scenarios: nonparametric inference and parametric inference.

2.1. Scenario 1: Nonparametric Inference—Bootstrap Method

To make the nonparametric bootstrap inference ([Efron and Tibshirani, 1993](#)) of *T/C* ratio γ, we take a log transformation of γ as

$$\theta = \log(\gamma) = \log(\mu_T) - \log(\mu_C)$$

An estimate of the standard error of $\hat{\theta} = \log(\widehat{\gamma})$ is given by

$$\widehat{se}(\hat{\theta}) \simeq \left(\frac{1}{\overline{V}_T^2} \frac{\widehat{\sigma}_T^2}{n} + \frac{1}{\overline{V}_C^2} \frac{\widehat{\sigma}_C^2}{m} \right)^{1/2} \tag{2}$$

Because the only change of interest is the tumor volume reduction after treatment, a one-sided confidence interval will be constructed.

A bootstrap *p*-value and $100(1 - \alpha)\%$ one-sided confidence upper limit of the bootstrap interval can be estimated by following bootstrap procedures.

1. Generate B independent bootstrap samples of relative tumor volume from each group, $V_C^{*b} = \{V_{1C}^{*b}, \ldots, V_{mC}^{*b}\}$ and $V_T^{*b} = \{V_{1T}^{*b}, \ldots, V_{nT}^{*b}\}$, $b = 1, \ldots, B$.

2. Compute the bootstrap replication $\widehat{\gamma}^{*b}$, where $\widehat{\gamma}^{*b} = \overline{V}_T^{*b}/\overline{V}_C^{*b}$ for $b = 1, \ldots, B$.

3. A $100(1 - \alpha)\%$ bootstrap *t*-interval upper limit is obtained directly from the bootstrap sample,

$$t^{*b} = \frac{\hat{\theta}^{*b} - \hat{\theta}}{\widehat{se}^{*b}}, b = 1, \ldots, B$$

where $\hat{\theta}^{*b} = \log(\hat{\gamma}^{*b})$ and $\widehat{se}^{*b}$ are calculated using [Eq. (2)](#) for the bootstrap Sample $\{V_T^{*b}, V_C^{*b}, b = 1, \ldots, B\}$ and $\hat{\theta} = \log(\hat{\gamma})$. Let the αth percentile of $\{t^{*b}, b = 1, \ldots, B\}$ be estimated by the value of $\hat{t}_\alpha$ such that $\#\{t^{*b} < \hat{t}_\alpha\}/B = \alpha$. Then an one-sided upper limit of the bootstrap t-interval is given by

$$\hat{\gamma}_U = \hat{\gamma}\exp(-\hat{t}_\alpha \widehat{se}(\hat{\theta}))$$

4. A bootstrap p-value for hypothesis ([1](#)) is obtained by

$$p\text{ - Value} = \#\{t^{*b} < t_{\text{obs}}\}/B$$

Where $t_{obs} = \log(\hat{\gamma})/\widehat{se}(\hat{\theta})$.

2.2. Scenario 2: Parametric Inference—Small-Sample Likelihood Method

In this scenario, we assume $X_d = \log(V_d) \sim N(\mu_d, \sigma_d^2)$ with unequal variances between two groups; that is, $\sigma_T^2 \neq \sigma_C^2$. The relationship between the mean μ_d and variance σ_d^2 of log-transformed variable X_d and mean θ_d and variance τ_d^2 of original variable V_d is given by $\theta_d = e^{\mu_d + \sigma_d^2/2}$ and $\tau_d^2 = e^{2\mu_d + \sigma_d^2}(e^{\sigma_d^2} - 1)$, also given in [Table 1](#). The T/C ratio $\gamma = \theta_T/\theta_C$ is given by $\gamma = \exp(\mu_T - \mu_C + 0.5(\sigma_T^2 - \sigma_C^2))$. Then hypothesis ([1](#)) is equivalent to the following hypothesis:

$$H_0^\dagger : \psi = 0 \text{vs. } H_a^\dagger : \psi < 0$$

where $\psi = \log(\gamma) = \mu_T - \mu_C + 0.5(\sigma_T^2 - \sigma_C^2)$. The corresponding log-likelihood function $\ell(\psi, \lambda)$ is given by

