PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2006; **5**: 231–233
Published online in Wiley InterScience (www.interscience.wiley.com).

Letter to the editor

Comment on 'Comparison of Sample Size Formulae for 2×2 Cross-over Designs Applied to Bioequivalence Studies', Siqueira Al, Whitehead A, Todd S and Lucini MM. *Pharmaceutical Statistics* 2005; 4:233–243

In a recent paper published in this journal, Siqueira *et al.* [1] compared a method for exact power calculation based on two non-central *t*-distributions with different approximations commonly used for bioequivalence assessment. The purpose of this letter is to point out that (i) the authors applied a method for exact power calculation which differs from the gold standard solution according to Owen [2], (ii) the approximate formulae have already been published by Hauschke *et al.* [3] and Bristol [4], and (iii) from a pharmacokinetic point of view, the presentation of the results lacks the possibility of a simple interpretation.

The authors assumed a two-period, two-sequence cross-over design and, according to international guidelines [5, 6], the following multiplicative model is assumed for concentration-related characteristics, e.g. the area under the concentration–time curve AUC or the maximum concentration $C_{\max}$. Let sequences and periods be indexed by i and k, i, $k = 1, 2$, respectively, and n_i subjects are randomized to sequence i. Let X_{ijk} denote the pharmacokinetic outcome on the jth subject in the ith sequence during period k; then the following multiplicative model is considered:

$$X_{ijk} = \exp(\mu_h + s_{ij} + \pi_k + e_{ijk})$$

where μ_h is the effect of formulation h, where $h = \mathrm{R}$ if $i = k$ and $h = \mathrm{T}$ if $i \neq k$, π_k is the effect of the kth period. The term s_{ij} denotes the random effect of the jth subject in sequence i and e_{ijk} is the random error term for subject j in period k and sequence i.

Carry-over effects are not included in the model as a factor, because in bioequivalence trials these effects are seldomly seen if a washout phase of 5–6 half-lives is chosen to ensure that all traces of the drug have been removed [7].

Taking logarithms of the pharmacokinetic outcomes transforms the multiplicative model on the original scale to the corresponding additive model on the logarithmic scale:

$$Y_{ijk} = \ln X_{ijk} = \mu_h + s_{ij} + \pi_k + e_{ijk}$$

It is assumed that the subject effects s_{ij} are independent and normally distributed with expected mean 0 and intersubject variance σ_s^2. The intrasubject residuals e_{ijk} are independent and normally distributed with expected mean 0 and variances σ_e^2. Furthermore, the random terms s_{ij} and e_{ijk} are assumed to be mutually independent. Let $\exp(\mu_\mathrm{T})/\exp(\mu_\mathrm{R}) = \exp(\mu_\mathrm{T} - \mu_\mathrm{R})$ be the ratio of the expected mean values of the test and the reference formulation on the original scale, and let the interval (θ_1, θ_2) denote the equivalence range, where $0 < \theta_1 < 1 < \theta_2$. The following test problem concerning equivalence is considered:

$$\mathrm{H}_0: \quad \exp(\mu_\mathrm{T} - \mu_\mathrm{R}) \leqslant \theta_1 \quad \text{or} \quad \exp(\mu_\mathrm{T} - \mu_\mathrm{R}) \geqslant \theta_2$$

$$\mathrm{H}_1: \quad \theta_1 < \exp(\mu_\mathrm{T} - \mu_\mathrm{R}) < \theta_2$$

After logarithmic transformation, the above test problem becomes:

$$\mathrm{H}_0: \quad \mu_\mathrm{T} - \mu_\mathrm{R} \leqslant \ln \theta_1 \quad \text{or} \quad \mu_\mathrm{T} - \mu_\mathrm{R} \geqslant \ln \theta_2$$

$$\mathrm{H}_1: \quad \ln \theta_1 < \mu_\mathrm{T} - \mu_\mathrm{R} < \ln \theta_2$$

