

# Statistical Analysis of In Vivo Anticancer Experiments: Tumor Growth Inhibition

**Ludwig A. Hothorn, PhD**  
 Professor of Biostatistics,  
 Unit Biostatistics,  
 University of Hannover,

*Tumor growth inhibition data in in vivo anticancer experiments are commonly analyzed using the treatment-to-control ratio (TCR). Parametric and nonparametric confidence interval approaches for this ratio are introduced, enabling a quantitative statistical decision. The growth curves are characterized by the area-under-the-curve technique, adjusted for animal-specific survival. Simple simultaneous ap-*

*proaches are proposed for complex designs, including several treatment or dose groups. This implementation makes decision making easier for the pharmacologists through the use of simple diagrams for the treatment-to-control ratios and their confidence intervals. Tumor inhibition and regression can be appropriately statistically analyzed by treatment-to-control ratios and their confidence intervals.*

## Key Words

Treatment-to-control ratio;  
 Tumor inhibition;  
 Confidence intervals

## Correspondence Address

Ludwig A. Hothorn, PhD,  
 Unit Biostatistics,  
 University of Hannover,  
 Herrenhaeuserstr. 2,  
 D-30419, Hannover,  
 Germany  
 (email: hothorn@biostat  
 .uni-hannover.de).

## INTRODUCTION

The effect of antineoplastic compounds can be tested in experiments in vivo by analyzing the appropriateness of an inhibition of tumor growth relative to a control. In some experiments, the tumor is characterized by repeated measures of tumor volume, and the corresponding number of survivors is recorded over time. The treatment-to-control ratio (TCR) of tumor volume at a selected measurement time is commonly used for characterizing the effectiveness of therapy. For example, the tumor growth rate of human neuroblastoma xenograft was inhibited to a TCR of 0.3 after 16 days of treatment with the synthetic fumagillin analogue TNP-470 in mice (1).

Two specific statistical problems arise in testing: (a) accounting for missing values caused by mortality and (b) analyzing the ratio to control instead of the widespread used difference to control. Inherently connected with the second problem is the question of biological relevance, that is, which TCR is meaningful as opposed to which has formal statistical significance ( $P < 0.05$ ).

Previous statistical approaches (2,3) have been based on a two-stage model of carcinogenesis, which estimates the number of tumors over time and tests group differences. Moreover, a latent variable model for susceptibility and multi-

plicity of tumor counts that takes informative censoring caused by mortality into account has been described (4). Such a model considers the continuous end point for tumor volume and not the number of tumors. For a paired design, the tumor volumes of the left flank of mice (treatment) were compared with the right (control) by a multivariate sign test (5). An overview of the statistical approaches from a pharmacological standpoint is available (6).

Treatment differences were tested using the parameter estimates for both short- and long-term tumor growth based on a Gompertzian bi-exponential mixed-effects model (7). Recently, small-sample tests were proposed for incomplete data with truncation because of very small tumor volumes and independent censoring caused by mortality (8). However, these tests were based on rather restrictive assumptions (eg, a multivariate normal distribution with a covariance matrix in Toeplitz format), and they yielded only crude P values. A small sample test for contrasts of time points was used for tumor inhibition and regrowth experiments in which the parameters were estimated by the EM algorithm assuming three forms of repeated measures: complete, informative missing, and missing at random (9). This idea was recently extended to a linear model approach for the analysis of different cycles of drug administration (10).

## SUMMARY MEASURE OF INCOMPLETE GROWTH CURVES

Depending on the design and duration of an inhibition study, some animals die spontaneously or have to be sacrificed in moribund status because of an unacceptably large tumor size. Not uncommonly, this causes a missing value problem. An increasing probability of missings over time and a possible higher mortality rate in the treated group(s) can occur. Ignoring these so-called informative missings may cause serious bias when using standard statistical tests. One possible analysis is the selection of the last time point with complete data and evaluating that point using standard tests. However, increasing values for effect ratios between treatment and control are expected with longer durations. Depending on the data, this approach may have an unacceptably high false-negative rate. Replacement of the missing tumor volumes with extrapolated values (eg, using predicted individual growth curves or group means) may work in some examples, but such replacement uses non-experimental data to characterize the antineoplastic activity.