$$\ell(\psi, \lambda) = -n\log\sigma_T - m\log\sigma_C + \{\psi + \mu_C - 0.5(\sigma_T^2 -$$

$$- \frac{1}{2\sigma_T^2}t_3 - \frac{1}{2\sigma_C^2}t_4 - \{\psi + \mu_C - 0.5(\sigma_T^2 - \sigma_C^2)\}^2$$

where $t = (t_1, t_2, t_3, t_4) = (\Sigma X_{Ti}, \Sigma X_{Cj}, \Sigma X_{Ti}^2, \Sigma X_{Cj}^2)$ is a minimum sufficient statistic and $\lambda = (\mu_C, \sigma_T, \sigma_C)$ is a vector nuisance parameter. The hypothesis can be tested by the signed log-likelihood ratio statistic r ([Cox and Hinkley, 1974](#)), which is simplified as

$$r(\psi) = \text{sgn}(\hat{\psi} - \psi)\left\{n\log\frac{\hat{\sigma}_{T\psi}^2}{\hat{\sigma}_T^2} + m\log\frac{\hat{\sigma}_{C\psi}^2}{\hat{\sigma}_C^2} + n(\overline{X}_T - \hat{\xi}_\psi)\left(1 - \frac{\hat{\sigma}_{C\psi}^2}{\hat{\sigma}_{T\psi}^2}\right)\right\}^{1/2}$$

where $\hat{\xi}_\psi = \psi + \widehat{\mu}_{C\psi} - 0.5(\widehat{\sigma}^2_{T\psi} - \widehat{\sigma}^2_{C\psi})$, and $(\widehat{\psi}, \widehat{\lambda}) = (\widehat{\psi}, \widehat{\mu}_C, \widehat{\sigma}_T, \widehat{\sigma}_C)$ and $(\psi, \widehat{\lambda}_\psi) = (\psi, \widehat{\mu}_{C\psi}, \widehat{\sigma}_{T\psi},$ $\widehat{\sigma}_{C\psi})$ are the maximum likelihood estimate and constrained maximum likelihood estimate for a given ψ, respectively. It is well known that under the null hypothesis of $H_0^\dagger$, r, is approximately distributed as a standard normal, but this is not accurate for small-samples. Therefore, a small-sample likelihood ratio statistic ([Barndorff-Nielsen and Cox, 1994](#)) is constructed as follows to make the statistical inference for ψ:

$$r^*(\psi) = r(\psi) + r(\psi)^{-1}\log\left\{\frac{u(\psi)}{r(\psi)}\right\}$$

where $r(\psi)$ is the signed log-likelihood ratio and $u(\psi)$ is given by [Fraser et al. (1999)](#) as follows:

$$u(\psi) = \frac{|\ell_{;t}(\widehat{\psi}, \widehat{\lambda}) - \ell_{;t}(\psi, \widehat{\lambda}_\psi)\ell_{\lambda;t}(\psi, \widehat{\lambda}_\psi)|}{|\ell_{\psi,\lambda;t}(\widehat{\psi}, \widehat{\lambda})|}\left\{\frac{|j(\widehat{\psi}, \widehat{\lambda})|}{|j_{\lambda\lambda}(\psi, \widehat{\lambda}_\psi)|}\right\}^{1/2}$$

where the sample-space derivatives $\ell_{;t}(\psi, \lambda)$ and mixed derivatives $\ell_{\lambda;t}(\psi, \lambda)$ are given by

$$\ell_{;t}(\psi, \lambda) = \left(\frac{1}{\sigma_T^2}\left\{\psi + \mu_C - \frac{1}{2}(\sigma_T^2 - \sigma_C^2)\right\}, \frac{\mu_C}{\sigma_C^2}, -\frac{1}{2\sigma_T^2}, -\frac{1}{2\sigma_C^2}\right)'$$

