COMPARING PROPERTIES OF VARIOUS TROUGH/PEAK RATIO ESTIMATORS

Mahjoob K., Kazempour M.K., Chi George Y.H.[1]
Division of Biometrics, US Food and Drug Administration
Contact Person: Kooros Mahjoob; 5600 Fishers Ln, Room 18B-45 (HFD-713), Rockville, MD 20857

ABSTRACT

The FDA Draft Guidelines for the study of antihypertensive drugs indicates that the trough/peak ratio (T/P) "should be no less than 1/2 to 2/3, depending on the magnitude of the effect." There is no generally agreed upon definition for such a ratio. The ratio may be estimated as the sample mean of the individual subjects' ratios or may be estimated as the sample trough-mean over the sample peak-mean. In this paper various models for constructing T/P are considered along with different ways of obtaining their variances and confidence intervals. The properties of these estimators are assessed and compared theoretically and numerically. Data from a clinical trial is utilized to evaluate the proposed ratios, variances and confidence intervals numerically.

Key Words: Trough/Peak ratio, hypertension, Ambulatory Blood Pressure Monitoring.

1. INTRODUCTION

The trough-to-peak ratio (T/P) is becoming a part of the drug evaluation in the clinical trials, especially in dose-ranging studies. The clinicians use this ratio as a yard stick to determine the therapeutically beneficial dose range of a new treatment and to prevent patients from being exposed to unnecessarily high dose just to gain needed duration. The terms "trough" and "peak" refer to the "minimum" and "maximum" effects of the study drug within a given interval of time, usually a 24-hour period post-dose for a sustained release once-a-day drug. A continuous monitoring of the patient within the dosing interval is required to determine the exact trough and/or peak values. However, in general the continuous determination of the response is impractical or impossible for office blood pressures. In practice the relationship between the pharmacokinetics (PK) time to maximum plasma drug concentration (T_{max}) and the pharmacodynamic (PD) time to maximum drug effect (T_{max} + time lag) may be used to determine the peak. The trough is assumed to occur at the end of the dosing interval.

In this work we restrict ourselves to T/P issues in hypertension (HT) studies. In HT studies the use of the ambulatory blood pressure monitoring (ABPM)[1] device may help to obtain a good approximation of the trough and peak values. There are studies on the accuracy, reliability and correlation between the ABPM data with office blood pressure[2-4]. In any event here the main focus is on the evaluation of various T/P estimation and not the accuracy and/or the validation of ABPM data.

There are several methods of modeling for the T/P estimators. These models and the methods of constructing confidence intervals may have a major impact on the final results. Customarily, the drug sponsors use Model II, see next section, to calculate the ratio and the Fieller's[5] theorem to construct the confidence interval for the ratio. Stewart and Hafner[6] reported an exact method in obtaining confidence intervals for the parallel, placebo-controlled trials. Other methods such as bootstrapping[7], Taylor expansion[8] can also be applied to construct confidence intervals. There is no general agreement on how the ratio is calculated. To our knowledge there is no mention of any comparison of the properties of these estimators and their confidence intervals in the literature.

In the next section, notations are introduced. Four different ways of computing the T/P ratios and three methods for constructing confidence intervals are presented in Section 3. A numerical example is discussed in Section 4. In Section 5 a brief discussion and the conclusion are presented. The mathematical derivations and the description of bootstrapping are reserved for the Appendices.

---

[1] The views expressed here are those of the authors and not of the US Food and Drug Administration.

2. NOTATION and MOTIVATION

The experiences accrued as a result of analyzing several antihypertensive clinical trials and observing the variations in the methods that the drug sponsors had used, motivated us to work on this problem. The main objective of a HT management is to reduce the patient's blood pressure and to maintain it within a desired level. Therefore, an effective once-a-day antihypertensive medication should demonstrate its effectiveness within the dosing interval; especially at the end of the dosing interval (trough). The "peak effect" of an antihypertensive drug should also be controlled to avoid over-dosing. The "Draft Guidelines for Antihypertensive Drug Products" of the Division of Cardio-Renal Drug Products of the Food and Drug Administration (FDA), indicates that "the drug effect at trough (measured as the difference from the placebo effect) should be no less than 1/2 to 2/3 of the peak effect, depending on the magnitude of the effect." The FDA has not yet publicly determined precisely how T/P should be measured.