An alternative approach, standard in clinical trials with repeated measures data (10,11), is the transformation of the individual growth curve into an area-under-the-curve (AUC) form using a simple trapezoidal rule. The AUC approach represents an integration transformation of the repeatedly measured tumor volumes into a univariate end point that can be evaluated with simple standard tests; that is, the AUC is a weighted mean of the repeated measures. The adjusted AUC can be calculated:  $AUC_{adjusted} = AUC / (t_i - t_1)$  with  $t_i$  from  $t_2$  to  $t_{final}$  ( $t_i$  represents the latest time with existing tumor volume), analogously to standardized AUCs in a clinical trial with informative missing values (12). No arguments for the assumption of a normal distribution for this transformed variable exist, therefore, a nonparametric approach can be recommended.

On the other hand, in several real studies, approximative normal distribution of the adjusted AUC was found. If baseline values are available, a

difference-to-baseline or ratio-to-baseline transformation can be performed before the AUC calculation to adjust for baseline heterogeneity. For details on choosing between difference- and ratio-to-baseline, see, for instance, (13).

Alternatively, a modification of the nonparametric Wilcoxon test can be used in which the Mann-Whitney counts for the remaining pairs of data determine the test statistics (14). However, only the estimation of a  $p$  value is possible using this approach; no confidence intervals are available.

A further alternative is the transformation of the data into time required to reach a predetermined tumor volume and use of a standard censored test (15), but the analysis of a derived variable depends on the threshold definition and does not allow direct interpretation of tumor inhibition. The evaluation of the final tumor volumes by the all-pairs multiple comparison Newman-Keuls procedure recently used in the analysis of a study for tumor inhibition of triplex-forming oligonucleotides can be seriously biased (16).

## CONFIDENCE INTERVALS FOR TREATMENT-TO-CONTROL RATIO

Confidence intervals are often preferred to  $P$  values because they are scale variant and can be interpreted directly in the medical measurement's units of measure. Moreover, they simultaneously offer information about the distance from the null hypothesis, the direction of the effect, and the variability. Confidence intervals for the difference are frequently used in biomedical research (eg, in the comparison of allogeneic transplant with chemotherapy for relapsed childhood acute lymphoblastic leukemia) (17). In answering some medical questions, it is more appropriate to use confidence intervals for the ratio to control or ratio to placebo (18); that is, percentage change versus control can be directly medically interpreted (19).

Sometimes, the ratio problem is transformed into a difference problem by taking the log of the end points, a technique well known in bioequivalence testing. This simple approach inherently assumes log-normal distributed end points (20).

Therefore, both a parametric approach, assuming normal distribution, and a nonparametric approach, assuming any continuous distribution, was worked out. A confidence interval for a ratio  $TCR = \bar{V}_{Treatment} / \bar{V}_{Control}$  is available for the parametric two-sample case (21). Because medical interest lies only in tumor volumes that decrease after treatment, the one-sided upper confidence interval  $(-\infty, \theta_{upper})$  is:

$$\theta_{upper} = \frac{\bar{V}_{Control}\bar{V}_{Treatment} + \sqrt{a\bar{V}_{Treatment}^2 + bV_{Control}^2 - ab}}{\bar{V}_{Control} - a}$$

$$a = \frac{MSE}{n_{Control}} t_{df=n_{Control}+n_{Treatment}-2, 1-\alpha}^2 \text{ and } b = \frac{MSE}{n_{Treatment}} t_{df, 1-\alpha}^2$$

where  $\hat{V}$  denotes the group mean of the tumor volumes,  $MSE$  is the common mean square error estimator,  $df$  is the degrees of freedom, and  $t_{df, 1-\alpha}$  is the quantile of the  $t$  distribution. The side condition  $\hat{V}_{Control} > a$  is simply a one-sided test for control mean values larger than zero; that is this approach is limited to nonzero control effects. Unlike some clinical studies in tumor inhibition experiments, this restriction is not problematic because the number of tumors in the control are expected to be higher than those in the treatment group.

Using this confidence interval approach, the treatment effect is characterized by the TCR estimator and its 95% upper confidence interval. When the high dose is compared with the control in the example data (see Table 1),  $TCR = 0.80$ , and the one-sided upper 95% confidence limit is 0.97. This can be medically interpreted as follows: ‘the treatment reveals a mean tumor volume reduction of 80% with at least a 97% reduction and a maximal false positive rate of 95%.