and

$$\ell_{\lambda;t}(\psi, \lambda) = \begin{pmatrix} \frac{1}{\sigma_T^2} & -\frac{2}{\sigma_T^3}\left(\psi + \mu_C + \frac{1}{2}\sigma_C^2\right) & \frac{\sigma_C}{\sigma_T^2} \\ \frac{1}{\sigma_C^2} & 0 & -\frac{2\mu_C}{\sigma_C^3} \\ 0 & \frac{1}{\sigma_T^3} & 0 \\ 0 & 0 & \frac{1}{\sigma_C^3} \end{pmatrix}$$

respectively. The observed nuisance information matrix is a symmetric matrix with upper right submatrix as follows:

$$j_{\lambda\lambda}(\psi, \lambda_\psi)$$

$$= \begin{pmatrix} \dfrac{n}{\widehat{\sigma}_{T\psi}^2} + \dfrac{m}{\widehat{\sigma}_{C\psi}^2} & -\dfrac{n}{\widehat{\sigma}_{T\psi}} - \dfrac{2n}{\widehat{\sigma}_{T\psi}^3}(\widehat{\xi}_\psi - \overline{X}_T) & \dfrac{n\widehat{\sigma}_{C\psi}}{\widehat{\sigma}_{T\psi}^2} - \dfrac{2m}{\widehat{\sigma}_{C\psi}^3}(\widehat{\mu}_{C\psi} - \overline{X}_C) \\[2ex] \star & n + \dfrac{2n}{\widehat{\sigma}_{T\psi}^2} & -\dfrac{n\widehat{\sigma}_{C\psi}}{\widehat{\sigma}_{T\psi}} - \dfrac{2n\widehat{\sigma}_{C\psi}}{\widehat{\sigma}_{T\psi}^3}(\widehat{\xi}_\psi - \overline{X}_T) \\[2ex] \star & \star & \dfrac{2m}{\widehat{\sigma}_{C\psi}^2} + \dfrac{n\widehat{\sigma}_{C\psi}^2}{\widehat{\sigma}_{T\psi}^2} + \dfrac{4n}{\widehat{\sigma}_{T\psi}^2}(\widehat{\xi}_\psi - \overline{X}_T) \end{pmatrix}$$

The determinants of the observed information matrix and mixed derivative matrix are given by $|j(\widehat{\psi}, \widehat{\lambda})| = 4n^2m^2/\widehat{\sigma}_T^4\widehat{\sigma}_C^4$ and $|\ell_{\psi,\lambda;t}(\widehat{\theta})| = 1/\widehat{\sigma}_T^5\widehat{\sigma}_C^5$. Detailed derivations of these quantities were given by Wu et al. (2002). The small-sample likelihood ratio statistic r^* is also approximately normally distributed but it is highly accurate even for small-samples (Barndorff-Nielsen, 1991). Therefore, a one-sided p-value for testing hypothesis (1) is given by

$$p - \text{Value} = P(r^* < r_0^*) = 1 - \Phi(r_0^*)$$

where r_0^* is the observed value of r^* under the null hypothesis. The one-sided upper limit of the 100(1 −α)% confidence interval of ψ is obtained by solving $r^*(\widehat{\psi}_U) = z_{1-\alpha}$; therefore, the upper limit of γ is given by

$$\widehat{\gamma}_U = \exp(\widehat{\psi}_U)$$

Table 1

Parameters of normal and log-normal distribution

Variable	Distribution	Mean	Variance	T/C ratio
$X_d = \log(V_d)$	Normal	μ_d	σ_d^2	
V_d	Log-normal	$\theta_d = e^{\mu_d + \sigma_d^2/2}$	$\tau_d^2 = e^{2\mu_d + \sigma_d^2}(e^{\sigma_d^2} - 1)$	$\gamma = \theta_T/\theta_C$

3. REAL TUMOR XENOGRAFT EXPERIMENTAL DATA

For an example, data from a real tumor xenograft experiment conducted by the Pediatric Preclinical Testing Program (PPTP) (Kolb et al., 2008) were used to illustrate the methods proposed in previous section. In this study, human antibody SCH 717454 at a dose of 0.5mg per mouse was administered twice per week for 4 weeks by intraperitoneal injection with a 2-week follow-up. Table 2 shows the relative tumor volumes of 16 mice (8 each in the control and treatment groups) from week 1 to week 5 with complete observations for tumor line BT-50. To check the normality assumption of the log tumor

