Published in final edited form as:

Stat Med. 2010 November 20; 29(26): 2669–2678. doi:10.1002/sim.4038.

INTERVAL ESTIMATION OF QUANTILE RATIOS APPLIED TO ANTI-CANCER DRUG SCREENING BY XENOGRAFT EXPERIMENTS

Cheng Cheng^{1,*} and Jianrong Wu¹

¹Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-2794, U.S.A

Abstract

The current practice in analyzing data from anti-cancer drug screening by xenograft experiments lacks statistical consideration to account for experimental noise, and a sound inference procedure is necessary. A novel confidence bound and interval procedure for estimating quantile ratios developed in this paper fills the void. Justified by rigorous large-sample theory and a simulation study of small-sample performance, the proposed method performs well in a wide range of scenarios involving right-skewed distributions. By providing rigorous inference and much more interpretable statistics that account for experimental noise, the proposed method improves the current practice of analyzing drug activity data in xenograft experiments. The proposed method is fully nonparametric, simple to compute, performs equally well or better than known nonparametric methods, and is applicable to any statistical inference of a “fold change” that can be formulated as a quantile ratio.

Key words and phrases

Quantile; Median; Quartile; Ratio; Fold change; Confidence interval; Confidence bound; Xenograft experiment

1. Introduction

In cancer drug development, demonstrating anti-cancer activity in preclinical animal models is important. The National Cancer Institute (NCI) has promoted a cancer drug screening program since the mid-1950s. The Pediatric Preclinical Testing Program (PTTP) recently conducted a cancer drug screening program for both *in vitro* and *in vivo* models [1]. In *in vivo* testing, 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 the drug is administered. 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, treatment-to-control ratio (T/C) at a given time point when complete observations are available on all mice is commonly used to quantify the antitumor activity of the drug, where the T and C represent the mean (or median) of relative tumor volumes (tumor volumes divided by their initial volumes) of treatment and control, respectively. The T/C ratio is converted to an antitumor activity rating by an arbitrary threshold. For example, a significant treatment effect is claimed if $T/C \geq 0.42$

*cheng.cheng@stjude.org.

by NCI definition and $T/C = 0.45$ by PPTP definition, without any statistical inference [2] [1]. Clearly the problem with this approach is that it ignores the noise in experimental data. A sound statistical procedure is thus necessary for analyzing this type of experiment.

A formal statistical approach is to estimate the standard error and obtain confidence upper bound or interval of the T/C ratio to account for experimental noise. The treatment effect can be assessed by confidence limits. An interval estimate for the T/C ratio of two means on the basis of assumed Normal distribution of the tumor volume was proposed [3]. The normality assumption is unreliable due to skewness of the tumor volume distribution. There have been several parametric methods for interval estimation of the ratio of two means from two random samples [4][5][6]. A generalized variable approach is given in [7]. Recently a regression model for quantile functions (not Koenker's regression quantiles [8]) was introduced for estimation of differences between two means from two sample data [9] [10]. Estimating the ratio of two means from bivariate (normally distributed) data was perhaps first considered by Fieller [11], and bootstrap and bias corrected bootstrap methods have been developed for this problem [12](Sections 10.3, 25.2–25.3). However, in vivo tumor volume data from xenograft experiments frequently exhibit a wide range of distributional characteristics sometimes with outliers. T/C ratio of means is not robust against to the outliers and the likelihood based confidence interval of the T/C ratio is not robust against the distribution assumption. Figure 1 shows such a data set; clearly no single parametric family can capture the diverse distributional behavior in the control and treatment groups. This figure will be further discussed in Section 4.

Therefore, a nonparametric, quantile-based interval approach is more preferred. Henceforth, by “quantile (or median, quartile, etc.) ratio” we mean the ratio of two quantiles from two groups. In this paper, we propose to use median ratio for quantification of the treatment effect, along with an interval estimator of the median ratio to assess significance of the treatment effect in tumor xenograft experiments. The proposed method is fully nonparametric. Furthermore, quantile ratios can have broad applications. In general, a biological experiment often makes observations of two positive random variables with right-skewed distributions. The investigator sometimes prefer the ratio (often referred as “fold change” in biomedical literature) than the difference of some location parameter (mean or median) of the two variables. The fold change indeed provides a meaningful comparison of two right-skewed, positive and continuous random variables. The nonparametric quantile ratio interval estimators developed herein are applicable to any inference of a “fold change” that can be formulated as a quantile ratio.

Theory and methods of quantile estimation have been studied thoroughly with a large body of literature. For quantile estimation based on uncensored data, significant works include asymptotic representation [13], asymptotic expansions [14], strong approximation of the quantile process [15], data modeling [16], estimators derived from subsampling arguments [17][18][19], estimators derived from kernel smoothing [20][21][22][23], and a more general class of estimators that subsumes all previous estimators [24]. Kaigh and Cheng [25] review and investigate approaches to estimating variances of quantile estimators for quantile interval estimation and comparison of two population medians.

Motivated by the fact that proper smoothing provides reduction in the asymptotic variance of the quantile estimator, we propose in this paper a direct and conservative estimator of this asymptotic variance, and then use it to construct confidence intervals or bounds for quantile ratios. There are two other completely nonparametric approaches to two-sample quantile interval estimation. The closed-form bootstrap variance estimator [25] was derived from a subsampling argument. The other is Koenker's regression quantiles [8]. The difference between two quantiles can be formulated as the regression coefficient corresponding to a

particular regression design matrix, and be estimated via L_p -norm ($0 < p < 1$) fitting. A confidence interval can be obtained by inverting a rank test of the regression coefficient [26]. Small-sample performance of these methods in interval estimation of quantile ratios are compared by a simulation study. The results demonstrate an interesting asymmetry in performance in contrast of estimating a quantile difference: all methods work reasonably well when the true quantile ratio is small (< 1), and loss of precision occurs when the true quantile ratio gets large, especially when a long-tail distribution is involved. Thus the performance depends on which population quantile is put in the numerator (respectively denominator). Note however for interval estimation of a quantile difference, the performance is invariant of which population quantile is before (respectively after) the subtraction.