The interpretation is simple. Whenever the upper confidence interval is lower than 1, a significant decrease in tumor volume exists; otherwise, no conclusion can be drawn. However, how relevant is a treatment with a TCR of only 80% and an upper confidence interval of only 97%? Formal decision making cutoff points for the TCR (not for its upper confidence interval) have been suggested (22):  $\leq 42\%$  for relevant antitumor activity and  $\leq 10\%$  for highly relevant activity.

Let us assume two experiments have the same TCR of 22% but different upper confidence intervals of 38% and 79%. The relevance of the second drug study is much vaguer (although still formally significant) compared with the first. Therefore, a cutpoint for the upper confidence limit (eg,  $\theta = 50\%$ ) should be defined to include the effect size *and* variability. The a priori definition of such a cut-off point is not needed for the confidence interval approach. The upper limit  $\theta_{upper}$  itself is an outcome of the experiment and can be interpreted by pharmacologists according to relevance. Sample size influences the variability; hence, an a priori selected sample size or an international guideline with an ‘at least xx animals/group recommendation’ should be used.

Generally, confirmatory decisions in statistical tests are arbitrary when arbitrary sample sizes are used. For the parametric confidence interval for the ratio to control, a related algorithm for sample size calculation is available (23). The following formula approximates a balanced sample size for a given false-negative rate  $\beta$ , a false-positive rate  $\alpha$ , the coefficient of variation  $CV_{Control} = \sqrt{\text{Variance}_{Control}} / \bar{V}_{Control}$  and a value from the alternative  $\theta^A = (\mu_{Treatment} / \mu_{Control})$  is:

$$n \geq (1 + \theta^2) \left( \frac{CV_{Control}}{\theta - \theta^A} \right)^2 (t_{\alpha, 2n-2} + t_{\beta, 2n-2})^2.$$

From the example data in Table 1, a  $CV_{control}$  of about 35% can be concluded. Assuming a maximum false-positive rate of 5%, a maximum false-negative rate of 30%,  $\theta = 0.75$ , and  $\theta^A = 0.5$  results in a necessary sample size of  $n_i = 16$ . For an explicit discussion of the dependency of power on variance, sample size, and growth delay for the simple tumor-doubling time model, see ref. 24.

It should be noted that variance heterogeneity, particularly larger variances in groups with fewer samples, violates the assumption of the estimation. Therefore, equal sample sizes are preferred in such experiments. Parametric intervals require the validity of the normal distribution assumption. Yet, sometimes skewed data or out-

TABLE 1

Tumor Volumes of a Study With Two Doses and a Control for Baseline From Weeks 17 to 31						
Group	T0	T17	T21	T25	T28	T31
0	297	690	1368	2530	3956	
0	105	143	211	174	234	355
0	187	635	1174	2050	2966	
0	128	529	865	913	1200	1545
0	234	790	1313	2000	3187	
0	100	438	770	1094	1679	1910
0	72	302	448	957	1223	1310
0	208	532	1091			
0	333	1229	1787	2020	1598	
0	173	640	777	840	987	1182
0	113	322	1005	1346	1869	2202
0	81	211	443	633	1100	1370
0	249	833	1234	1164	1806	1893
0	138	386	952	1296	1641	2470
0	162	611	1100	1874	2364	
0	133	335	665	1185	1602	2341
0	217	519	909	1105	2260	2771
0	131	398	667	1020	1024	1809
0	64	198	315	497	778	897
0	346	792	1076	1850	2418	2941
0	68	261	508	652	1285	1754
1	66	171	298	705	889	1123
1	143	414	558	726	1033	1302
1	180	573	923	835	1229	1663
1	133	416	698	1034	1441	1776
1	161	327	804	1078	1242	1754
1	75	108	116	106	129	115
1	190	478	772	1281	1590	
1	175	632	1285	537		
1	121	258	474	733	896	1381
1	97	299	453	627	1218	1698
1	94	213	592	834		
1	87	407	550	876	1146	
1	254	764	1437	1636	1950	
1	292	1136	1584	1994	1723	2190
1	220	451	934	1580	2302	2818
1	190	455	1189	1589	2255	2365