To present T/P models the following convention and notations are adopted. Consider a placebo-controlled HT trial with one active treatment. The treatment is given once-a-day, the changes from baseline in diastolic blood pressure (dbp) is the primary efficacy variable. It is assumed that hourly ABPM measurements are obtained for 24-hours at the baseline and at the end of the treatment period. Further it is assumed that:

- there is a population of size 2N hypertensive patients, from which N could randomly be assigned to the placebo (d=0) and N to the active treatment (d=1),

- from a random sample of size 2n of these patients, n are assigned to placebo and n to the active treatment.

- $\Delta Y_{id(h)}$, $\Delta Y_{id(T)}$, and $\Delta Y_{id(P)}$ = Change from baseline in dbp of the i^{th} subject who is receiving medication d at Hour h post-dose, at trough, and at peak respectively; where i=1,2,.., N; d = 0, 1; and h=1, 2,..., 24.

The sample and population mean changes from baseline at Hour h post-dose, at trough, and at peak are:

$\Delta \bar{Y}_{d(h)}$, $\Delta \bar{Y}_{d(T)}$, $\Delta \bar{Y}_{d(P)}$ for sample,

$\Delta\mu_{d(h)}$, $\Delta\mu_{d(T)}$, $\Delta\mu_{d(P)}$ for population,

where

$$\Delta \bar{Y}_{d(h)} = \frac{1}{n}\sum_{i=1}^{n} \Delta Y_{id(h)} \;\&\; \Delta \mu_{d(h)} = \frac{1}{N}\sum_{i=1}^{N} \Delta Y_{id(h)}.$$

3. TROUGH/PEAK RATIO: POINT ESTIMATES, CONFIDENCE INTERVALS

a) Point Estimate

There are various ways to formulate and interpret the trough-to-peak ratio. Some formulations are more popular than the others and have been used more frequently in the submitted New Drug Application (NDA). These methods are classified into two categories:

i) DELTA models, which use only the active treatment values $\Delta Y_{id(T)}$,$\Delta Y_{id(P)}$.

These models are appropriate when there is no placebo arm in the trial or if there is no desire to make adjustment for the placebo effect.

ii) DOUBLE-DELTA models, which use post-baseline values adjusted for placebo:

$$\Delta\Delta\mu_{(T)} = \Delta\mu_{1(T)} - \Delta\mu_{0(T)} \;\&\; \Delta\Delta\mu_{(P)} = \Delta\mu_{1(P)} - \Delta\mu_{0(P)}$$

In the following the T/P models are presented as population parameters; i.e. based on data from the entire population of hypertensive patients. The corresponding estimators of these T/P are based on samples of size n with similar mathematical expressions, except with N replaced by n.

DELTA Models

The concern is to measure the magnitudes of trough, peak, and T/P ratio adjusted for baseline. In these models, for various reasons, no placebo adjustment is considered. The trough and peak are computed within each arm (i.e. placebo & active).

Three models considered here.

$$R(I) = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta Y_{i1(T)}}{\Delta Y_{i1(P)}}, \qquad (1)$$

$$R(II) = \frac{\Delta \mu_{1(T)}}{\Delta \mu_{1(P)}}, \qquad (2)$$

$$R(III) = \frac{\underset{h}{Min}\left[\Delta \mu_{1(h)}\right]}{\underset{h}{Max}\left[\Delta \mu_{1(h)}\right]}. \qquad (3)$$

The estimators of Models II & III are not unbiased estimators. Model I generates an unbiased estimator. This model is used in PK but it is not popular among the drug sponsors for PD. It can be proved (see Appendix 1) that, in average this ratio is smaller than the true ratio.

DOUBLE-DELTA Models

In DOUBLE-DELTA models the data of both arms, placebo and active treatment, are used. One model is discussed here but there are also other possible models.

$$R(IV) = \frac{\underset{h}{Min}\left[\Delta\Delta \mu_{(h)}\right]}{\underset{h}{Max}\left[\Delta\Delta \mu_{(h)}\right]}. \qquad (4)$$

b) Confidence Interval

Three different methods for constructing confidence intervals (CI) on the T/P are discussed.

Bootstrapping

The bootstrapping, also known as re-sampling, was first used in 1960's. B. Efron, in late 1970's, re-introduced this method and provided the theory to support it[7]. In this paper bootstrapping is employed to estimate the T/P's, the biases of the estimates, the standard errors of the biases, the CI's, and the standard errors of the lower and upper bounds of CI's. The procedure will be discussed in Appendix 2.

Fieller's Method

This method is most often used in the clinical literature. In this method it is assumed that the numerator and the denominator of the trough-to-peak ratio are normally distributed. Therefore a transformation into a random variable with the Snedecor F distribution is possible . This method should be applied only to those ratios where the normality assumptions are satisfied.