According to Schuirmann [8], equivalence can be concluded at nominal level of significance α, if

$$T_1 = \frac{\bar{Y}_\mathrm{T} - \bar{Y}_\mathrm{R} - \ln \theta_1}{\hat{\sigma}_e \sqrt{\frac{1}{2}(1/n_1 + 1/n_2)}} > t_{1-\alpha,\, n_1+n_2-2}$$

and

$$T_2 = \frac{\bar{Y}_\mathrm{T} - \bar{Y}_\mathrm{R} - \ln \theta_2}{\hat{\sigma}_e \sqrt{\frac{1}{2}(1/n_1 + 1/n_2)}} < -t_{1-\alpha,\, n_1+n_2-2}$$

where $t_{1-\alpha, n_1+n_2-2}$ is the $(1-\alpha)$-quantile of the central t-distribution with n_1+n_2-2 degrees of freedom, $\bar{Y}_\mathrm{T}$ and $\bar{Y}_\mathrm{R}$ are the least square means of the test and reference treatment and $\hat{\sigma}_e^2$ is the mean square MS_e from the $ANOVA$ after logarithmic transformation.

Copyright © 2006 John Wiley & Sons, Ltd.

Pharmaceutical STATISTICS

For power and sample size determination, a balanced cross-over design is assumed, that is $n = 2n_1 = 2n_2$. The power of the test procedure is the probability that the null hypothesis H_0 is rejected if the alternative hypothesis H_1 is true:

$$P(T_1 > t_{1-\alpha,n-2} \text{ and } T_2 < -t_{1-\alpha,n-2} \mid \ln\theta_1 < \mu_T - \mu_R < \ln\theta_2, \sigma_e)$$

Owen [2] has shown that the random vector (T_1, T_2) has a bivariate noncentral t-distribution with $n-2$ degrees of freedom and noncentrality parameters

$$\delta_1 = \frac{\mu_T - \mu_R - \ln\theta_1}{\sigma_e\sqrt{2/n}} \quad \text{and} \quad \delta_2 = \frac{\mu_T - \mu_R - \ln\theta_2}{\sigma_e\sqrt{2/n}}$$

and that the above expression for the exact power can be calculated by the difference of two definite integrals, which depends on δ_1 and δ_2.

Instead of calculating the difference of the two definite integrals, Siqueira *et al*. [1] based their power calculations on the cumulative distribution functions of two non-central t-distributions and called it the 'gold standard' solution. However, as already derived by Wang and Chow [9], this approach is only an approximation of the 'gold standard' according to Owen [2].

A further critical point is the graphical presentation of the power curves and the corresponding tabulation of required sample sizes. The focus of their presentation is on the logarithmic scale and not on the original scale. Hence, Siqueira *et al*. [1] have shown the results as a function of the difference $\mu_T - \mu_R$ and $\sigma_d^2 = 0.5\sigma_e^2$. From a pharmacokinetic viewpoint, an interpretation of these values on the logarithmic scale might be difficult and therefore the corresponding values should be presented on the original scale. For example, the value $\mu_T - \mu_R = 0.10$ for the underlying pharmacokinetic characteristic AUC refers to $\exp(0.10) = 1.105$ and the interpretation is that there is an increase in absorption of about 11%.

Furthermore, it is more convenient to express the variability in terms of the intrasubject coefficient of variation on the multiplicative scale [10], that is $CV_e = \sqrt{\exp(\sigma_e^2) - 1}$. This is illustrated in Figure 1 which gives an impression of the attained exact power curves for commonly used sample sizes assuming a CV_e of 20% and the (0.8, 1.25) equivalence range.

The calculation of the exact power requires either the direct evaluation of the bivariate noncentral t-distribution or the calculation of the difference of two definite integrals, which might not be accessible to practitioners. For that reason, Siqueira *et al*. [1] present approximate formulae for sample size determination. However, it should be noted that these formulae have already been provided by Hauschke *et al*. [3] using the central t-distribution and by Bristol [4] using the normal distribution.

In summary, Siqueira *et al*. [1] compared different approximate approaches with a method which is solely an approximation of the 'gold standard' method provided by Owen [2]. Furthermore, the approximation formulae and the comparisons of the exact and approximate sample sizes have been published already in the 1990s [3,4]. Finally, it is of utmost importance for a straightforward interpretation that the statistical presentation of the results should be performed on the original scale.