TABLE 1

<i>Continued</i>						
Group	T0	T17	T21	T25	T28	T31
1	158	467	964	999	1530	1654
1	54	120	357	870	1726	1013
1	188	892	1291	1464	2136	
1	166	339	395	446	796	772
1	78	389	735	1086	1547	1200
2	76	156	361	568	602	993
2	66	171	298	705	889	1123
2	143	414	558	726	933	1302
2	180	573	923	835	1229	1663
2	133	416	698	1034	1441	1776
2	161	327	804	1078	1242	1754
2	75	108	116	106	129	
2	190	478	772	850	951	
2	175	632	577	537		
2	121	258	474	733	896	1381
2	97	299	453	627	1018	1698
2	94	213	592	734		
2	87	407	550	876	1046	
2	254	764	1437	1636	1850	
2	292	1136	1584	1664	1723	2190
2	220	451	934	1580	2302	2818
2	190	455	1189	1188	1055	2365
2	158	467	964	999	1430	1654
2	54	120	357	870	1026	1013
2	188	892	1291	1464	1734	
2	166	339	395	446	796	772
2	78	389	735	1086	1547	1200

liers can be observed. Moreover, heavily tied data occur (eg, zero number of tumors or odd numbers of metastases). Therefore, an alternative nonparametric confidence interval approach in analogy to *t* test/Wilcoxon's U test is proposed (25). For the parallel group design, the appropriateness of the log transformation of the ratios, the use of asymptotic nonparametric

confidence intervals (26), and the back-transformation of the ratios was demonstrated.

### SIMULTANEOUS COMPARISON OF SEVERAL TREATMENT OR DOSE GROUPS VERSUS A CONTROL

A design with a control and *k* treatment or dose groups is frequently used. Here, the multiplicity

problem occurs; that is, the experimentwise false-positive rate should be controlled at the predefined  $\alpha$  level of 5%. For this many-to-one comparison problem, multiplicity-adjusted confidence intervals are recommended. Recently, the use of the simple Bonferroni adjustment was proposed (27), that is, the use of the quantile  $t_{df,1-\alpha/k}$ , as long as the coefficient of variation of the mean of the control

$CV_{\text{Control}}^{\text{mean}} = \frac{\sqrt{\text{Variance}_{\text{Control}}}}{\sqrt{n_{\text{Control}} \bar{V}_{\text{Control}}}}$  is small. Since the volumes of the control are large in tumor inhibition experiments, this criterion will frequently be fulfilled.

In designs with dose groups, the question arises of whether a global trend exists at all, and if so, what is the minimal effective dose? The dose-response analysis can be performed stepwise using the unadjusted intervals (28). Start with the comparison of the high dose with the control. If this upper confidence interval is smaller than 1, conclude that a trend exists up to  $D_{\text{high}}$  and continue with the comparison versus  $D_{\text{high}-1}$ ; otherwise, stop (see the example section). The lowest significant dose in the procedure is the minimal effective dose

### SPECIFIC COMPARISON OF THE MORTALITY-TIME RELATIONSHIPS

In addition to the analysis of tumor volume, the comparison of the mortality-time relationships can be used for demonstrating antitumor activity. In an inhibition study, the primary end point is the tumor volume, whereas in a survival study, the primary end point is time to sacrifice. Although both study types are practiced, they are rarely combined because of ethical limitations. There is an extensive body of literature on survival analysis. Some animal anticancer experiments only count the individual days until mortality and the survivors' day of terminal sacrifice without noting cause-of-death information (ie, spontaneous death is assumed to be tumor-related without consideration of any competing risks). The standard non-parametric log rank test reduces to the Savage scores test, and Gehan's test reduces to the U test if the censor-

ing point is equal over the groups. In such inhibition studies a simple, Wilcoxon U test can be used for comparing the survival times. Exact permutative versions are recommended because the sample sizes are small to moderate.

### AN EXAMPLE

A new potential antineoplastic compound was investigated in a study on a special mice strain inoculated with certain human tumor cells. After an initial growth phase for establishing measurable tumor volumes, 64 animals were randomized to three treatment groups: a control (0) and two dose groups (1,2). The baseline tumor volumes were measured. Tumor volumes were observed on days 17, 21, 25, 28, and 31; the individual raw tumor volume data are given in Table 1. The box plots of the tumor volumes in Figure 1 indicate variance heterogeneity over time, symmetric or skewed distribution, and outliers; the box plots of the baseline tumor volumes reveal some heterogeneity between the randomized groups.