Let R= T/P be the parameter of interest and r=t/p be an estimator of R. Also let the numerator and denominator of r be linear functions of normal random variables. The conditional distribution of t-rp given r, is normally distributed.

If r is fixed and E(t-rp)=0, then $(t-rp)^2/\hat{V}(t-rp)$ is distributed as a Snedecor F with 1 and ν degrees of freedom. The variance term in the denominator of the F ratio can be expanded as $\hat{V}(t-rp)=\hat{V}(t)+r^2\hat{V}(p)-2r\hat{cov}(t,p)$. To construct a 1-α confidence interval for R, one can solve the quadratic equation

$$F_{1,\nu;\alpha}=(t-rp)^2/[\hat{V}(t)+r^2\hat{V}(p)-2r\hat{cov}(t,p)],$$

or

$$[p^2-\hat{V}(p)F_{1,\nu;\alpha}]r^2-2[pt-\hat{Cov}(t,p)F_{1,\nu;\alpha}]r+[t^2-\hat{V}(t)F_{1,\nu;\alpha}]=0,$$

to obtain the two roots with respect to r.

The above quadratic equation in a simple form, with obvious notation, can be represented as

$$Ar^2-2Br+C=0.$$

Then a (1-α) 100% lower and upper confidence bounds are:

$$R_l=\frac{B-(B^2-AC)^{1/2}}{A} \quad \& \quad R_u=\frac{B+(B^2-AC)^{1/2}}{A},$$

were $F_{1,\nu;\alpha}$ is the α percentile value from an F table with 1 and ν degrees of freedom; $\hat{Cov}(t,p)$ is an estimator of the covariance of t and p. For detail on the restrictions and the underlying assumptions the reader is referred to references 5 and 9.

Taylor Expansion

This Taylor expansion approach is directly taken from an article by Hubert et.al[8]. Based on this method the variance for the trough-to-peak ratio is calculated as:

$$V(t/p) = (1/p)^2 V[t] + (t/p^2)^2 V[p] + 2(t/p^3) Cov[t,p].$$

and a $(1-\alpha)100\%$ can be constructed as:

$$(t/p - t_{\alpha/2,\nu}\sqrt{V(t/p)},\ \ t/p + t_{\alpha/2,\nu}\sqrt{V(t/p)})$$

where $t_{\nu;\alpha}$ is the α percentile value from a Student t table with ν degrees of freedom.

4. NUMERICAL EXAMPLE

Data from a placebo-controlled HT trial on a sustained release once-a-day capsules given at 8:00 a.m. is used here. We have selected the data of one of the active treatments and placebo arms. There were 60 patients in each arm, but 29 patients from placebo and 29 patients from the active arm participated in ABPM study (58 patients). The ABPM observations were obtained at the baseline and the end of week 8 post treatment. The baseline demographics and other descriptive data of the two arms were comparable; they are not of our main interest and are not displayed here.

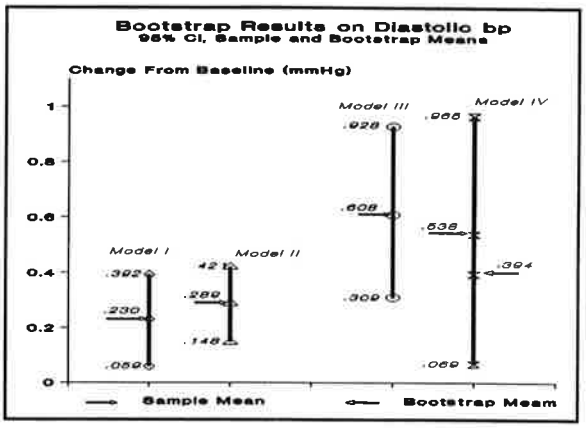
The results of the bootstrap procedure (see Appendix 2) are reported in Table I and also depicted in Figure 1. The results show that all statistics for Models I and II are in close agreement and the same is true for Models II and IV. Models I and II generated smaller ratios, smaller biases (no bias for Model I), smaller SE of ratios, shorter CI's, and smaller SE of lower and upper limits of CI's than their counterparts in Models II and IV. Model IV has the largest bias and the largest 95% CI. From the point estimators (the 2nd column), the ratio estimator of Model I is smaller than that of Model II. This is in agreement with the results derived in Appendix 1.