volume, the Shapiro–Wilk tests gave *p*-values of 0.646 and 0.794 for the control and treatment groups, respectively, at week 4 and 0.500 and 0.845 at week 5. Q–Q plots are shown in [Figure 1](#). Both Shapiro–Wilk tests and Q–Q plots suggested that the normal distribution assumption of the log tumor volume was satisfied. Furthermore, *F*-tests were conducted to check the equal variance assumptions. Because of the nonrobustness of the *F*-test against normality assumption, permutation tests were also conducted to assess the equal variance of two groups. The *F*-tests and permutation tests respectively gave *p*-values of 0.814 and 0.801 for week 4 and 0.023 and 0.008 for week 5. Therefore, equal variance assumptions were satisfied for log tumor volumes between control and treatment at week 4 but not at week 5. For week 4 data, the estimate (standard error) of the *T/C* ratio was 0.948 (0.097), the two-sample *t*-test gave a *p*-value of 0.682, and the 95% confidence interval upper limit was 1.991. Neither the *p*-value nor the upper limit showed a significant treatment effect at week 4. For week 5 data, the estimate (standard error) of the *T/C* ratio was 0.508 (0.076), the small-sample likelihood ratio test gave a *p*-value of 0.007 (Bonferroni adjusted *p*-value = 0.014), and the 95% confidence interval upper limit was 0.927. Both the *p*-value and upper limit showed a significant treatment effect at week 5. The Shapiro–Wilk test may lack power to assess the log-normal distribution assumption with small samples. Therefore, the nonparametric bootstrap inference is attractive if the log-normal distribution assumption is not guaranteed. The bootstrap *t*-tests gave *p*-values of 0.327 and 0.002 (Bonferroni adjusted *p*-value = 0.004) for week 4 and week 5 data, respectively. The upper limits of bootstrap *t*-intervals were 1.157 and 0.684 for week 4 and week 5 data, respectively. Therefore, both parametric small-sample likelihood inference and nonparametric bootstrap inference gave the same conclusions. The results are summarized in [Table 3](#).
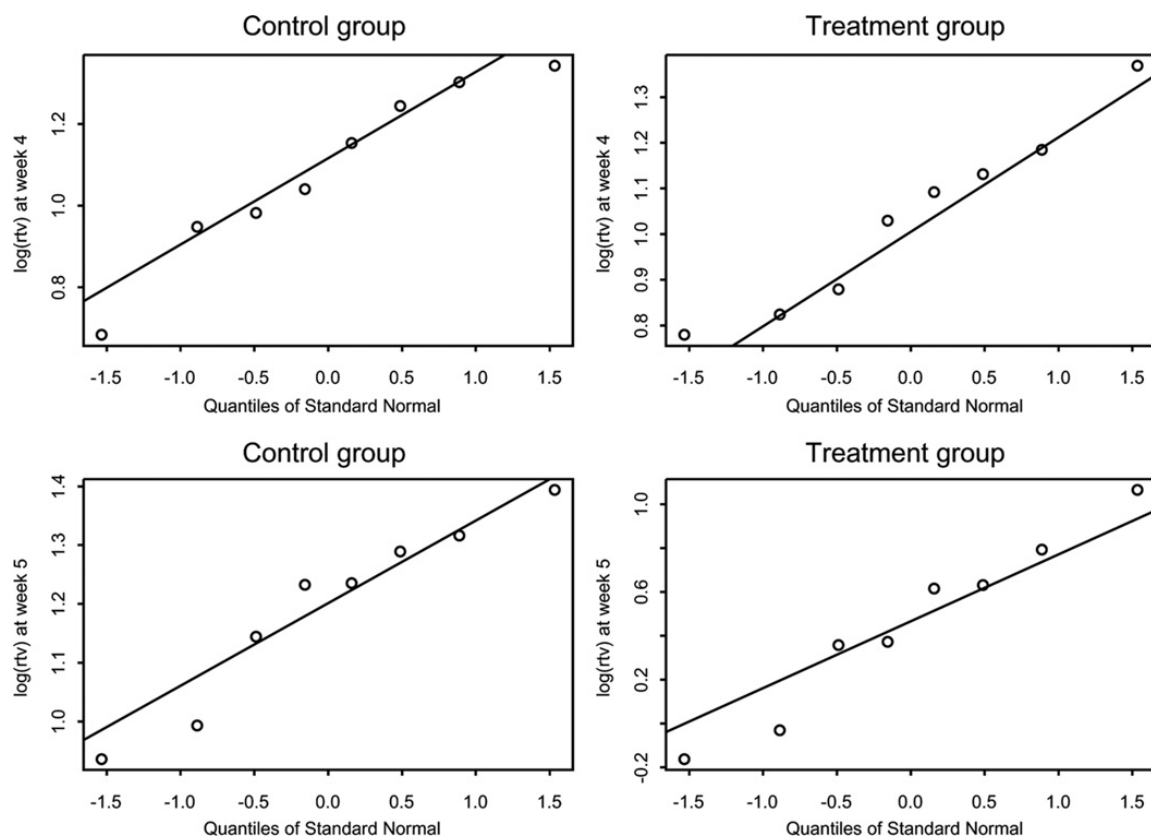