The analysis begins as a simple control versus high dose two-group comparison. The TCR for the final time point (T31) alone is 0.875 with an upper confidence interval of 1.12, indicating no significant change in the high-dose group with respect to the control. However, six values are missing in the control and seven in the high-dose group because of mortality, so analysis of the final time point is biased. Depending on the mortality pattern, this bias can be positive or negative. Using the adjusted AUC approach until the final time point TCR of 0.756 results in an upper confidence interval of 0.984, indicating a significant change. One-sided confidence intervals are appropriate because only decreases in tumor volume are of interest. The corresponding one-sided upper parametric interval for the adjusted AUC approach is 0.943. Due to the baseline heterogeneity, differences to baseline were used for analysis. The box plots indicate skewed distributions with outliers; hence, the nonparametric approach seems to be appropriate for these data. The associated TCR is 0.743 with an upper confidence interval of 0.899. The more sensitive estimates are not surprising, but

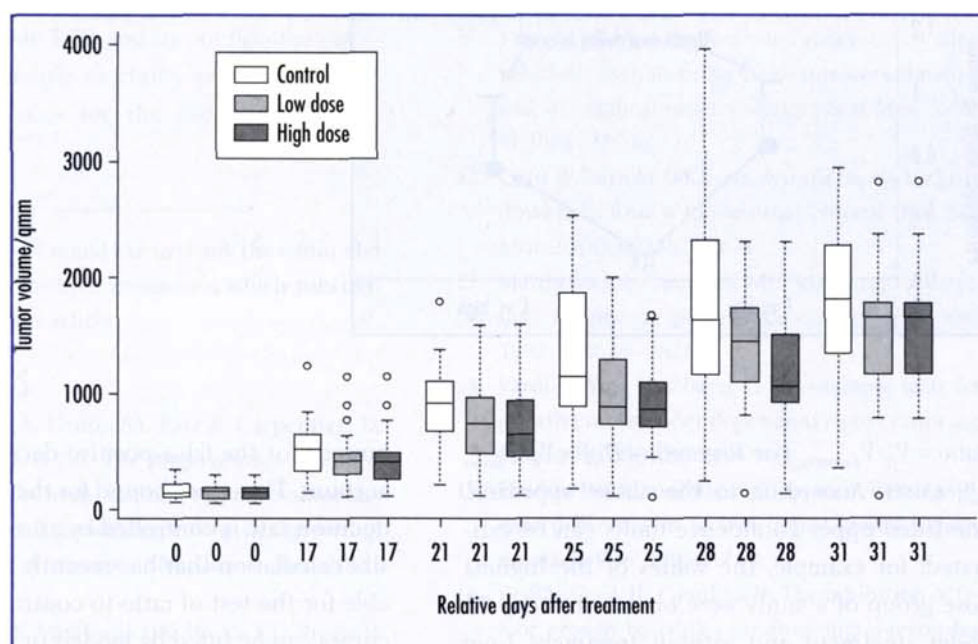


FIGURE 1

Box-plots for tumor volumes (baseline and repeated measures).

like the  $t$  test versus  $U$  test selection problem, care is needed in choosing between the parametric and nonparametric approaches. The high dose reveals a significant reduction in tumor volume according to the tumor growth curves in comparison with the control of about 74% with a false-positive rate of 90% to 95%. Using the above 75% threshold, this effect is significant but not biologically relevant.

Now, the entire data set will be analyzed. It can be assumed a priori that if a treatment effect in the low-dose group occurs, such an effect must also hold for all higher doses. Therefore, the upper  $(1 - \alpha)$  confidence intervals for the high-dose and the low-dose groups versus the control can be estimated. Figure 2 presents both the TCR and the upper confidence intervals.

The decision procedure works as follows. Start with the upper confidence interval of the highest dose versus the control. If it is smaller than 1, the highest dose is effective, so continue with the next comparison. Otherwise, stop with the conclusion that neither dose is significant. If the upper confidence interval of the low dose is smaller than 1, this low dose is not significant. Therefore, stop with the conclusion that only the high dose is significant and therefore is the

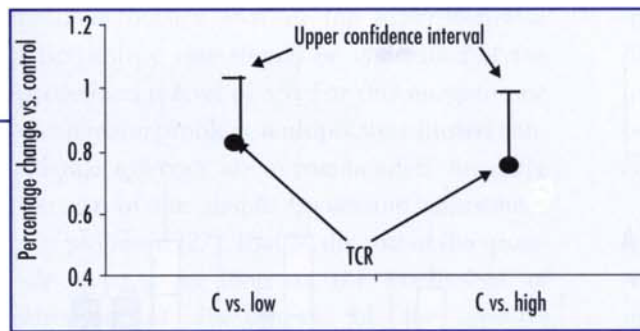
minimum significant (effective) dose. This procedure can also be performed with the relevance criteria  $\theta$ . Using  $\theta = 75\%$  the procedure stops in the first comparison with the conclusion that neither dose is effective. Simultaneous confidence intervals for  $\theta$  are available (28).