TABLE I: Bootstrap Results

MODEL		Sample Ratio	Boot-strap Ratio	Bias	SE Boot-strap Ratio	95% Confidence Interval for the Ratio (Adjusted for bias)		
						L	U	U-L
DELTA	I	0.230	0.230	-0.000	0.085	0.059 (.007)	0.392 (.006)	0.333 (.009)
	II	0.289	0.288	-0.001	0.070	0.148 (.007)	0.421 (.005)	0.273 (.008)
	III	0.608	0.594	-0.014	0.156	0.309 (.012)	0.928 (.015)	0.618 (.019)
DOUBLE DELTA	IV	0.538	0.394	-0.144	0.226	0.069 (.027)	0.965 (.032)	.896 (.050)

*The parentheses represent the SE values.

FIGURE 1: Graphical Presentation of the Results in Table I



Comparison of the confidence intervals for Models I and II are presented in Table II and also is depicted in Figure 2.

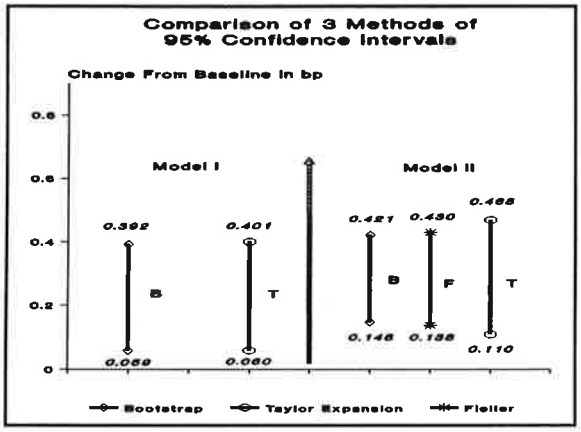
Models I and II fulfill the underlying normality assumption in Fieller's theorem and the Taylor expansion and their CI's are calculated and compared with the bootstrapping method. For Model II, the Taylor expansion has the largest 95% CI; 31% more than bootstrapping and 23% larger than the Fieller's. For Model I the Fieller's theorem was not applicable.

TABLE II: Comparison of CI's

Model	95% Confidence Interval for the True Ratio								
	Bootstrap			Fieller			Taylor Expansion		
	L	U	U-L	L	U	U-L	L	U	U-L
I*	0.059	0.392	0.333	NA	NA	NA	0.060	0.401	0.341
II	0.148	0.421	0.273	0.138	0.430	0.292	0.110	0.468	0.358

*: For Model III the Taylor Expansion confidence interval is the Student t 95% C.I.; L = Lower Limit and U = Upper Limit of the CI.

FIGURE 2: Graphical Presentation of the CI's presented in Table II.



DISCUSSION and CONCLUSION

In this limited work, Models I and II demonstrated good characteristics. Both had the least biases and the shortest CI's. The variations in their lower limits and upper limits are small, therefore the confidence intervals are more robust compared to Models III and IV. The point estimators of the ratios are alarmingly different in Models (I,II) and (III, IV). The latter models produce point estimates that are about twice their counterparts in Model I and II. This might be data-dependent and reflective of the impact of the baseline in this trial. More works need to be done in this area to shed light on the problem.

All three methods of constructing confidence intervals are in close agreement. For Model I the lower bounds range from 0.059 to 0.060 and upper bounds range from 0.331 to 0.341.

For Model II the lower bounds range from 0.110 to 0.148 and upper bounds range from 0.273 to 0.358. The 95% CI of the Taylor expansion included the Fieller's and that included the bootstrap; see Figure 2. Within that small variation the bootstrap method is the best among the three methods of constructing confidence intervals. Obviously the bootstrap procedure is a computer intensive procedure, especially for larger clinical trials. Fieller's confidence intervals are the next best. These confidence intervals should be used only when the underlying assumptions hold[5,9]. The Taylor expansion procedure seems to be more conservative (i.e. wider confidence intervals) among the three methods evaluated.

Although Model II is theoretically biased it performed the best in terms of the point estimate, bias, the CI, and the robustness. The bias produced by this model is negligible. The variation within the lower limit and the upper limit were the least (i.e. more robust). If Model II is chosen the choice of the three methods of constructing CI will not be as crucial. Most often the models used in the NDA submissions are this model along with the Fieller's method. Models III and IV are the least favorable among these four models. In addition to alarmingly large point estimates produced by these two models, they also have the largest variances, the largest biases, the most variation in their limits, and furthermore didn't satisfy the underlying assumptions in the Fieller's theorem.