[Figure 1](#)

Q–Q plot of log relative tumor volume for week 4 data and week 5 data.

Table 2

Relative tumor volumes of BT-50 SCH 717454 tumor xenograft model

Group	Week	Mouse 1	2	3	4	5	6	7	8
Control	1	1.02	1.08	1.15	1.03	1.92	1.47	1.77	1.20
	2	1.77	1.64	2.68	1.89	1.95	1.71	2.45	1.53
	3	3.13	3.23	2.68	2.36	2.45	2.06	2.69	1.89
	4	3.47	3.68	2.58	2.67	3.17	2.83	3.83	1.98
	5	3.63	4.03	3.44	2.70	3.43	3.14	3.73	2.55
Drug	1	1.38	1.63	1.21	1.70	1.07	1.18	1.58	1.65
	2	2.56	1.81	1.73	2.10	1.56	1.45	2.69	2.13
	3	2.69	2.13	1.8	2.00	2.06	2.09	2.54	2.96
	4	2.80	3.27	2.41	3.10	2.18	2.28	2.98	3.93
	5	1.88	2.21	0.85	1.45	1.43	0.97	2.90	1.85

Table 3

Data analysis summary of BT-50 SCH 717454 tumor xenograft data

Week	F-test (P-test[a])	T/C ratio (se)	Likelihood method $\widehat{\gamma}_U$	p-Value	Bootstrap method $\widehat{\gamma}_U$	p-Value
4	0.814 (0.801)	0.948 (0.097)	1.991	0.682	1.157	0.327
5	0.023 (0.008)	0.508 (0.076)	0.927	0.007	0.684	0.002

[a]Permutation test based on 2,000 random permutation samples.

4. SAMPLE SIZE

Preclinical tumor xenograft experiments are often conducted without adequate consideration of the study power and sample size. The number of mice used in tumor xenograft studies tends to be quite arbitrary. For example, 5 and 7 (Thompson et al., 1999), 10 (Houghton et al., 2007), and 21 (Hothorn, 2006) mice per group were used in respected studies. There is little literature available regarding the

sample size and power calculations for tumor xenograft experiments. To provide some guidelines for statistical design for such experiments, sample size and power calculations are discussed based on the hypothesis of *T/C* ratio for tumor growth inhibition studies.

To formulate the sample size calculation, assume equal numbers of mice (*n*) to be tested in each group, and the log tumor volume $X_d = \log(V_d)$ follows a normal distribution with mean μ_d and common variance σ^2 between the two groups. Then it is easy to see that the following relationships hold ([van Belle and Martin, 1993](#)):

$$\gamma = \exp(\mu_T - \mu_C) \quad \text{and} \quad \sigma^2 = \log(1 + CV^2)$$

where *CV* is the common coefficient of variation of V_d.