The proposed method is applied to the analysis of xenograft experiments. By providing rigorous inference and much more interpretable statistics that account for experimental noise, this method substantially improves the current practice of analyzing such experiments.

This paper is organized as follows. The proposed and other nonparametric methods are described in Section 2. The simulation study and results are given in Section 3. Applications to anti-cancer drug screening with xenograft models are presented in Section 4. Concluding remarks are made in Section 5.

2. The Estimation Methods

2.1 Quantile estimators and confidence intervals

We begin by a brief review of point and interval estimation of a single quantile. Let $X_1, \dots, X_n$ be an independent identically distributed (i.i.d.) sample from a distribution on the real line, with cumulative distribution function (cdf) F , probability density function (pdf) f , quantile function $Q(u) = F^{-1}(u) = \inf\{t: F(t) \geq u\}$, and quantile density function (qdf) defined as $q(u) = \frac{d}{du}Q(u) = \frac{1}{f(Q(u))}$, $u \in (0, 1)$. Essentially all estimators of the u th population quantile $Q(u)$ are in forms of some linear combination of the order statistics $X_{1:n}, X_{2:n}, \dots, X_{n:n}$. For example, the traditional sample quantile defined by inverting the empirical cdf, $\tilde{Q}(u) = X_{[\lceil un \rceil]:n}$ ($\lceil a \rceil$ is the smallest integer greater than or equal to a), gives weight 1 to the $\lceil un \rceil$ th order statistic. In general an estimator of $Q(u)$ can be written as

$\hat{Q}(u) = \sum_{j=1}^n w(u, j) X_{j:n}$ with some choice of the weight function $w(u, \cdot)$. A reasonable

restriction on the weights is that $w(u, j) \geq 0$ for all j and $\sum_{j=1}^n w(u, j) = 1$. Such a weighted average procedure is often referred as “smoothing.” A class of weights can be derived from subsampling arguments [19]. When $w(u, \cdot)$ is so chosen that it is differentiable in u , a natural quantile density estimator is

$$\hat{q}(u) = \frac{d}{du} \hat{Q}(u) = \sum_{j=1}^n \left[\frac{d}{du} w(u, j) \right] X_{j:n}.$$

A very broad class of quantile and quantile density function estimators using a generalized kernel representation has been studied [24][27].

Henceforth we focus on a particular choice, the Bernstein polynomial estimator, for its nice asymptotic and finite-sample properties [28]: Estimate $Q(u)$ by

$$\widehat{Q}(u) = \sum_{j=1}^n \left[\binom{n-1}{j-1} u^{j-1} (1-u)^{n-j} \right] X_{j:n}$$

and estimate $q(u)$ by

$$\widehat{q}(u) = \frac{d}{du} \widehat{Q}(u) = \sum_{j=1}^{n-1} \left[\binom{n-2}{j-1} u^{j-1} (1-u)^{n-1-j} \right] \frac{X_{j+1:n} - X_{j:n}}{1/(n-1)}.$$

An advantage over the sample quantile $\tilde{Q}(u)$ that proper smoothing can provide to $\widehat{Q}(u)$ is an appreciable reduction in asymptotic variance (as $n \rightarrow \infty$) which translates into a gain in the Hodges-Lehmann efficiency [24]. The asymptotic variance of the sample quantile is σ^2/n with $\sigma^2 = [q(u)]^2 u(1-u)$ ([29] P.77). It can be shown (see Appendix) that under reasonable smoothness conditions of the quantile function and the weight function, $\widehat{Q}(u)$ is

asymptotically normal with mean $Q(u)$ and an asymptotic variance σ_n^2/n ; that is, $\frac{\widehat{Q}(u) - Q(u)}{\sigma_n/n^{1/2}}$ converges in distribution to $N(0, 1)$. Furthermore, the asymptotic variance $\sigma_n^2 = \sigma^2 - \xi_n + b_n$ with $\xi_n > 0$ for all n . In the Bernstein polynomial case, $\xi_n = O(n^{-1/2})$ and $b_n = O(n^{-1})$; thus reduction in asymptotic variance is expected. These facts thus suggest the following procedure to construct a confidence interval (CI) for $Q(u)$: First estimate σ^2 by plugging in the qdf estimator $\widehat{q}(u)$ to give $\widehat{\sigma}^2 = [\widehat{q}(u)]^2 u(1-u)$; next an asymptotically $100(1-\alpha)\%$ CI of $Q(u)$ is given by

$$\widehat{Q}(u) \pm (\widehat{\sigma}/n^{1/2}) \Phi^{-1}(1-\alpha/2) = \widehat{Q}(u) \pm \widehat{q}(u) [u(1-u)/n]^{1/2} \Phi^{-1}(1-\alpha/2),$$

where $\Phi^{-1}(\cdot)$ is the $N(0, 1)$ quantile function.

