The Validity of the Test of Individual Equivalence Ratios

J. T. Gene Hwang; Weizhen Wang

Biometrika, Vol. 84, No. 4. (Dec., 1997), pp. 893-900.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199712%2984%3A4%3C893%3ATVOTTO%3E2.0.CO%3B2-V

Biometrika is currently published by Biometrika Trust.

Biometrika (1997), **84**, 4, *pp.* 893–900
Printed in Great Britain

The validity of the test of individual equivalence ratios

BY J. T. GENE HWANG
Department of Mathematics, Cornell University, Ithaca, New York 14583, U.S.A.
e-mail: hwang@math.cornell.edu

AND WEIZHEN WANG
Department of Mathematics and Statistics, Wright State University, Dayton, Ohio 45435, U.S.A.
e-mail: wwang@euler.math.wright.edu

SUMMARY

We evaluate the validity of the test of individual equivalence ratios (TIER), a term coined by Anderson & Hauck (1990). The test was also proposed in Wellek (1989). It is proved to be a valid test under a class of unimodal symmetric distributions; this includes normal distributions. Since most bioequivalence studies involve data which are normal, or at least unimodal and symmetric, the test is typically valid. It is, however, not always valid, as shown by counter examples.

Some key words: Crossover design; Symmetric distribution; Unimodal distribution.

1. INTRODUCTION

In bioequivalence studies, the goal is to demonstrate that a tested (T) drug has characteristics similar to a reference (R) drug. Typically, these two treatments are applied to patients and blood samples are compared in terms of bioavailabilities which are characterised by pharmacokinetic variables such as the area under the concentration versus time curve, the maximum concentration, and the time when the peak concentration is achieved. Since the cost of seeking approval from the U.S. Food and Drug Administration, for example, of a new drug through bioequivalence is less than one percent of what is needed to develop an original drug, the bioequivalence approach has attracted enormous commercial attention.

There are three definitions relating to bioequivalence: bioequivalence in average, bioequivalence in distribution and individual bioequivalence. The first two definitions focus only on the equivalence of the averages or the distributions of the two treatments. The last definition, however, refers to equivalence within an individual subject. Individual bioequivalence, unlike the other types of bioequivalence, could guarantee switchability; i.e. a patient could switch between the two treatments. This idea of individual bioequivalence was popularised in a very influential paper by Anderson & Hauck (1990). See also Anderson (1993) for interesting pictorial explanations of how individual bioequivalence differs from the other types.

Although current U.S. Food and Drug Administration guidance (FDA, 1992) is based on bioequivalence in average, many discussions have pointed toward the more appealing notion of individual bioequivalence. In this paper, we discuss the test of individual equival-

ence ratios developed in Anderson & Hauck (1990) and also proposed in Wellek (1989, 1993).

To describe the test, let X_{ij} be the measured bioavailability corresponding to the ith subject, $i = 1, 2, \ldots, n$, and jth formulation, $j = T, R$. It is assumed that $Y_{ij} = \log X_{ij}$ satisfies

$$Y_{ij} = F_j + U_{ij} + e_{ij}, \tag{1·1}$$

where F_j is the population average of the jth formulation, U_{ij} is the deviation of the ith individual effect from the population average for the jth formulation, and e_{ij} is the within-subject error. It is assumed that $\{e_{ij}\}$ are independent of $\{U_{ij}\}$. However, U_{iT} and U_{iR} may be correlated. It is further assumed that, for each j, U_{ij} $(1 \leqslant i \leqslant n)$ are independently and identically distributed.

Let

$$Y_i = Y_{iT} - Y_{iR} = m_i + \varepsilon_i, \tag{1·2}$$

where

$$m_i = F_T - F_R + U_{iT} - U_{iR}, \quad \varepsilon_i = e_{iT} - e_{iR},$$

Anderson & Hauck (1990) and Welleck (1989) considered testing the hypotheses

$$H_0 : P_0 \leqslant P_{\min} \quad \text{versus} \quad H_1 : P_0 > P_{\min}, \tag{1·3}$$

where $P_0 = \text{pr}(|m_i| < \Delta)$ and Δ and $P_{\min}$ are prespecified quantities. As an example, the regulatory agency may specify that $\Delta = \log(1{\cdot}25)$ and $P_{\min} = 0{\cdot}8$. This corresponds to a definition of bioequivalence in which at least 80% of the subjects' bioavailabilities on new formulation in the original scale, that is the antilog of $F_T + U_{iT}$, are within 25% of that of the standard formulation.

