Published in final edited form as:

Pharm Stat. 2010 January ; 9(1): 46–54. doi:10.1002/pst.369.

Interval Approach to Assessing Antitumor Activity for Tumor Xenograft Studies

Jianrong Wu¹ and Peter J. Houghton²

¹ Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678, USA

² Department of Molecular Pharmacology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678, USA

Abstract

In preclinical cancer drug screening tumor xenograft experiments, the tumor growth inhibition ratio (T/C) is commonly used to assess the antitumor activity of the agents. Unfortunately, this measurement can discard useful data and result in a high false-negative rate. Furthermore, the degree of antitumor activity based on the T/C ratio is assessed on the basis of an arbitrary cutoff point that does not reflect variations in different tumor lines. To overcome these drawbacks, we propose an adjusted area-under-the-curve (aAUC) ratio to quantify tumor growth inhibition. A nonparametric bootstrap t -interval of the aAUC ratio is also proposed for assessing the significance of the anti-tumor activity of the agents. The proposed method is then applied to a real tumor xenograft study.

Keywords

area-under-the-curve; antitumor activity; bootstrap; confidence interval; xenografts

1. Introduction

In preclinical cancer drug screening tumor xenograft experiments, human cancer cells are engrafted into mice to produce xenograft models. Tumor-bearing mice are randomized into control (C) and treatment (T) groups, and the maximum tolerated dose of each drug is administered. The volume of each tumor (one tumor per mouse) is measured at the initiation of study and periodically throughout the study. Mice are euthanized when their tumor volume reaches four times its initial volume, thus resulting in incomplete longitudinal tumor volume data. Because of the incompleteness of the experimental data, the T/C ratio (Corbett, et al., 2003, Houghton, et al., 2007) calculated at a given time with complete observations is commonly used to quantify tumor growth inhibition, where T and C represent the means of the relative tumor volumes of the control and treatment mice, respectively. Here, the relative tumor volume of each tumor is defined as the tumor volume divided by its initial volume. This approach could, however, discard some observed data and result in a high false-negative rate. Furthermore, the degree of antitumor activity is assessed on the basis of an arbitrary cutoff point (Corbett, et al., 2003, Houghton, et al. 2007). Several new statistical methods have recently been developed for assessing antitumor activity in tumor xenograft models. Tan et al. (2002) proposed a small-sample t -test via the EM algorithm. They assumed a multivariate normal distribution for the repeated log tumor volumes with a Toeplitz covariance matrix. Due to the strong model assumption, their method has limited application to drug screening tumor xenograft data. Furthermore, the method yields only a p -value and does not quantify tumor growth inhibition. Vardi et al. (2001) proposed a nonparametric two-sample U-test. The proposed methodology is a fully nonparametric approach, which is generally applicable to drug

screening tumor xenograft data. The method, however, assesses the treatment effect by the cross-treatment difference instead of the ratio and yields a p -value only; no confidence intervals are available. Liang (2006) proposed a non-parametric approach to compare antitumor effects in two treatment groups. The approach is, in essence, a comparison of two tumor volume curves and yields a p -value only. Some sophisticated models have been proposed to fit the tumor growth curves; for example, a biexponential model (Demidenko, 2004, Liang and Sha, 2004), a LINEXP model (Demidenko, 2006), and a nonparametric model (Liang, 2005). However, it is often difficult to model tumor growth curves due to a short study period, small sample size, and diverse tumor growth patterns in tumor xenograft experiments. Hothorn (2006) proposed an interval approach for the T/C ratio. However, Hothorn's interval is obtained on the basis of an assumed normal distribution of the tumor volume. Recently, the Pediatric Preclinical Testing Program (PPTP) established panels of childhood cancer xenografts and tumor lines for *in vivo* screening (Houghton, et al, 2007). The objective of the PPTP study was to identify novel agents with significant antitumor activity. The PPTP has produced a large amount of tumor xenograft screening data. Therefore, the development of an appropriate methodology for analysis of tumor xenograft data is important for the evaluation of existing and new childhood anticancer agents. It is also the motivation of this paper. Here we propose an adjusted area-under-the-curve (aAUC) ratio to quantify tumor growth inhibition. The aAUC ratio, unlike the T/C ratio, uses all observed data and retains high efficiency for evaluating antitumor activity. Furthermore, a nonparametric bootstrap t -interval (Efron and Tibshirani, 1993) is developed for the aAUC ratio to assess the significance of antitumor activity of the tested agent, which avoids the specific assumptions about the distribution of tumor volume and the tumor growth curve.