Kaigh and Cheng [25] review and investigate subsampling based estimators of the variance of a quantile estimator which can be used to construct confidence intervals (or one-sided bounds) for quantiles and quantile differences. Let

$$b_r(j) = \int_{(j-1)/n}^{j/n} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt.$$

Following the subsampling argument [19], the bootstrap distribution of the r th sample order statistic is $\Pr(X_{r:n} = X_{j:n}) = b_r(j)$, $j = 1, \dots, n$. Then if one estimates $Q(u)$ by the $r(u) = \lceil un \rceil$ th sample quantile, a bootstrap estimator of its variance can be computed in closed form as [25]

$$\widehat{\sigma}_{Boot}^2 = n^{-1} \sum_{j=1}^n b_{r(u)}(j) (X_{r(u):n} - X_{j:n})^2.$$

This procedure is henceforth referred as closed-form bootstrap (CFBoot) method. Rather than the sample order statistics, one can alternatively bootstrap smoothed order statistics

[25]. A specific choice is as follows. First transform the order statistics $X_{1:n}, \dots, X_{n:n}$ into the smoothed order statistics (see [19] for general form)

$$\widehat{Q}_{j:n} = \sum_{t=1}^n \left[\frac{\binom{j+t-2}{j-1} \binom{2n-t-j}{n-j}}{\binom{2n-1}{n}} \right] X_{t:n}, \quad j=1, \dots, n,$$

and estimate $Q(u)$ by $\widehat{Q}_{r(u):n}$ with $r(u) = \lceil un \rceil$. The closed-form bootstrap variance estimator is given $\widehat{\sigma}_{KC}^2 = n^{-1} \sum_{j=1}^n b_{r(u)}(j) (\widehat{Q}_{r(u):n} - \widehat{Q}_{j:n})^2$. This method is henceforth referred as the “KC91” procedure.

2.2 Interval estimation of a quantile ratio

Let $X_{kj}, j = 1, \dots, n_k$, be a random sample from a continuous distribution on $(0, \infty)$ with quantile function $Q_k(\cdot)$, $k = 1, 2$. The goal is to estimate $Q_2(u)/Q_1(u)$ for a fixed $u \in (0, 1)$, with a CI or an upper (or lower) confidence bound. It is more convenient to deal with instead $\log(Q_2(u)) - \log(Q_1(u))$ first, and then exponentiate. A quantile function is transformation equivariant, i.e., if X has quantile function $Q(\cdot)$ and $g(\cdot)$ is a strictly monotone function, then $g(X)$ has quantile function $g(Q(\cdot))$. Therefore the log-transformed quantile $\log(Q_k(u))$ can be naturally estimated based on the log-transformed data $\log(X_{kj}), j = 1, \dots, n_k$ using the estimators outlined above, for $k = 1, 2$. Thus let $\widehat{\ell}Q_k(u)$ and $\widehat{\ell}q_k(u)$ be respectively the Bernstein polynomial estimators of $\log(Q_k(u))$ and the corresponding quantile density based on the log-transformed data, let $\widehat{\sigma}_k^2 = [\widehat{\ell}q_k(u)]^2 u(1-u)$, $k = 1, 2$, and let $\widehat{\sigma} = (\widehat{\sigma}_2^2/n_2 + \widehat{\sigma}_1^2/n_1)^{1/2}$; then a $100(1 - \alpha)\%$ CI for $\log(Q_2(u)) - \log(Q_1(u))$ is given by $[\widehat{\ell}Q_2(u) - \widehat{\ell}Q_1(u)] \pm \widehat{\sigma}\Phi^{-1}(1 - \alpha/2)$, and it follows that the CI for $Q_2(u)/Q_1(u)$ is

$$\frac{\exp\{\widehat{\ell}Q_2(u)\}}{\exp\{\widehat{\ell}Q_1(u)\}} \exp\{\pm \widehat{\sigma}\Phi^{-1}(1 - \alpha/2)\}.$$

Similarly a $100(1 - \alpha)\%$ one-sided upper confidence bound for the quantile ratio is

$$\frac{\exp\{\widehat{\ell}Q_2(u)\}}{\exp\{\widehat{\ell}Q_1(u)\}} \exp\{\widehat{\sigma}\Phi^{-1}(1 - \alpha)\}.$$

This method is henceforth called the Bernstein polynomial plug in (BPP) procedure. The CFBoot and KC91 procedures can be extended to interval estimation of quantile ratios in the same manner.

An alternative approach is Koenker's regression quantile model [8]. Consider the regression model $X_{ki} = \beta_0 + \beta_1 Z_{ki} + \varepsilon_{ki}$, $i = 1, \dots, n_k$, $k = 1, 2$, where $Z_{ki} = 0$ if $k = 1$ and $Z_{ki} = 1$ if $k = 2$, and ε_{ki} are random errors (not necessarily i.i.d). In the solution $(\widehat{\beta}_0, \widehat{\beta}_1)$ that minimizes the fitting criterion $\sum_{k=1}^2 \sum_{i=1}^{n_k} E_{ki}(u - I(E_{ki} < 0))$, $[0 < u < 1, E_{ki} = X_{ki} - \beta_0 - \beta_1 Z_{ki}]$, $\widehat{\beta}_1$ is an

estimator of $Q_2(u) - Q_1(u)$. A rank test of the null hypothesis $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ can be inverted to obtain a confidence interval [26]. For quantile ratio estimation one simply applies this procedure to log-transformed data and then back transform, as described earlier.

3. A Simulation Study

3.1 Simulation setup

This simulation study compares the proposed BPP method, regression quantiles, CF-Boot, and KC91 on their performances to produce 95% confidence interval (CI) and upper confidence bound (UCB). Ratios of the 3 quartiles and the 90th percentile in typically small sample scenarios are considered. We consider a variety of cases constructed from four right-skewed population distributions on the positive half of the real line, with various tail behaviors ranging from finite to heavy right tail. Details of 6 simulated cases are given in Table 1. These 6 cases represent a wide range of scenarios determined by the tail behavior of the population quantile function in the numerator and denominator of the quantile ratios.

Performance measurements include coverage probability (i.e., accuracy), median confidence interval length (precision), and inter-quartile range (IQR) of confidence interval length (stability). Because of the directional nature of UCB, the precision is measured by the “bound length” defined as the computed UCB minus the true quantile ratio; thus a negative value indicates that the UCB fails to cover the true quantile ratio. Median of the bound length is estimated. Moreover, rather than IQR, the 1st and the 99th percentiles of the bound length are estimated to assess stability of the UCB; the interval between these two percentiles is herein referred as inter-percentile range (IPR). Two sample size configurations for the denominator and numerator group are considered: $(n_1, n_2) = (10, 10)$ and $(n_1, n_2) = (30, 25)$. These choices represent typically small, but not unreasonably small, sample sizes in most applications. Random samples in each group are generated by evaluation of the corresponding population quantile function on Uniform(0,1) random numbers. Ten thousand (10,000) simulation trials are run for each case and each sample size configuration to estimate the performance measurements. All methods are implemented with R. Regression quantiles and their CI (UCB) are computed using the `rq` function with options `ci=T` and `iid=F`.