### ANALYSIS OF TUMOR RE-GROWTH

A more relevant objective is the proof of tumor regression instead of only inhibition. For the treatment groups  $i$  of an inhibition study, a tumor regression factor  $TRF_i = (V_{T_0j} - V_{T_{final_i}}) / V_{T_0}$  can be defined. Parametric and nonparametric confidence intervals for this one-sample variable are simple to estimate. If the lower bound is positive, then significant tumor regression occurs in this treatment group. Another possibility is to demonstrate that no or only moderate tumor regrowth took place after substance administration of a successful tumor inhibitor. The baseline value (ie, the day of the last tumor inhibition measurement under treatment) will be compared with repeated measures of a period after substance administration. The comparison for each time point  $j$  versus baseline within the treatment groups  $i$  seems to be appropriate; that is, a paired comparison problem for *Regrowth* –

FIGURE 2

TCR and upper confidence intervals for the dose-response analysis.



Ratio =  $V_{ji}/V_{Baseline,i}$  or  $Regrowth - Diff = V_{ji}/V_{Baseline,i}$  exists. According to the above approach, one-sided upper confidence limits can be estimated; for example, the values of the highest dose group of a study were observed at day 31 under treatment and without treatment from day 35 until day 63. Since some animals had immeasurable tumor volumes (ie,  $V_{Baseline} = 0$ ), the regrowth differences and their parametric lower 95% confidence limits were calculated for a one-sided, one-sample  $t$  test and are given in Table 2.

If an a priori selected regrowth difference of up to  $100 \text{ mm}^3$  is tolerated as equivalent, no meaningful regrowth took place until day 49 because its lower confidence limit is still less than the threshold.

**CONCLUSIONS**

Biostatistical design and evaluation of tumor growth inhibition inherently accompanies in vivo anticancer experiments. The TCR of the tumor volume is a commonly used measure for effect size. Inhibition percentages are easily interpretable. The confidence interval for this ratio to control enables quantitative decision making that takes effect size, variability, and upper

bounds for the false-positive decision rate into account. The upper bound for the false negative decision rate is controlled by an a priori sample size calculation that has recently become available for the test of ratio to control. The growth curves can be fitted by models or more robustly characterized by the AUC approach, adjusted for animal-specific survival. Modifications for simultaneous comparisons of several treatment groups versus control and identification of the minimal effective dose in dose-response experiments are available. Tumor regression experiments can be analyzed analogously; however, only raw data or ratio-to-baseline transformation guarantee the necessary side condition of a nonzero control effect.

The final result is a straightforward diagram for the TCRs and their confidence intervals that can be easily interpreted, particularly with respect to relevance.

The proposed approach is simple and appropriate for pharmacological interpretation. It is directed toward experiments in tumor growth inhibition/regression with some mortality but not for a combination of tumor growth and survival. In contrast to the existing published approaches, the proposed approach focuses on

TABLE 2

Analysis of Regrowth Data (in $\text{mm}^3$ Tumor Volume)									
Day without treatment	35	38	42	45	49	52	56	59	63
Difference to baseline	4.1	4.9	64.4	108.8	165.1	249.6	360.8	469.5	693.3
Lower confidence interval-limit	-0.74	4.9	33.3	60.0	97.0	147.0	213.3	284.9	419.1



the interpretable TCR and its confidence interval based on simple mortality-adjusted integral summary statistics for the individual growth curves.

*Acknowledgments*—I would like to thank the editor and the two referees for their comments, which substantially improved this article.