2. Adjusted AUC Ratio and Interval Estimate

To make full use of the observed but incomplete tumor volume data, an aAUC is proposed. Tumor growth inhibition is then quantified by the ratio of the aAUCs of the treatment and control groups. To assess the significance of the treatment effect, two nonparametric bootstrap intervals are proposed in this section: a bootstrap bias-corrected (BC) interval and a bootstrap t -interval.

Consider a typical tumor growth experiment with two groups: control and treatment. Let $\xi(t)$ be the relative tumor volume growth curve at time $t \in T = [0, t^*]$ for a mouse; the associated right censoring time is τ . The censoring occurs because a mouse dies due to toxicity before the end of study or is euthanized when its tumor volume reaches four times its initial volume. Then the area-under-the-curve (AUC) could be calculated up to τ as

$$AUC = \int_{T(\tau)} \xi(t) \nu(dt),$$

where $T(a) = \{t: t \in T, t \leq a\}$ and ν is a suitable measure on T . The raw AUCs are not comparable due to uneven censoring. Therefore, the AUC is adjusted by the length of the interval between the initiation of the experiment and the last time point with existing tumor volume measurements, and the resulting the aAUC is

$$aAUC = \frac{1}{\tau} \int_{T(\tau)} \xi(t) \nu(dt).$$

It is often difficult to model the growth curve $\xi(t)$ explicitly due to diverse tumor growth patterns in xenograft experiments. Therefore, an empirical approach is used to calculate the

aAUC. Specifically, assuming that there are $k + 1$ measured relative tumor volumes ξ_i for a tumor at time t_i ($i = 0, \dots, k$), the AUC is calculated up to the last time point with existing tumor volume measurements using a simple trapezoidal rule,

$$\text{AUC} = \int_{t_0}^{t_k} \xi(t) dt \simeq \frac{1}{2} \sum_{i=0}^{k-1} (\xi_i + \xi_{i+1})(t_{i+1} - t_i).$$

Then the aAUC can be calculated as

$$\text{aAUC} = \frac{1}{t_k - t_0} \text{AUC}.$$

The normality assumption of the aAUC is unreliable due to the skewness of the aAUC and small sample size. Therefore, a nonparametric approach is more appropriate and useful.

To simplify the notation, let X_C and X_T be the aAUC of the control and treatment groups and μ_C and μ_T be the corresponding means. We can define the aAUC ratio, γ , as the ratio of the means of the aAUCs of the treatment and control groups; that is,

$$\gamma = \frac{\mu_T}{\mu_C},$$

which quantifies the tumor growth inhibition of the agent and can be estimated by

$$\hat{\gamma} = \frac{\bar{X}_T}{\bar{X}_C},$$

where $\bar{X}_C$ and $\bar{X}_T$ are the sample means of the control and treatment groups, respectively. To construct the interval of 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 $\hat{\theta} = \log(\hat{\gamma})$ can be obtained by the Delta method as

$$\widehat{se}(\hat{\theta}) \simeq \left(\frac{1}{\bar{X}_T^2} \frac{\widehat{\sigma}_T^2}{n_T} + \frac{1}{\bar{X}_C^2} \frac{\widehat{\sigma}_C^2}{n_C} \right)^{1/2}, \quad (1)$$

where $\widehat{\sigma}_C^2$, n_C and $\widehat{\sigma}_T^2$, n_T are the sample variances and sample sizes of the control and treatment groups, respectively. Because the only change of interest is the tumor volume reduction after treatment, a one-sided confidence interval will be constructed.

The $100(1-\alpha)\%$ one-sided upper confidence limits of bootstrap intervals can be estimated by following bootstrap procedures.