3.2 Performance for median ratios

CI performances are compiled in Table 2. For both sample size configurations $(n_1, n_2) = (10, 10)$ and $(n_1, n_2) = (30, 25)$ BPP and regression quantiles perform largely equally well. For $(n_1, n_2) = (10, 10)$ CFBoot and KC91 do not generally perform as well as the other two, either for poor coverage probability or for less precision; when the sample size is increased to (35,20), both methods become overly conservative, giving coverage probabilities much higher than the nominal 95% level at the cost of greatly reduced precision and stability.

Results for UCBs are compiled in Table 3. The proposed BPP method performs well in all cases for both sample size configurations. Notably in all but one case regression quantiles give coverage probabilities below the nominal 95% level, with median bound lengths slightly smaller than BPP’s. Again CFBoot and KC91 are inferior for sample size (10,10) and become overly conservative as the sample size is increased to (35,20).

In all cases and sample size configurations, across simulation runs the `rq` function frequently gives the warning message “Solution may be nonunique”, but the returned solution usually works fine.

3.4 CI performance for other quantile ratios

Results on the 1st quartile and the 90th percentile ratios are compiled in Tables S1 and S2 in Supplement. For the 1st quartile ratio, at both sample size configurations BPP and regression quantiles both perform well; CFBoot and KC91 perform poorly primarily due to the underestimation of variance. For the 90th percentile ratio at the sample size (10,10) BPP performs poorly and quantile regression method cannot provide a confidence interval although the point estimate is fine. Interestingly CFBoot and KC91 are able to give reasonable CIs except for Case 5. At the sample size (30,25) regression quantiles give slightly better coverage probability than BPP with small increase in median CI length in all cases; and CFBoot and KC91 again are very conservative now. See Supplement for more discussion.

In all cases and all methods considered here, notably the precision and stability of CI or UCB decrease as the quantile ratio increase. Location and scale of a right-skewed distribution are intimately related; increase in location results in increase in scale. The distributions of the point estimators (pivots) of quantile ratios may have inherited this characteristic, entailing that reliable and precise estimation of large quantile ratios is more difficult to achieve, and the quality of the estimator is not invariant to which quantile is put in the numerator (respectively denominator) — a better (worse) interval estimator of the quantile ratio can be obtained for the reciprocal than for the ratio itself. Obtaining a rigorous analytical handle on this phenomenon remains an open problem.

4. An Application to Xenograft Experiments

The data are taken from a published single-agent tumor xenograft study conducted by the PPTP [1]. In this study, vincristine was administered intraperitoneally at a dosage of 1 mg/kg every 7 days for 6 weeks and evaluated in 24 tumor cell lines as a single agent. For each tumor line, nearly 20 tumor-bearing mice were equally randomized into control and treatment groups. The volume of each tumor is measured at the initiation of the study and weekly for 6 weeks. Mice are euthanized when the tumor volume reaches four times its initial volume. Table 4 shows 20 solid tumor cell lines together with the T/C ratios (ratio of two means) and antitumor activity ratings published in [1]. The T/C ratio was calculated either on day 21 or earlier when all mice in both groups still had measurable tumor volumes. Table 4 also shows the estimated median ratios (MR), log median ratios along with standard errors, and 95% one-sided upper confidence bounds (UCB) of the median ratios by the BPP procedure described in Section 2. The PPTP study used a rather arbitrary threshold for the T/C ratio antitumor activity rating. Agents producing a $T/C \leq 0.15$ were considered highly active; those with $0.15 < T/C \leq 0.45$ were considered to have intermediate activity; and those with $T/C > 0.45$ were considered to have low activity. This approach clearly does not account for the noise in the experimental data. Assessment of antitumor activity based on the confidence bound approach is much more interpretable in the presence of experimental noise: Whenever the upper confidence bound is lower than 1, a significant antitumor activity is observed; otherwise, no significant antitumor activity can be claimed. For the tumor lines KT-13 and Rh28 vincristine is rated to have high activity by PPTP's T/C ratio; the 95% one-sided upper confidence bound of the median ratio clearly show strong statistical evidence to support this rating after accounting for experimental noise. PPTP rates the agent to have low activity in five tumor lines, KT-16, KT-14, EW5, SK-N-AS and NB-1691; the 95% UCBs of median ratio do not provide sufficient evidence that the agent is active in these tumor cell lines either. For a number of tumor lines vincristine is rated to have low activity by PPTP's T/C ratio; whereas the 95% UCBs of median ratio show evidence that the agent is active, such as in the lines BT-29, Rh10, Rh30, Rh30R and NB-1771. In several tumor lines vincristine is rated to have intermediate activity by PPTP's T/C ratio; whereas the 95%

UCBs of median ratio show strong evidence that the agent is highly active, such as in the lines BT-50, SKNEP, Rh18, SJ-BT39 and Rh41.

Notably there is a large discrepancy between the T/C ratio and the median ratio for the BT-50 cell line. Figure 1 shows the raw and log-transformed tumor volume order statistics in the control and treatment group. There appears to be two subgroups in the treatment group: a majority that responded well to the treatment and a minority that did not. Clearly the T/C ratio is not robust, and not a single family of parametric models can capture the diverse distributional behavior in these data. The regression quantiles method gave the median ratio estimate 0.03, 95% UCB 0.78, and CI (0.01, 1.05); the rq function gave a warning “Solution may be nonunique.” The BPP method gave the median ratio estimate 0.04, 95% UCB 0.2, and CI (0.005, 0.28).