The test of individual equivalence ratios is to use the statistic

$$X = \text{number of } i\text{'s such that } |Y_i| < \Delta \tag{1·4}$$

and consider the p-vale $\text{pr}(X \geqslant x)$, where x is the realisation of (1·4) and X has a binomial $B(n, P_{\min})$ distribution. The individual bioequivalence is declared if the p-value is less than α, typically taken to be 0·05.

There are, however, criticisms about the test. We shall say that a test is α valid or valid if the type I error is no greater than a prespecified level α. Note that the test of individual equivalence ratios is obviously valid for testing against

$$H_0^W : P_W \leqslant P_{\min}, \tag{1·5}$$

where $P_W = \text{pr}(|Y_i| < \Delta)$. Here the subscript W refers to the inclusion of the within error ε_i in the probability evaluation. However, it may not be valid for testing against H_0 in (1·3). Therefore, the validity of the test of individual equivalence ratios is in question. Although Schall & Luus (1993, p. 118) stated that P_W is in general smaller than P_0 and claimed the test to be valid in general, no specific condition had been established which implied the validity of the test of individual equivalence ratios. In this paper, we show that the test is valid provided that $P_{\min} \geqslant \frac{1}{2}$, that m_i and ε_i are symmetric, and that m_i has a unimodal distribution; see Theorem 1. It is, however, not true that the test is always valid; we give some counter examples in § 3. These counter examples, as implied by Theorem 1, have to involve nonunimodal distributions. In bioequivalence studies, such distributions are rare, and hence the validity of the test of individual equivalence ratios is the norm.

Our Theorem 1 has several other applications. In a related approach, Liu & Chow (1997) use normal assumptions to derive different procedures. The null hypothesis they consider is different from H_0 and is similar to H_0^W. They showed that their test is valid for their hypothesis. To the present authors, the hypothesis H_0 appears more appropriate since the measurement error, not included in H_0, should be irrelevant. An application of Theorem 1, however, shows that Liu & Chow's test is also valid for H_0 under their normal assumptions. Similar 'approximate' results can be established for the nearly unbiased tests proposed by W. Wang in his 1995 Ph.D. dissertation and in a Cornell technical report by W. Wang and J. T. G. Hwang.

Although the test of individual equivalence ratios has been shown to be typically valid, it does not mean that we endorse the use of it. As pointed out by an associate editor, the test of individual equivalence ratios has other problems including the three discussed below. First, as with virtually any test, the subjects are typically healthy; there is doubt as to whether or not the inference can be applied to patients, and the only way out may be to experiment with the drugs on the patients. Secondly, the test of individual equivalence ratios dichotomises the data and the test may be too inefficient. Apparently, this prompted the studies of Liu & Chow (1997) and the technical report by Wang and Hwang mentioned above. The latter study provides a nearly unbiased test that improves greatly upon the test of individual equivalence ratios. Finally, the test of individual equivalence ratios applies only to a 2×2 crossover design. For such a design, it is not possible to separate m_i and ε_i. Therefore it seems better to use a higher-order design such as a four-period crossover design with subjects allocated at random in each pair of periods to receive either the sequence TR or RT. One theoretically interesting aspect of the present paper, however, is that, even though m_i cannot be separated from ε_i, nontrivial tests exist for a hypothesis involving only m_i.

2. A GENERAL THEOREM OF VALIDITY

We shall show that under fairly general assumptions a test that is valid for $H_0^W : P_W \leqslant P_{\min}$ is valid for

$$H_0 : P_0 \leqslant P_{\min}. \tag{2·1}$$

A sufficient condition is that

$$H_0 \text{ implies } H_0^W. \tag{2·2}$$

We shall say that a random variable X has a symmetric distribution about its mean μ if $X - \mu$ and $-(X - \mu)$ have the same distribution. Also a random variable Z is said to have a unimodal distribution at a if its probability density function is nonincreasing for $z \geqslant a$ and is nondecreasing for $z \leqslant a$.

Now we may state the theorem.

THEOREM 1. *Assume model* (1·2) *where* m_i *and* ε_i *are independent. Suppose that the distribution of* ε_i *is symmetric about zero and* m_i *has a symmetric unimodal probability density function at a. If* $P_{\min} \geqslant \frac{1}{2}$, *then* (2·2) *holds. Consequently, any valid test for* H_0^W *is valid for* H_0.

Since the test of individual equivalence ratios is valid for H_0^W, the theorem obviously implies that the test of individual equivalence ratios is valid under the assumption in Theorem 1. In particular, it is valid under the special case where m_i and ε_i are normally distributed and ε_i has a zero mean. Note that measured bioavailability, after the logarithm

transformation, often appears normal, or at least symmetric and unimodal. Hence the test of individual equivalence ratios is typically valid.