1. Generate B independent bootstrap samples of aAUC from each group,
 $X_C^{*b} = \{X_{1C}^{*b}, \dots, X_{n_{1C}}^{*b}\}$ and $X_T^{*b} = \{X_{1T}^{*b}, \dots, X_{n_{1T}}^{*b}\}$, $b = 1, \dots, B$.
2. Compute the bootstrap replication $\hat{\gamma}^{*b}$, where $\hat{\gamma}^{*b} = \bar{X}_T^{*b} / \bar{X}_C^{*b}$ for $b = 1, \dots, B$.
3. A $100(1-\alpha)\%$ bootstrap bias-corrected (BC) percentile interval upper limit is obtained directly from bootstrap distribution $\hat{G}(s) = \#\{\hat{\gamma}^{*b} < s\}/B$ of $\{\hat{\gamma}^{*b}, b = 1, \dots, B\}$ as

$$\hat{\gamma}_{upper} = \hat{G}^{-1}(\alpha_1),$$

where $\alpha_1 = \Phi(2z_0 + z_{1-\alpha})$ and $z_0 = \Phi^{-1}(\hat{G}(\hat{\gamma}))$, $\Phi(\cdot)$ is the standard normal distribution.

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

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

where $\hat{\theta}^{*b} = \log(\hat{\gamma}^{*b})$ and $\widehat{se}^{*b}$ is calculated using (1) for the bootstrap sample $\{X_T^{*b}, X_C^{*b}, b = 1, \dots, B\}$. Let the α^{th} percentile of $\{t^{*b}, b = 1, \dots, B\}$ be estimated by the value of $\hat{t}_\alpha$ such that $\#\{t^{*b} < \hat{t}_\alpha\} = \alpha$, then upper limit of the bootstrap t -interval is given by

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

3. Simulation Studies

In this section, we designed simulation studies to investigate the power to distinguish treatments by using three endpoints: a) the relative tumor volume at the last time point with complete data from the two groups (RTVC), b) the aAUC, and c) the cross-treatment difference (CTD) defined by Vardi et al. (2001). We also conducted a simulation to study the coverage probability of the proposed nonparametric bootstrap intervals for the aAUC ratio.

3.1 Comparison of Power

Two simulation scenarios are considered to assess the powers of the three endpoints mentioned above, one with equal censorship (scenario I), meaning the censoring times for the two groups have the same distribution. However, this kind of homogeneity assumption may be unrealistic in some circumstances. For example, treatment group mice could have a slower growth rate and higher mortality rate than control group mice. Thus, it is desirable to consider heterogeneous censorship (scenario II), with different censoring distributions for the two groups. To yield simulation data similar to the real tumor xenograft data, tumor volumes for both scenarios were generated by following exponential growth curves, which are the typical for control mice. The error terms reflect measurement errors and resulting correlated longitudinal tumor volumes. The σ measures the departure from the exponential form of the growth curve. We generated 10 tumor growth curves for each group with tumor volume at each week for a 6-week study period. Tumor volumes were treated as missing when the tumor volume exceeded four times its initial volume.

1. *Scenario I:* Group 1 censored growth curves are generated from

$$\xi(t) = \exp\left(\delta t + \sigma \sum_{s=1}^t e_s\right), \quad \tau = \inf_{t \in \{1, 2, 3, 4, 5\}} \{t: \xi(t) \geq 4 \text{ or } t=6\},$$

and Group 2 censored growth curves are generated from

$$\zeta(t) = \exp\left(\lambda t + \sigma \sum_{s=1}^t \varepsilon_s\right), \quad \eta = \inf_{t \in \{1, 2, 3, 4, 5\}} \{t: \zeta(t) \geq 4 \text{ or } t=6\},$$

2. *Scenario II*: the same growth curves and censoring times are generated according to Scenario I, with the exception of τ , which is replaced by

$$\tau = \inf_{t \in \{1, 2, 3, 4, 5\}} \{t: \xi(t) \geq 4 \text{ or } t \leq C\},$$

where C is a random time independent of $\{\xi(t), t \leq 6\}$, and distributed as the integer part of $6U^{1/2} + 1$ with a uniform random variable U in $(0, 1)$.

Both scenarios assume that $\{e_t, \varepsilon_t, t \geq 1\}$ are independent $N(0, 1)$ variables, with $\xi(0) = \zeta(0) = 1$, $\sigma = 0.3, 0.5$, and $\delta = 0$ and λ is specified in Table 1.