5. Conclusion

We have developed a sound and rigorous inference procedure to assess treatment effect in anti-cancer drug screening by xenograft experiments. Our method consists of a novel interval estimator of quantile ratios for right-skewed distributions, justified by a rigorous large-sample theory and a simulation study. The BPP procedure performs well both in a variety of simulated cases and in the application to xenograft experiments, even for small sample size. For CI and UCB of median ratios BPP performs the best among all the methods compared in the simulation study; and BPP is a completely nonparametric approach.

When the population distributions are right-skewed, it is easier to achieve adequate precision and accuracy of confidence bounds and intervals for small (fractional) population quantile ratios. Thus in practice when there is an expected direction of effect implied by the subject matter, it would be helpful to formulate the population quantile ratio in such a way that it is a fraction. The xenograft experiment is such an example — one expect that tumor growth in the treatment group is no faster than that in the control group even if the agent has no effect.

Theory of quantile estimation shows that the variance of any reasonable estimator of the u th quantile is directly proportional to how fast the quantile function changes at u , i.e, the magnitude of the quantile density $q(u)$. Consequently, the precision of interval estimation depends on $q(u)$; the higher the quantile density, the less precise is the quantile estimator. Additionally, estimators of the variance of a quantile point estimate, especially the closed-form bootstrap variance estimator, can be sensitive to outliers. Thus as always, it is a good practice to examine the original and log-transformed data graphically before the formal analysis. For example, sample quantile plots (order statistics) provide valuable information on the behavior of the quantile density $q(u)$, as well as the existence of potential outliers.

Given the small sample size in the xenograft application it is natural to consider finite sample, exact methods for quantile interval estimation. A recent development for a single quantile can be found in [30] (see also references therein). Extension of these methods to two-sample problems remains an open problem.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank two referees and an associate editor for their careful reading of the manuscript and constructive comments which have helped improve the paper substantially. This research is supported in part by the National Cancer Institute support grants NO1-CM-42216 and CA21765, and American Lebanese and Syrian Associated Charities (ALSAC).

References

1. Houghton PJ, Morton CL, Gorlick R, et al. The Pediatric Preclinical Testing Program: description of models and early testing results. *Pediatric Blood and Cancer*. 2007; 49:928–940. [PubMed: 17066459]
2. Corbett TH, White K, Polin L, et al. Discovery and preclinical antitumor efficacy evaluations of LY32262 and LY33169. *Investigational New Drugs*. 2003; 21:33–45. [PubMed: 12795528]
3. Hothorn L. Statistical analysis of in vivo anti-cancer experiments: tumor growth inhibition. *Drug Information Journal*. 2006; 40:229–238.
4. Zhou XH, Tu W. Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics and Data Analysis*. 2000; 35:201210.
5. Chen YH, Zhou XH. Interval estimates for the ratio and difference of two lognormal means. *Statistics in Medicine*. 2006; 25:4099–4113. [PubMed: 16479559]
6. Wu J, Jiang G, Wong ACM, Sun X. Likelihood analysis for the ratio of means of two independent log-normal distributions. *Biometrics*. 2002; 58:463469.
7. Weerahandi S. Generalized confidence intervals. *Journal of American Statistical Association*. 1993; 88:899–905.
8. Koenker RW, Bassett GW. Regression quantiles. *Econometrica*. 1978; 46:33–50.
9. Dominici F, Cope L, Naiman DQ, Zeger SL. Smooth quantile ratio estimation. *Biometrika*. 2005; 92:543–557.
10. Dominici F, Zeger SL. Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable diseases. *Biostatistics*. 2005; 6:505–519. [PubMed: 15872022]
11. Fieller EC. Some problems in interval estimation. *Journal of the Royal Statistical Society B*. 1954; 16:175–185.
12. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. Chapman & Hall/CRC; New York: 1993.
13. Bahadur RR. A note on quantiles in large samples. *Annals of Mathematical Statistics*. 1966; 37:577–580.
14. Reiss R-D. Asymptotic expansions for sample quantiles. *Annals of Probability*. 1976; 4:249–258.
15. Csörgő M, Révész P. Strong approximations of the quantile process. *Annals of Statistics*. 1978; 6:882–894.
16. Parzen E. Nonparametric statistical data modeling (with comments). *Journal of American Statistical Association*. 1979; 74:105–131.
17. Harrell FE, Davis CE. A new distribution-free quantile estimator. *Biometrika*. 1982; 69:635–640.
18. Kaigh WD, Lachenbruch PA. A generalized quantile estimator. *Communications in Statistics, Series A*. 1982; 11:2217–2238.
19. Kaigh WD, Cheng C. Subsampling quantile estimators and uniformity criteria. *Communications in Statistics, Series A*. 1991; 20:539–560.
20. Falk M. Asymptotic normality of the kernel quantile estimator. *Annals of Statistics*. 1985; 13:428–433.
21. Yang SS. A smooth nonparametric estimator of a quantile function. *Journal of American Statistical Association*. 1985; 80:1004–1011.
22. Zelterman D. Smooth nonparametric estimation of the quantile function. *Journal of Statistical Planning and Inference*. 1990; 26:339–352.
23. Sheather SJ, Marron JS. Kernel quantile estimators. *Journal of American Statistical Association*. 1990; 85:410–416.
24. Cheng C, Parzen E. Unified estimators of smooth quantile and quantile density functions. *Journal of Statistical Planning and Inference*. 1997; 59:291–307.
25. Kaigh WD, Cheng C. Subsampling quantile estimators standard errors with applications. *Communications in Statistics, Series A*. 1991; 20:977–995.
26. Koenker, RW. Confidence intervals for regression quantiles. In: Mandl, P.; Huskova, M., editors. *Asymptotic Statistics*. Springer-Verlag; New York: 1994. p. 349-359.