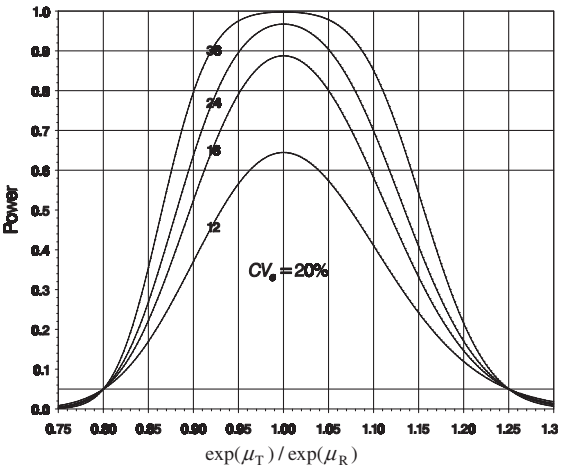

Figure 1. Probability of correctly concluding bioequivalence (power) as a function of the ratios $\exp(\mu_T)/\exp(\mu_R)$ from the alternative (0.8, 1.25); power curves refer to a total sample size of n = 12, 18, 24, and 36 subjects, α = 5% and CV_e = 20%.

REFERENCES

1. Siqueira AL, Whitehead A, Todd S, Lucini MM. Comparison of sample size formulae for 2×2 cross-over designs applied to bioequivalence studies. *Pharmaceutical Statistics* 2005; **4**:233–243.
2. Owen DB. A special case of a bivariate non-central t-distribution. *Biometrika* 1965; **52**:437–446.
3. Hauschke D, Steinijans VW, Diletti E, Burke M. Sample size determination for bioequivalence assessment using a multiplicative model. *Journal of Pharmacokinetics and Biopharmaceutics* 1992; **20**:557–561.
4. Bristol DR. Probabilities and sample sizes for the two one-sided tests procedure. *Communications in Theory and Methods* 1993; **22**:1953–1961.
5. Committee for Proprietary Medicinal Products 2001. *Note for guidance on the investigation of bioavailability and bioequivalence*. EMEA: London.
6. Food and Drug Administration 2001. *Guidance for industry. Statistical approaches to establishing bioequivalence*. Center for Drug Evaluation and Research: Rockville, USA.
7. Senn S, D'Angelo G, Potvin D. Carry-over in cross-over trials in bioequivalence: theoretical concerns and empirical evidence. *Pharmaceutical Statistics* 2004; **3**:133–142.

Copyright © 2006 John Wiley & Sons, Ltd.

8. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 1987; **15**:657–680.
9. Wang H, Chow S-C. On statistical power for average bioequivalence testing under replicated crossover designs. *Journal of Biopharmaceutical Statistics* 2002; **12**:295–309.
10. Hauschke D, Steinijans VW, Diletti E, Schall R, Luus HG, Elze M, Blume H. Presentation of the intrasubject coefficient of variation for sample size planning in bioequivalence studies. *International Journal of Clinical Pharmacology and Therapeutics* 1994; **32**:376–378.

Dieter Hauschke and Dietrich Knoerzer
Department of Biometry,
ALTANA Pharma, Byk-Gulden-Str. 2,
78467 Konstanz, Germany
E-mail: dieter.hauschke@altanapharma.com
(DOI: 10.1002/pst.211)

Rejoinder

Hauschke and Knoerzer correctly point out that what we have called the 'gold standard' method is an approximation to that presented by Owen. Our use of this approximation should not affect the results in Tables V and VI of our paper. Comparisons with empirical power calculations based on 1 000 000 simulations of the two one-sided tests and with nQuery Advisor which states that it uses Owen's approach, showed that the power calculations in Figure 1 of our paper were accurate to the third decimal place for $n = 12$ and 24. For $n = 6$, the same accuracy is maintained except in situations where the power falls below about 0.5. Since a power of at least 0.8 is usually required in a bioequivalence study, the approximate gold standard method and Owen's method will be equivalent in practical terms.

The purpose of our paper was to compare the sample size formulae which had already appeared in the literature. We presented a more extensive comparison than has been previously published, as it is based on nine different formulae. A variety of ways of writing the sample size formula for a bioequivalence trial have appeared in the literature, and we have selected a different set of parameters from those proposed by Hauschke and Knoerzer. Relationships exist between the different parameterizations, and those performing the calculations will want to choose ones appropriate for their particular situation. It was not our intention to provide a comprehensive guideline.

Arminda Lucia Siqueira
Anne Whitehead
Susan Todd
Maria M Lucini

Copyright © 2006 John Wiley & Sons, Ltd.