To make the power comparisons meaningful, a randomization test was used for three endpoints to test equal treatment effects, which yields type I errors at a fixed level, say 0.05. Let X_C and X_T be the one of three endpoints for the control and treatment groups, respectively, and S_{n_C, n_T} be the corresponding test statistic, where n_C and n_T are the sample sizes of the control and treatment groups, respectively. In the following simulations, a two-sample t -statistic,

$$S_{n_C, n_T}(X_C, X_T) = (\bar{X}_C - \bar{X}_T) / \sqrt{\hat{\sigma}_C^2/n_C + \hat{\sigma}_T^2/n_T},$$

was used for the endpoints of RTVC and aAUC,

and two-sample U-statistic, $S_{n_C, n_T}(X_C, X_T) = \sum_{i=1}^{n_C} \sum_{j=1}^{n_T} D_{ij} / (n_C n_T)$, was used for the endpoint of cross-treatment difference, where D_{ij} is the cross-treatment difference defined by Vardi et al. (2001). The algorithm to calculate a p -value using the randomization test is as follows:

1. Choose M independent index vectors $g^*(1), g^*(2), \dots, g^*(M)$, each of which contains n_C of $X_C = \{X_{1C}, \dots, X_{n_{CC}}\}$ indexes and n_T of $X_T = \{X_{1T}, \dots, X_{n_{TT}}\}$ indexes, which are randomly selected from the set of all possible permuted vectors.
2. Evaluate test statistic S_{n_C, n_T} for each randomization sample of $(g^*(b), X_C, X_T)$ obtained in step 1, such that

$$S_{n_C, n_T}^*(b) = S_{n_C, n_T}(g^*(b), X_C, X_T), \quad b = 1, \dots, M,$$

where $S_{n_C, n_T}(g^*(b), X_C, X_T)$ is a realization of S_{n_C, n_T} by taking the first n_C vectors of the combined sample (X_C, X_T) according to the index $g^*(b)$ as the X_C and the rest as X_T .

3. A two-sided p -value is given as

$$\hat{p} = \frac{\sum_{b=1}^M I(|S_{n_C, n_T}^*(b) - S_{m, n}^{(\cdot)}| \geq |S^0 - S_{n_C, n_T}^{(\cdot)}|)}{M},$$

where $S^0 > 0$ is the observed value of S_{n_C, n_T} and $S_{n_C, n_T}^{(\cdot)} = \sum_{b=1}^M S_{n_C, n_T}^*(b)/M$.

The proportion of rejections of the null hypothesis $H_0: \delta = \lambda$, or equal treatment effect, represents the empirical power of the test. For a prespecified significance level $\alpha = 0.05$ and sample size of 10 in each group, Table 1 lists the results based on 10,000 simulation runs and 2,000 randomization samples. The last two columns of the table represent the average proportions of missing measurements (MP) of the two groups and the average of the last time points with complete data (LTC) of the two groups. For both scenarios, Table 1 shows that the powers of the RTVC endpoint are much lower than those of the aAUC or CTD endpoints when the LTC is less than 4 weeks if $\sigma = 0.3$ and 2 weeks if $\sigma = 0.5$. That shows that using the tumor volume at the last time point with complete data could be very inefficient if some mice died or were euthanized early in the experiment. The empirical power of the proposed aAUC is comparable to that of the cross-treatment difference endpoint, which has been shown to be highly efficient by Vardi et al. (2001).

3.2 Comparison of Coverage Probability

The coverage probability is the probability that a confidence interval captures the true parameter, and it is estimated here as the proportion of cases in a simulation in which the calculated interval includes the true value. We chose to use the parameter configurations to represent different types of antitumor activity, from low to high activity. We calculated one-sided upper confidence limits. The sample size considered in the simulation was $n=10$ for each group, which is a typical sample size for tumor xenograft experiments. The standard deviations (σ_C, σ_T) of a normal or log-normal distribution used were (0.05, 0.05), (0.1, 0.1), (0.5, 0.5), (0.1, 0.2), and (0.2, 0.1). The mean parameters of a normal or log-normal distribution and shape (α) and rate (λ) parameters of a gamma distribution for the simulation study are listed in Table 2. For each parameter configuration, we generated 10,000 random samples from a normal distribution $N(\mu_g, \sigma_g^2)$, or a log-normal distribution $LN(\mu_g, \sigma_g^2)$, or a gamma (α_g, λ_g) distribution, where $g = C, T$, and used 2,000 bootstrap samples. The simulated 95% coverage probabilities are shown in Table 3. The simulated results showed that the bootstrap BC interval had a low coverage probability, whereas the coverage probabilities of the bootstrap t -interval were close to the target nominal level (0.95), regardless of whether the underlying distribution was normal or skewed log-normal or gamma distributions.