27. Cheng C. Uniform consistency of generalized kernel estimators of quantile density. *Annals of statistics*. 1995; 23:2285–2291.
28. Cheng C. The Bernstein polynomial estimator of a smooth quantile function. *Statistics and Probability Letters*. 1995; 24:321–330.
29. Serfling, RJ. *Approximation Theorems of Mathematical Statistics*. Wiley; New York: 1980.
30. Ozturk O, Deshpande JV. Rank-set sample nonparametric quantile confidence intervals. *Journal of Statistical Inference and Planning*. 2006; 136:570–577.

Appendix

A.1 Asymptotic Normality and variance

For notational convenience, write the Bernstein polynomial quantile estimator into an

integral form: $\widehat{Q}(u) = \int_0^1 \tilde{Q}(t) K_n(u, t) d\mu_n(t)$, where $K_n(u, t) = \binom{n-1}{[nt]-1} u^{[nt]-1} (1-u)^{n-[nt]}$, $t \in (0, 1)$; $\mu_n(\cdot)$ is a step function with jumps at j/n ($j = 1, \dots, n$): $d\mu_n(t) = 1$ for $t \in \{j/n : j = 1, \dots, n\}$, and $= 0$ otherwise. Under the conditions on the quantile function $Q(\cdot)$ outlined in [24], Theorem 2.1 therein implies that $\widehat{Q}(u)$ is Asymptotically Normal with mean $Q(u)$ and

asymptotic variance $\sigma_n^2 = E \left[\left(\int_0^1 q(t) B_n(t) K_n(u, t) d\mu_n(t) \right)^2 \right]$, where the expectation is taken with respect to $B_n(\cdot)$, a Brownian bridge on $[0, 1]$. Next, properties of Bernstein polynomial imply that in general if $g(\cdot)$ is a twice differentiable function on $(0, 1)$ then

$\widehat{g}(u) = \int_0^1 g(t) K_n(u, t) d\mu_n(t) = g(u) + O(n^{-1})$ as $n \rightarrow \infty$, for any fixed $u \in (0, 1)$. Hence with $\sigma^2 = [q(u)]^2 u(1-u)$,

$$\begin{aligned} \sigma_n^2 &= E \left[\int_0^1 q(t) B_n(t) K_n(u, t) d\mu_n(t) \int_0^1 q(s) B_n(s) K_n(u, s) d\mu_n(s) \right] \\ &= \int_0^1 \int_0^1 q(t) q(s) E [B_n(t) B_n(s)] K_n(u, t) K_n(u, s) d\mu_n(t) d\mu_n(s) \\ &= \int_0^1 \int_0^1 q(t) q(s) [\min\{t, s\} - st] K_n(u, t) K_n(u, s) d\mu_n(t) d\mu_n(s), \end{aligned}$$

and after a long calculation,

$$\begin{aligned} \sigma_n^2 &= \int_0^1 t q(t) K_n(u, t) d\mu_n(t) \int_0^1 (1-s) q(s) K_n(u, s) d\mu_n(s) - \int_0^1 q(t) \left[\int_0^t (t-s) q(s) K_n(u, s) d\mu_n(s) \right] K_n(u, t) d\mu_n(t) \\ &= \sigma^2 - \xi_n + b_n, \quad n \rightarrow \infty, \end{aligned}$$

where $\xi_n = \int_0^1 q(t) \left[\int_0^t (t-s) q(s) K_n(u, s) d\mu_n(s) \right] K_n(u, t) d\mu_n(t)$. In the Bernstein polynomial case, $b_n = O(n^{-1})$ and $\xi_n = O(n^{-1/2})$; details can be found in [24]. Note that $\xi_n > 0$ for all n and dominates the bias term b_n .

A.2 Asymptotic coverage probability

Consider a single quantile first. Recall that $\widehat{\sigma}^2 = [\widehat{q}(u)]^2 u(1-u)$. The quantile density estimator $\widehat{q}(u)$ converges to $q(u)$ in probability [27]; thus $\widehat{\sigma}^2$ converges to σ^2 in probability.

The Asymptotic Normality in A.1, along with Slutsky's theorem implies that $\frac{\widehat{Q}(u) - Q(u)}{\widehat{\sigma}/n^{1/2}}$ converges in distribution to $N(0, 1)$. For the same reason, the normalized log-quantile difference estimator (cf. Section 2.2 for notation)

$\left[\left(\widehat{\ell Q_2}(u) - \widehat{\ell Q_1}(u) \right) - [\log(Q_2(u)) - \log(Q_1(u))] \right] / \widehat{\sigma}$ converges in distribution to $N(0, 1)$.
Therefore for UCB

$$\lim_{n \rightarrow \infty} \Pr \left(\log \frac{Q_2(u)}{Q_1(u)} \leq \left(\widehat{\ell Q_2}(u) - \widehat{\ell Q_1}(u) \right) + \widehat{\sigma} \Phi^{-1}(1 - \alpha) \right) = \lim_{n \rightarrow \infty} \Pr \left(\frac{Q_2(u)}{Q_1(u)} \leq \frac{\exp\{\widehat{\ell Q_2}(u)\}}{\exp\{\widehat{\ell Q_1}(u)\}} \exp\{\widehat{\sigma} \Phi^{-1}(1 - \alpha)\} \right) = 1 - \alpha.$$

Likewise, for CI

$$\lim_{n \rightarrow \infty} \Pr \left(\log \frac{Q_2(u)}{Q_1(u)} \in \left(\widehat{\ell Q_2}(u) - \widehat{\ell Q_1}(u) \right) \pm \widehat{\sigma} \Phi^{-1}(1 - \alpha/2) \right) = \lim_{n \rightarrow \infty} \Pr \left(\frac{Q_2(u)}{Q_1(u)} \in \frac{\exp\{\widehat{\ell Q_2}(u)\}}{\exp\{\widehat{\ell Q_1}(u)\}} \exp\{\pm \widehat{\sigma} \Phi^{-1}(1 - \alpha/2)\} \right) = 1 - \alpha.$$

Figure 1. Linearly interpolated tumor volume order statistics in the control(circle) and treatment (bullet) group respectively, from the BT-50 experiment.