Define

$$T = \sqrt{\frac{n}{2}} \, \frac{\overline{X}_T - \overline{X}_C}{S}$$

where $\overline{X}_d$ and S^2 are the sample mean and pooled sample variance, respectively. Under the alternative H_a for the hypothesis ([1](#)), *T* follows a noncentral *t*-distribution with a noncentral parameter δ and degree freedom of $2n - 2$,

$$T = \frac{\sqrt{\frac{n}{2}}[\overline{X}_T - \overline{X}_C - (\mu_T - \mu_C)]/\sigma + \sqrt{\frac{n}{2}}(\mu_T - \mu_C)/\sigma}{S/\sigma} \sim T_{2n-2}(\delta)$$

where the noncentral parameter δ is given by

$$\delta = \frac{\sqrt{\frac{n}{2}}(\mu_T - \mu_C)}{\sigma} = \frac{\sqrt{\frac{n}{2}}\log(\gamma)}{[\log(1 + CV^2)]^{1/2}}$$

Therefore, the power or sample size can be calculated from the following equation:

$$\text{Power} = P(T_{2n-2}(\delta) < t_{\alpha,2n-2})$$

where $t_{\alpha,2n-2}$ is the αth percentile of *t*-distribution with degree freedom of $2n - 2$. The power or sample size calculation has been implemented in SAS/PROC POWER. The sample sizes per group calculated for various *CV* and *T/C* values are listed in [Table 4](#). From this calculation, we can conclude that for a moderate *CV* (0.6) value, groups of 9 mice each have at least 80% power to detect a 50% reduction of mean relative tumor volume. Investigators can select the sample size for their experiments based on the hypothesis of effect size and historical information of the coefficient of variation of the tumor volume.

Table 4

Number of animals per group to attain a power of 80% for a one-sided test with α = 0.05

				T/C				
CV	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
0.1	2	2	2	2	2	2	3	4
0.2	2	2	2	2	3	3	5	11
0.3	2	2	2	3	4	5	10	23
0.4	2	2	3	4	5	8	16	38
0.5	2	3	3	5	7	12	23	57
0.6	3	3	4	6	9	16	31	78
0.7	3	3	5	7	12	20	40	100
0.8	3	4	6	9	14	25	49	124
0.9	3	4	6	10	16	29	59	149
1.0	3	5	7	11	19	34	69	173

Note. *CV*, common coefficient of variation of the relative tumor volume of two groups.

5. DISCUSSION AND CONCLUSION

Two methods have been proposed to make a statistical inference for the tumor growth inhibition *T/C* ratio. If a log-normal distribution of tumor volume can be assumed, then the small-sample likelihood inference is suitable. If the underlying distribution of tumor volume is difficult to assess, then the nonparametric bootstrap method can be used. Furthermore, sample sizes were derived for the purpose of statistical design of drug screening tumor xenograft experiments. The proposed methods are easy to implement. An S-plus code for the analysis is available from the author. However, antitumor activity assessment based on the *T/C* ratio could result in substantial efficiency loss if at least one observation missing occurs early in the experiment. The methods proposed by Vardi et al. (2001) and Tan et al. (2002) are more efficient for handling the missing data. However, neither method quantifies the treatment effect, yielding only a *p*-value with no confidence interval. Longitudinal modeling of tumor xenograft data is attractive (Heitjan et al., 1993). It incorporates the correlation structure of the repeated observation and is valid under the assumption of data missing at random. However, longitudinal modeling requires more restrictive assumptions, such as normality (log-normality), autoregressive correlation, and growth curve. Finally, a reviewer pointed out that in practice it is more important to investigate the homogeneity of the *T/C* ratio across several tumor models. Johnson et al. (2001, p. 1430) pointed out that "activity in multiple xenograft models is a useful predictor of clinical activity." The work of Voskoglou-Nimikos et al. (2003) appears to support this conclusion.