4. A Real Tumor Xenograft Model

For a real example, we used data from a published single-agent tumor xenograft study conducted by the PPTP (Houghton, et al, 2007). In this study, vincristine was administered intraperitoneally at a dosage of 1 mg/kg every 7 days for 6 weeks and evaluated in 38 tumor cell lines as a single agent. For each tumor cell line, nearly 20 mice were equally randomized into control and treatment groups. Table 4 shows 22 solid tumor cell lines together with the T/C ratios and antitumor activity ratings published by Houghton et al. The T/C ratio was calculated on day 21 or when all mice in both groups still had measurable tumor volumes if less than 21 days. Table 4 also shows the estimated aAUC ratios, standard errors, and 95% upper limits of the bootstrap t -intervals. The PPTP study used arbitrary cutoff points for the T/C ratio antitumor activity rating. Agents producing a $T/C \leq 0.15$ were considered highly active; those with $T/C > 0.15$ but ≤ 0.45 were considered to have intermediate activity; and those with $T/C > 0.45$ were considered to have low activity. The drawback to the use of such arbitrary cutoff point is obvious: it fails to consider the variations among different cell lines. Assessment of antitumor activity based on the interval approach is more intuitive. Whenever the upper confidence limit is lower than 1, significant antitumor activity is observed; otherwise, no significant antitumor activity can be claimed. From Table 4, we see that the antitumor activity assessments of the two methods produced agreement for 16 tumor lines, but not for

four cell lines, Rh10, Rh30, Rh30R, and NB-1771, in which the T/C ratio showed a low activity but the aAUC ratio led to an active evaluation. In one cell line, D212, the T/C ratio showed an intermediate activity but the aAUC ratio led to an inactive evaluation. The discrepancy between the T/C ratio and the aAUC ratio for cell lines Rh10, Rh30, and Rh30R was because the T/C ratio was calculated early in the study and thus failed to capture the antitumor activity seen later (Figure 1). In contrast, the discrepancy between the T/C ratio and the aAUC ratio for cell lines NB-1771 and D212 was because the T/C ratio did not account for the variability of the different tumor lines, therefore leading to false-negative (NB-1771) and false-positive (D212) antitumor activity evaluations.

5. Discussion and Conclusion

The analysis of tumor xenograft experimental data presents several statistical challenges, such as incomplete longitudinal observations, small samples, and diverse tumor growth patterns. Existing methods either are based on rather restrictive normality assumptions or discard useful data and result in a loss of efficiency. The proposed aAUC ratio fully uses the experimental data and is therefore more efficient for evaluating antitumor activity. Furthermore, the proposed bootstrap t -interval is a nonparametric approach, which avoids the underlying distribution assumption and has a good coverage probability even for sample size as small as 10 per group. The upper limit of the bootstrap t -interval defines the effect of size and variability and therefore avoids the use of an arbitrary cutoff point. For the final conclusion, the proposed aAUC ratio not only efficiently quantifies tumor growth inhibition but also gives a simple assessment of antitumor activity for tumor xenograft experiments.

Acknowledgments

The authors are thankful to the editor and anonymous referees whose careful reading and constructive comments improved this article. This work was supported in part by National Cancer Institute (NCI) support grants CA21765 and NO1-CM-42216 and by the American Lebanese Syrian Associated Charities (ALSAC).