Table 1
Six cases in the simulation study: Population quantile functions and the ratios of true quartiles and 90th percentile.

Case	Denominator Q_1	Numerator Q_2	0.5	0.25	0.75	0.9
1	$3 \exp\{Q_{\text{Normal}}(:, 0, 1)\}$	$Q_{\text{Beta}}(:, 1.2, 6)$	0.0452	0.0425	0.0404	0.0326
2	$[Q_{\text{Gamma}}(:, 1.8, 0.9)]^{1/2}$	$Q_{\text{Beta}}(:, 1.2, 6)$	0.1057	0.0681	0.1447	0.1761
3	$3 \exp\{Q_{\text{Normal}}(:, 0, 1)\}$	$Q_{\text{Gamma}}(:, 2, 1)$	0.5594	0.6290	0.4572	0.3599
4	$[Q_{\text{Gamma}}(:, 1.8, 0.9)]^{1/2}$	$[Q_{\text{Gamma}}(:, 1.8, 0.9)]^{1/2}$	1.0000	1.0000	1.0000	1.0000
5	$[Q_{\text{Gamma}}(:, 1.8, 0.9)]^{1/2}$	$3 \exp\{Q_{\text{Normal}}(:, 0, 1)\}$	2.3157	1.5587	3.5889	5.4794
6	$Q_{\text{Beta}}(:, 1.2, 6)$	$[Q_{\text{Gamma}}(:, 1.8, 0.9)]^{1/2}$	9.4598	7.0253	12.1467	14.7325

Distribution name is given in the subscript and the parameters are given in the parentheses after the semicolon; for example, $3 \exp\{Q_{\text{Normal}}(:, 0, 1)\}$ means the population quantile function is 3 times the $\log\text{-normal}(0, 1)$ quantile function.

Table 2

CI performance for median ratio at 95% nominal level.

Case	Summary Stat.	BPP: Proposed	Reg. Quant.	CFBoot	KC91
$n_1 = n_2 = 10$					
1	Cover.Prob(se)	0.958(0.002)	0.960(0.001)	0.951(0.002)	0.930(0.002)
	Med.Length(IQR)	0.105(0.077)	0.108(0.078)	0.151(0.157)	0.090(0.067)
2	Cover.Prob(se)	0.946(0.002)	0.952(0.002)	0.941(0.002)	0.922(0.002)
	Med.Length(IQR)	0.174(0.102)	0.180(0.098)	0.254(0.281)	0.156(0.111)
3	Cover.Prob(se)	0.953(0.002)	0.959(0.001)	0.941(0.002)	0.920(0.002)
	Med.Length(IQR)	1.173(0.847)	1.212(0.870)	1.475(1.295)	0.969(0.668)
4	Cover.Prob(se)	0.954(0.002)	0.957(0.002)	0.957(0.002)	0.933(0.002)
	Med.Length(IQR)	0.886(0.404)	0.942(0.464)	1.184(0.727)	0.824(0.404)
5	Cover.Prob(se)	0.942(0.002)	0.942(0.002)	0.915(0.002)	0.900(0.002)
	Med.Length(IQR)	4.083(2.949)	4.216(3.029)	4.229(4.041)	3.081(2.152)
6	Cover.Prob(se)	0.947(0.002)	0.9465(0.002)	0.9451(0.002)	0.926(0.002)
	Med.Length(IQR)	16.27(13.34)	16.42(14.21)	26.61(26.69)	17.70(15.64)
$n_1 = 30, n_2 = 25$					
1	Cover.Prob(se)	0.957(0.002)	0.946(0.002)	0.997(0.0005)	0.997(0.0005)
	Med.Length(IQR)	0.061(0.029)	0.058(0.028)	0.138(0.091)	0.105(0.055)
2	Cover.Prob(se)	0.949(0.002)	0.945(0.002)	0.997(0.0005)	0.997(0.0005)
	Med.Length(IQR)	0.109(0.047)	0.104(0.044)	0.265(0.177)	0.201(0.108)
3	Cover.Prob(se)	0.954(0.002)	0.941(0.002)	0.995(0.0006)	0.993(0.0007)
	Med.Length(IQR)	0.684(0.324)	0.646(0.320)	1.284(0.657)	1.032(0.445)
4	Cover.Prob(se)	0.957(0.002)	0.9481(0.002)	0.998(0.0004)	0.997(0.0004)
	Med.Length(IQR)	0.544(0.182)	0.522(0.189)	1.100(0.424)	0.899(0.300)
5	Cover.Prob(se)	0.949(0.002)	0.944(0.002)	0.991(0.0009)	0.990(0.0009)
	Med.Length(IQR)	2.547(1.309)	2.486(1.289)	4.143(2.319)	3.459(1.642)
6	Cover.Prob(se)	0.95(0.002)	0.943(0.002)	0.998(0.0004)	0.997(0.0005)
	Med.Length(IQR)	9.184(4.752)	9.020(5.032)	23.43(13.70)	18.65(9.54)

Estimated coverage probability (Cover.Prob) is given along with standard error (se); median CI length (Med.Length) is given along with IQR of CI length. The upper half shows $n_1 = n_2 = 10$, and the lower half shows $n_1 = 30, n_2 = 25$.

Table 3

UCB performance for median ratio at 95% nominal level.