## REFERENCES

- Shusterman S, Grupp SA, Barr R, Carpentieri D, Zhao H, Maris JM. The angiogenesis inhibitor TNP-470 effectively inhibits human neuroblastoma xenograft growth, especially in the setting of subclinical disease. *Clin Cancer Res*. 2001; 7:977–984.
- Freedman LS, Midthune DN, Brown CC, Steele V, Kelloff GJ. Statistical analysis of animal cancer chemoprevention experiments. *Biometrics*. 1993; 49:259–268.
- Kokoska SM, Hardin JM, Grubbs CJ, Hsu C. The statistical analysis of cancer inhibition/promotion experiments. *Anticancer Res*. 1993;13: 1357–1363.
- Dunson DB, Haseman JK. Modeling tumor onset and multiplicity using transition models with latent variables. *Biometrics*. 1999;55:965–970.
- Koziol JA, Maxwell DA. A distribution-free test for paired growth curve analyses with application to an animal tumour immunotherapy experiment. *Stat. Med*. 1992;1: 83–89.
- Hanfelt JJ. Statistical approaches to experimental design and data analysis of in vivo studies. *Breast Cancer Res Treat*. 1997; 46: 279–302.
- Liang H, Sha NJ. Modeling antitumor activity by using a non-linear mixed-effects model. *Math Biosci*. 2004;89:61–73.
- 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.
- Tan T, Fang HB, Tian GL. Repeated-measures models with constrained parameters for incomplete data in tumour xenograft experiments. *Stat. Med*. 2005;24:109–119.
- Pham B, Cranney A, Boers M, Verhoeven AC, Wells G, Tugwell P. Validity of area-under-the-curve analysis to summarize effect in rheumatoid arthritis clinical trials. *J Rheumatol*. 1999; 26:712–716.
- Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*. 1992; 11:1685–1704.
- Qian W, Parmar MKB. Analysis of messy longitudinal data from a randomized clinical trial. *Stat Med*. 2000;19:2657–2674.
- Matthews JN, Campbell MJ. Adjusting for baseline: change or percentage change. *Stat Med*. 1992;11:1624–1626.
- Vardi Y, Ying Z, Zhang C. Two-sample tests for growth curves under dependent right censoring. *Biometrika*. 2001;80:949–960.
- Heitjan DF, Kunselman S. Time until first significant difference in in vivo tumor growth experiments. *In Vivo*. 1995;9:7–9.
- Re RN, Cook JL, Giardina JF. The inhibition of tumor growth by triplex-forming oligonucleotides. *Cancer Lett*. 2004;209:51–53.
- Harrison G, Richards S, Lawson S, et al. Comparison of allogeneic transplant versus chemotherapy for relapsed childhood acute lymphoblastic leukaemia in the MRC UKALL R1 trial. *Ann Oncol*. 2000;11:999–1006.
- Roehmel J. Therapeutic equivalence investigations: statistical considerations. *Stat Med*. 1998; 17:1703–1714.
- Feuerstein TJ, Rossner R, Schumacher M. How to express an effect mean as percentage of a control mean? *J Pharmacol Toxicol Methods*. 1997;37: 187–190.
- Liu JP, Weng CS. Evaluation of log-transformation in assessing bioequivalence. *Commun Stat*. 1994; A23:421–434.
- Fieller E. Some problems in interval estimation. *J R Stat Soc*. 1954;B16:175–185.
- Corbett Th, Valeriote F. Tumor models and the discovery and secondary evaluation of solid tumor active agents. *Int J Pharmacognosy*. 1995;33: 102–122.
- Hauschke D, Kieser M, Hothorn LA. Proof of safety in toxicology based on the ratio of two means for normally distributed data. *Biometr J*. 1999; 41:295–304.
- Iversen PW. Comparison of in vivo tumor growth models. *Proc Annu Meeting Am Stat Assoc*. 2001:5–9.
- Hothorn T, Munzel U. Non-parametric confidence interval for the ratio. Report University of Erlangen, Department Medical Statistics. 2002. Available at: [www.uni-erlangen.de](http://www.uni-erlangen.de).

26. Hodges JL, Lehmann EL. Estimates of location based on rank tests. *Ann Math Stat.* 1963; 34:598–611.
  27. Dilba G, Bretz F, Guiard V, Hothorn LA. Simultaneous confidence intervals for ratios for the comparison of several treatments with a control. *Meth Inf Med.* 2004;43:465–469.
  28. Bretz F, Hothorn LA, Hsu J. Identifying effective and/or safe doses by stepwise confidence intervals for ratios. *Stat Med.* 2003;22:847–858.
-