References

- 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]
- Demidenko, E. *Mixed Model: Theory and Applications*. Wiley; New York: 2004.
- Demidenko E. The assessment of tumor response to treatment. *Appl Statist* 2006;55:365–377.
- Efron, B.; Tibshirani, RJ. *An introduction to the bootstrap*. Chapman and Hall; New York: 1993.
- Hothorn L. Statistical analysis of in vivo anticancer experiments: tumor growth inhibition. *Drug Information Journal* 2006;40:229–238.
- 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]
- Liang H. Modeling antitumor activity in xenograft tumor treatment. *Biometrical Journal* 2005;47:1–11.
- Liang H. Comparison of antitumor activities in tumor xenograft treatment. *Contemporary clinical trials* 2007;28:115–119. [PubMed: 16814615]
- Liang H, Sha NJ. Modeling antitumor activity by using a nonlinear mixed-effects model. *Mathematical Biosciences* 2004;189:61–73. [PubMed: 15051414]
- 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;58:612620.
- Vardi Y, Ying ZL, Zhang CH. Two-sample tests for growth curves under dependent right censoring. *Biometrika* 2001;88:949960.

Figure 1. Solid lines are tumor volume profiles of control mice, and dotted lines are tumor volume profiles of treated mice.

Table 1
Power evaluation of three endpoints based on 10,000 simulations for two scenarios

Parameter		Empirical Power				Censoring		
σ	γ	RTVC	aAUC	CTD	MP	LTC		
Scenario I	0.3	0.00	0.051	0.052	0.051	0.006	5.338	
		0.10	0.305	0.321	0.335	0.018	4.531	
		0.15	0.532	0.599	0.622	0.031	3.995	
		0.20	0.717	0.833	0.852	0.049	3.512	
		0.25	0.843	0.952	0.962	0.073	3.140	
		0.30	0.924	0.990	0.994	0.102	2.850	
Scenario II	0.5	0.00	0.051	0.051	0.049	0.057	2.951	
		0.10	0.103	0.137	0.143	0.080	2.587	
		0.20	0.231	0.395	0.416	0.113	2.275	
		0.30	0.403	0.690	0.723	0.154	2.030	
		0.40	0.583	0.877	0.911	0.197	1.843	
		0.50	0.730	0.947	0.979	0.238	1.691	
	0.3	0.00	0.050	0.052	0.054	0.009	5.095	
		0.10	0.311	0.338	0.336	0.021	4.373	
		0.20	0.532	0.633	0.632	0.033	3.884	
		0.30	0.716	0.859	0.859	0.052	3.448	
		0.40	0.844	0.962	0.965	0.076	3.101	
		0.50	0.923	0.992	0.992	0.105	2.825	
	0.5	0.00	0.049	0.054	0.053	0.065	2.812	
		0.10	0.105	0.166	0.147	0.089	2.493	
		0.20	0.229	0.442	0.422	0.122	2.223	
		0.30	0.405	0.740	0.732	0.163	1.989	
		0.40	0.578	0.898	0.916	0.206	1.814	
		0.50	0.725	0.955	0.980	0.247	1.661	

dC/nC and dT/nT : number of deaths/mice initially in study for control group and treatment group, respectively.

T/C : the ratio of means of relative tumor volumes at the last time point with complete data of treatment to control.

aAUC: the ratio of means of adjusted area under the curve of treatment to control.

CTD: cross-treatment difference.

MP: average proportions of missing measurements.

LTC: average of last time point with complete data.

Table 2

Parameter configurations for the simulation study

Parameter	Design					
	Normal			Log-normal		
	I	II	III	I	II	III
(μ_C, μ_T)	(10, 8)	(10, 5)	(10, 2)	(1.0, 0.8)	(1.5, 0.8)	(2.0, 0.5)
γ	0.8	0.5	0.2	0.78	0.5	0.22
Gamma						
Parameter	I	II	III	IV	V	VI
$(a_C, \lambda, C, a_T, \lambda_T)$	(1,1,1,1,10)	(10,1,1,1,1)	(1,1,1,1,5)	(5,1,1,1,1)	(1,2,1,8)	(8,1,2,1)
γ	0.1	0.1	0.2	0.2	0.25	0.25
	VII	VIII	IX	X	XI	XII
$(a_C, \lambda, C, a_T, \lambda_T)$	(1,1,1,2)	(2,1,1,1)	(2,2,3,4)	(4,3,2,2)	(1,1,9,10)	(10,9,1,1)
γ	0.5	0.5	0.75	0.75	0.90	0.90

Table 3

95% Coverage probability based on 10,000 simulations for normal, log-normal and gamma data and 2,000 bootstrap samples used