ACKNOWLEDGMENTS

REFERENCES

1. Atadja P, Gao L, Kwon P, Trogani N, Walker H, Hsu M, Yeleswarapu L, Chandramouli N, Perez L, Versace R, Wu A, Sambucetti L, Lassota P, Cohen D, Bair K, Wood A, Remiszewski S. Selective growth inhibition of tumor cells by a novel histone deacetylase inhibitor, NVP-LAQ824. *Cancer Res.* 2004;64:689–695. [PubMed] [Google Scholar]
2. Barndorff-Nielsen OE. Modified signed log likelihood ratio. *Biometrika.* 1991;78:557–563. [Google Scholar]
3. Barndorff-Nielsen OE, Cox DR. *Inference and Asymptotics.* London: Chapman and Hall; 1994. [Google Scholar]
4. Bissery MC, Guenard D, Gueritte-Voegelein F, Lavelle F. Experimental antitumor activity of taxotere (RP 56976, NSC 629503), a taxol analogue. *Cancer Res.* 1991;51:4845–4852. [PubMed] [Google Scholar]
5. Corbett TH, White K, Polin L, Kushner J, Paluch J, Shih C, Grossman CS. Discovery and preclinical antitumor efficacy evaluations of LY32262 and LY33169. *Investig. New Drugs.* 2003;21:33–45. [PubMed] [Google Scholar]
6. Cox DR, Hinkley DV. *Theoretical Statistics.* London: Chapman and Hall; 1974. [Google Scholar]
7. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* New York: Chapman and Hall; 1993. [Google Scholar]
8. Fraser DAS, Reid N, Wu J. A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika.* 1999;86:249–264. [Google Scholar]
9. Heitjan DF, Manni A, Santen RJ. Statistical analysis of *in vivo* tumor growth experiments. *Cancer Res.* 1993;53:6042–6050. [PubMed] [Google Scholar]
10. Houghton PJ, Morton CL, Tucker C, Payne D, Favours E, Cole C, Gorlick R, Kolb EA, Zhang W, Lock R, Carol H, Tajbakhsh M, Reynolds CP, Maris JM, Courtright J, Keir ST, Friedman HS, Stopford C, Zeidner J, Wu J, Liu T, Billups CA, Khan J, Ansher S, Zhang J, Smith MA. The pediatric preclinical testing program: Description of models and early testing results. *Pediatr. Blood Cancer.* 2007;49:928–940. [PubMed] [Google Scholar]
11. Hothorn L. Statistical analysis of *in vivo* anticancer experiments: Tumor growth inhibition. *Drug Inform. J.* 2006;40:229–238. [Google Scholar]
12. Johnson JI, Decker S, Zaharevitz D, Rubinstein LV, Venditti JM, Schepartz S, Kalyandrug S, Christian M, Arbuck S, Hollingshead M, Sausville EA. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. *Br. J. Cancer.* 2001;84:1424–1431. [PMC free article] [PubMed] [Google Scholar]
13. Kolb EA, Gorlick R, Houghton PJ, Morton CL, Lock R, Carol H, Reynolds CP, Maris JM, Keir ST, Billups CA, Smith MA. Initial testing (stage 1) of a monoclonal antibody (SCH 717454) against the IGF-1 receptor by the pediatric preclinical testing program. *Pediatr. Blood Cancer.* 2008;50:1190–1197. [PubMed] [Google Scholar]
14. Tan M, Fang HB, Tian GL, Houghton PJ. Small-sample inference for incomplete longitudinal data with truncation and censoring in tumor xenograft models. *Biometrics.* 2002;58:612–620. [PubMed] [Google Scholar]
15. Thompson J, George EO, Poquette CA, Cheshire PJ, Richmond LB, de Graaf SS, Ma M, Stewart CF, Houghton PJ. Synergy of topotecan in combination with vincristine for treatment of pediatric solid tumor xenografts. *Clin. Cancer Res.* 1999;5:3617–3631. [PubMed] [Google Scholar]

16. van Belle G, Martin DC. Sample size as a function of coefficient of variation and ratio of means. *Biometrics.* 1993;58:612–620. [Google Scholar]
17. Vardi Y, Ying ZL, Zhang CH. Two-sample tests for growth curves under dependent right censoring. *Biometrika.* 2001;88:949–960. [Google Scholar]
18. Voskoglou-Nimikos T, Pater JL, Seymour L. Clinical predictive value of the in vitro cell line, human xenograft and mouse allograft preclinical cancer models. *Clin. Cancer Res.* 2003;9:4227–4239. [PubMed] [Google Scholar]
19. Wn J, Jiang G, Wong ACM, Sun X. Likelihood analysis for the ratio of means of two independent log-normal distributions. *Biometrics.* 2002;58:463–469. [PubMed] [Google Scholar]