Case	Summary Stat.	BPP: Proposed	Reg. Quant.	CFBoot	KC91
$n_1 = n_2 = 10$					
1	Cover.Prob(se)	0.957(0.002)	0.923(0.002)	0.958(0.002)	0.928(0.002)
	Med.Length(IPR)	0.059(-0.011,0.264)	0.049(-0.016,0.255)	0.090(-0.014,0.691)	0.048(-0.015,0.245)
2	Cover.Prob(se)	0.957(0.002)	0.962(0.001)	0.928(0.002)	0.909(0.002)
	Med.Length(IPR)	0.095(-0.023,0.333)	0.100(-0.022,0.341)	0.144(-0.043,1.236)	0.077(-0.037,0.371)
3	Cover.Prob(se)	0.952(0.002)	0.921(0.002)	0.960(0.001)	0.932(0.002)
	Med.Length(IPR)	0.648(-0.149,2.983)	0.551(-0.206,3.418)	0.876(-0.139,4.891)	0.527(-0.159,2.395)
4	Cover.Prob(se)	0.956(0.002)	0.922(0.002)	0.962(0.001)	0.945(0.002)
	Med.Length(IPR)	0.441(-0.127,1.483)	0.387(-0.176,1.578)	0.623(-0.124,2.350)	0.410(-0.141,1.430)
5	Cover.Prob(se)	0.949(0.002)	0.939(0.002)	0.899(0.003)	0.896(0.003)
	Med.Length(IPR)	2.309(-0.556,10.20)	2.227(-0.629,10.75)	2.221(-0.927,15.08)	1.559(-0.803,7.784)
6	Cover.Prob(se)	0.948(0.002)	0.923(0.002)	0.982(0.001)	0.966(0.001)
	Med.Length(IPR)	9.050(-1.917,58.25)	8.462(-2.621,60.32)	16.36(-0.663,122.69)	10.89(-1.549,64.54)
$n_1 = 30, n_2 = 25$					
1	Cover.Prob(se)	0.960(0.001)	0.943(0.002)	0.996(0.0006)	0.995(0.0007)
	Med.Length(IPR)	0.032(-0.007,0.111)	0.028(-0.009,0.103)	0.082(0.005,0.326)	0.060(0.003,0.192)
2	Cover.Prob(se)	0.964(0.001)	0.942(0.002)	0.994(0.0007)	0.993(0.0007)
	Med.Length(IPR)	0.056(-0.013,0.167)	0.049(-0.019,0.155)	0.153(0.008,0.629)	0.112(0.005,0.353)
3	Cover.Prob(se)	0.961(0.001)	0.942(0.002)	0.995(0.0006)	0.994(0.0007)
	Med.Length(IPR)	0.356(-0.087,1.201)	0.323(-0.117,1.15)	0.743(0.058,2.342)	0.577(0.032,1.637)
4	Cover.Prob(se)	0.959(0.001)	0.941(0.002)	0.997(0.0004)	0.998(0.0004)
	Med.Length(IPR)	0.260(-0.065,0.727)	0.235(-0.097,0.699)	0.573(0.075,1.387)	0.460(0.060,1.034)
5	Cover.Prob(se)	0.960(0.001)	0.941(0.002)	0.984(0.001)	0.986(0.001)
	Med.Length(IPR)	1.364(-0.318,4.717)	1.198(-0.426,4.338)	2.287(-0.119,7.927)	1.884(-0.074,5.886)
6	Cover.Prob(se)	0.951(0.002)	0.938(0.002)	0.999(0.0002)	0.998(0.0003)
	Med.Length(IPR)	4.722(-1.248,18.43)	4.586(-1.488,18.35)	13.73(2.573,52.13)	10.81(2.032,32.71)

Estimated coverage probability (Cover.Prob) is given along with standard error (se); median bound length (Med.Length) is given along with inter-percentile range (IPR) of bound length. The upper half shows $n_1 = n_2 = 10$, and the lower half shows $n_1 = 30, n_2 = 25$.

Table 4

Antitumor activity evaluation in the PPTP vincristine xenograft study.

Tumor line	Number of mice (n_C , n_T)	T/C ratio		Median Ratio (MR)		
		T/C	Activity	MR	log(MR)(s.e.)	95% UCB
BT-29	(10, 9)	0.64	Low	0.65	-0.43(0.24)	0.97
KT-16	(10, 10)	0.50	Low	0.67	-0.41(0.31)	1.10
KT-14	(10, 9)	0.89	Low	0.90	-0.10(0.26)	1.37
KT-13	(10, 10)	0.07	High	0.03	-3.54(1.12)	0.18
SKNEP	(10, 10)	0.25	Int	0.25	-1.37(0.24)	0.38
EW5	(10, 10)	0.79	Low	0.87	-0.14(0.22)	1.26
Rh10	(10, 10)	0.48	Low	0.55	-0.59(0.19)	0.76
Rh28	(10, 10)	0.14	High	0.06	-2.73(1.15)	0.43
Rh30	(10, 9)	0.51	Low	0.56	-0.57(0.15)	0.72
Rh30R	(10, 10)	0.68	Low	0.64	-0.45(0.21)	0.90
Rh41	(10, 10)	0.34	Int	0.30	-1.21(0.22)	0.43
Rh18	(10, 10)	0.16	Int	0.13	-2.03(0.22)	0.19
BT-50	(10, 9)	0.37	Int	0.04	-3.27(1.01)	0.20
SJ-BT39	(10, 10)	0.34	Int	0.33	-1.11(0.18)	0.44
D645	(10, 10)	0.43	Int	0.35	-1.05(0.41)	0.69
D212	(10, 10)	0.36	Int	0.44	-0.82(0.21)	0.63
NB-1771	(9, 9)	0.61	Low	0.58	-0.55(0.18)	0.78
NB-1691	(10, 10)	0.80	Low	0.75	-0.29(0.23)	1.09
CHLA-79	(9, 9)	0.68	Low	0.67	-0.40(0.22)	0.96
SK-N-AS	(10, 10)	0.94	Low	0.88	-0.13(0.15)	1.13
						NS

T/C ratio is the ratio of mean in treatment group to mean in control group. PPTP ratings are High, Low, and Intermediate (Int). Median ratio and 95% UCB are computed by the proposed BPP method. NS stands for "Not Significant" (i.e., not sufficient statistical evidence for activity).