(σ_G, σ_T)	Design	Normal		Log-Normal		Gamma	
		BC	t	BC	t	BC	t
(0.05, 0.05)	I	0.933	0.951	0.931	0.951	I	0.922 0.939
	II	0.929	0.950	0.931	0.951	II	0.872 0.927
	III	0.926	0.951	0.929	0.950	III	0.922 0.939
(0.1, 0.1)	I	0.932	0.950	0.930	0.951	IV	0.885 0.927
	II	0.927	0.950	0.930	0.950	V	0.922 0.939
	III	0.923	0.949	0.930	0.950	VI	0.905 0.940
(0.5, 0.5)	I	0.930	0.949	0.926	0.944	VII	0.921 0.939
	II	0.926	0.949	0.926	0.944	VIII	0.904 0.937
	III	0.921	0.948	0.926	0.944	IX	0.935 0.949
(0.1, 0.2)	I	0.925	0.948	0.916	0.943	X	0.916 0.942
	II	0.923	0.947	0.916	0.943	XI	0.954 0.959
	III	0.922	0.950	0.916	0.943	XII	0.875 0.929
(0.2, 0.1)	I	0.930	0.949	0.936	0.953		
	II	0.932	0.952	0.936	0.954		
	III	0.927	0.949	0.935	0.953		

BC: Bootstrap bias-corrected interval.

t : Bootstrap t -interval.

Table 4

Antitumor activity evaluation for PPTP vincristine xenograft study

Tumor line	Number of mice		T/C ratio		Adjusted AUC ratio	
	$(d_C/n_C, d_P/n_P)$	T/C	Activity	$\hat{\gamma}$ (s.e)	$\hat{\gamma}_{upper}^\dagger$	Activity
BT-29	(0/10,0/9)	0.64	Low	0.95 (0.071)	1.08	Inactive
KT-16	(0/10,0/10)	0.50	Low	0.94 (0.194)	1.33	Inactive
KT-14	(0/10,1/10)	0.89	Low	0.92 (0.091)	1.11	Inactive
KT-13	(0/10,0/10)	0.07	High	0.19 (0.038)	0.26	Active
SKNEP	(0/10,0/10)	0.25	Int	0.43 (0.084)	0.58	Active
EW5	(0/10,0/10)	0.79	Low	0.79 (0.112)	1.02	Inactive
Rh10 [*]	(0/10,0/10)	0.48	Low	0.61 (0.108)	0.80	Active
Rh28	(0/10,0/10)	0.14	High	0.11 (0.027)	0.18	Active
Rh30 [*]	(0/10,0/10)	0.51	Low	0.46 (0.057)	0.66	Active
Rh30R [*]	(0/10,0/10)	0.68	Low	0.64 (0.055)	0.74	Active
Rh41	(0/10,0/10)	0.34	Int	0.54 (0.070)	0.69	Active
Rh18	(0/10,0/10)	0.16	Int	0.51 (0.054)	0.63	Active
BT-50	(0/10,0/9)	0.37	Int	0.44 (0.097)	0.77	Active
SI-BT39	(0/10,0/10)	0.34	Int	0.33 (0.021)	0.37	Active
D645	(0/10,1/10)	0.43	Int	0.52 (0.085)	0.67	Active
D212 [‡]	(0/10,0/10)	0.36	Int	0.64 (0.187)	1.00	Inactive
NB-1771 [*]	(0/9,0/9)	0.61	Low	0.71 (0.082)	0.88	Active
NB-1691	(0/10,0/10)	0.80	Low	0.77 (0.105)	1.00	Inactive
NB-EBc1	(0/8,0/8)	0.77	Low	0.95 (0.121)	1.18	Inactive
CHLA-79	(0/9,0/9)	0.68	Low	0.90 (0.111)	1.10	Inactive
NB-1643	(0/8,0/8)	0.85	Low	0.94 (0.110)	1.27	Inactive
SK-N-AS	(0/10,0/10)	0.94	Low	0.95 (0.090)	1.12	Inactive

^{*} The T/C ratio shows low activity, but the adjusted AUC ratio leads to an active evaluation.

[‡] The T/C ratio shows intermediate activity, but the adjusted AUC ratio leads to an inactive evaluation.

[‡] 2,000 bootstrap samples were used to calculate the upper limits of boot-strap *t*-intervals.