REFERENCES

1. *Weber M., Cheung D., Graettinger W., Lipson J. (1988), Characterization of Antihypertensive Therapy by Whole-Day Blood Pressure Monitoring. JAMA 259;3281-3285.*

2. *Neutel J., Smith D., Lefkowitz, M., Kazempour M., Weber M. (1993) "Application of Ambulatory Blood Pressure Monitoring in Differentiating Between Antihypertensive Agents." American Journal of Medicine, 49, 181-187.*

3. *White W., Morganroth J. (1989), Usefulness of Ambulatory Monitoring of Blood Pressure in Assessing Antihypertensive Therapy, 63; The Journal of Cardiology 63; 94-98.*

4. *O'Brien E., O'Malley K., Cox J., Stanton A., (1991), Ambulatory Blood Pressure Monitoring in the Evaluation of Drug Efficacy, American Heart Journal 121; 999-1006.*

5. *Finney, D.(1978), Statistical Method in Biological Assay, 3rd ed., Griffin, London.*

6. *Stewart W. and Hafner, K., Statistical Analysis of Trough /Peak Ratios, Drug Information Journal Vol 25, 1991, 405-409.*

7. *Efron B., The Jackknife, the Bootstrap, and Other Re-Sampling Plan, SIAM, 1982.*

8. *Hubert J., Bohidar N., Peace K., Assessment of pharmacological Activity, Biopharmaceutical Statistics for Drug Development, Marcel Deker, New York and Basel.*

9. *Kendall M., Stuart A. (1960), The Advanced Theory of Statistics, Vol. 2, Chapter 20, Hafner Publication, New York.*

Appendix 1

Let τ and ρ be the true trough and peak values. Also let assume that their observed values are t_i and p_i which can be modeled as

$$t_i = \tau + \epsilon_i$$
$$p_i = \rho + \epsilon_i,$$

where $\epsilon_i \sim N(0,\sigma^2)$.

Then R(I) can be written as

$$R(I) = \frac{1}{N}\sum_i^N \frac{t_i}{p_i} = \frac{1}{N}\sum_i^N \frac{\tau+\epsilon_i}{\rho+\epsilon_i}$$
$$= \frac{1}{N}\sum_i^N \left(\frac{\tau}{\rho}+\frac{\epsilon_i}{\rho}\right)\left(1-\frac{\epsilon_i}{\rho}+\frac{\epsilon^2{}_i}{\rho^2}-\frac{\epsilon^3{}_i}{\rho^3}+\ldots\right)$$
$$= \frac{1}{N}\sum_i^N \left[\frac{\tau}{\rho}+\frac{\epsilon_i}{\rho}\left(1-\frac{\tau}{\rho}\right)-\frac{\epsilon_i^2}{\rho^2}\left(1-\frac{\tau}{\rho}\right)+\ldots\right]$$
$$\approx \frac{\tau}{\rho}-\left(1-\frac{\tau}{\rho}\right)\frac{\sigma^2}{\rho^2} .$$

By definition $\tau \le \rho$ or τ/ρ is smaller than 1. Therefore, the second part of the above equation is always positive and the ratio calculated through this approach is always less than the true value, τ/ρ .

Appendix 2

Bootstrapping takes place in three steps:

Step 1: From the set of n subjects, n subjects were re-sampled with replacement and their diastolic blood pressure measures were used to construct the bootstrap t/p ratio r^*.

Step 2: Procedure in Step 1 was repeated 1000 times and 1000 bootstrapped ratios r_1^*, r_2^*,.. ., r_b^*,..., r_{1000}^* were generated. These ratios were used to compute:

The mean ratio as

$$\hat{r}_\cdot^* = 1/1000\sum_{b=1}^{1000} \hat{r}_b^* .$$

The bias, $\hat{Bias} = \hat{r}_\cdot^* - \hat{r}$, in which $\hat{r}$ is the sample mean ratio.

The standard error (SE) of the ratios

$$\hat{SE}\,(\hat{r}^*) = \left[\frac{1}{999}\sum_{j=1}^{1000}\left(\hat{r}_j^* - \hat{r}_\cdot^*\right)^2\right]^{1/2}.$$

The lower (L) and upper limits (U) of the 95% CI's as the 2.5th, the 97.5th percentiles of the sorted 1000 bootstrap ratios.

Step 3: The procedures in Step 2 were repeated 100 times to generate 100 repetition of all the statistics in Step 2. The results were used to compute: the mean and SD's of the 100 L's and U's, the mean of the 100 bootstrap mean ratios $\hat{r}_\cdot^*$, the mean of 100 $\hat{Bias}(\hat{r}^*)$, and the mean of the 100 $\hat{SE}(\hat{r}^*)$.

The values reported in Tables I and II are the results obtained in Step 3 of bootstrapping.