When the assumptions of Theorem 1 fail, the conclusion is false. The next section includes counterexamples in which the test of individual equivalence ratios fails catastrophically. However, distributions of m and ε like those assumed for these counterexamples do not seem to arise in bioequivalence studies.

Finally we demonstrate the relevance of Theorem 1 to the other tests. Liu and Chow's test applies to the following model as well as to more complicated ones:

$$Y_{ij} = F_j + S_i + \varepsilon_{ij}, \tag{2·3}$$

where F_j, for $j = T$ or R, is the fixed formulation effect and S_i is the effect of the ith subject and $i = 1, \ldots, n$. They test

$$H_0 : P_{\mathrm{LC}} \leqslant P_{\min} \quad \text{versus} \quad H_1 : P_{\mathrm{LC}} > P_{\min}, \tag{2·4}$$

where $P_{\mathrm{LC}} = \mathrm{pr}(|Y_{iT} - Y_{iR}| < \Delta)$. It is assumed that S_i and ε_{ij} are independently normally distributed with common mean zero and variances σ_S^2 and σ_j^2, respectively.

To describe their procedure, let

$$\bar{Y} = (\sum Y_i)/n, \quad S^2 = \sum (Y_i - \bar{Y})^2/\{(n-1)n\}.$$

Note that S^2 is an unbiased estimator for the variance of $\bar{Y}$. Their procedure rejects H_0 if

$$|\bar{Y}| < \Delta - S t_{\alpha,v}(n^{\frac{1}{2}} z_{P^*}), \tag{2·5}$$

where $t_{\alpha,v}(\eta)$ is the α upper critical value of a noncentral t distribution with $v = n - 1$ degrees of freedom and noncentrality η, and z_{P^*} is the $P^* = \frac{1}{2}(1 - P_{\min})$ upper critical value of a standard normal distribution.

If one attempts to understand model (2·3) from Anderson & Hauck's (1990) viewpoint, the error term ε_{ij} in (2·3) can be considered to be the sum of two errors,

$$\varepsilon_{ij} = d_{ij} + e_{ij}, \tag{2·6}$$

where e_{ij} is the error incurred in measuring the bioavailabilities and d_{ij} is the deviation of the ith individual subject effect from S_i. Substituting ε_{ij} in (2·3) by $d_{ij} + e_{ij}$, we may write model (2·3) as (1·1) with

$$U_{ij} = S_i + d_{ij}. \tag{2·7}$$

Note that the U_{ij}'s defined here satisfy the assumptions of the U_{ij}'s described in the paragraph after (1·1).

Anderson & Hauck's hypothesis (1·3), however, is slightly different from Liu & Chow's hypothesis (2·4), since, in (2·4), the probability is evaluated with respect to e_{ij} as well, whereas P_0 defined right after (1·3) involves no e_{ij}, which seems more reasonable. Liu & Chow have demonstrated that their test (2·5) is valid for testing (2·4). However, is it valid for (1·3)? The answer is yes, and is stated in the following corollary, which follows directly from Theorem 1.

COROLLARY 1. *Let U_{ij} be as defined in* (2·7) *and m_i and ε_i be defined as in the paragraph containing* (1·2). *Suppose that m_i and ε_i are independent and are normally distributed where the mean of ε_i is zero. Then the test with the rejection region* (2·5) *is valid for testing* (1·3) *if $P_{\min} \geqslant \frac{1}{2}$.*

A similar comment applies to the nearly unbiased tests in Chapter 7 of Wang's Ph.D.

thesis and the technical report of Wang and Hwang mentioned in § 1. These tests are much more efficient in power than those of Anderson & Hauck (1990) and Liu & Chow (1997). These tests are approximately valid and hence we can only conclude that they are approximately valid for Anderson & Hauck's problem.

3. Counterexamples

In this section we exhibit some examples in which the test of individual equivalence ratios fails to be valid when the assumptions of Theorem 1 are violated.

Example 1. Assume that m and ε are statistically independent and have the following probability distributions: m takes the two values $-0{\cdot}55$ and $1{\cdot}05$ with probabilities $0{\cdot}8$ and $0{\cdot}2$; ε takes the two values $-0{\cdot}15$ and $1{\cdot}35$ with probabilities $0{\cdot}9$ and $0{\cdot}1$. Note that $E(\varepsilon) = 0$. Hence $Y = m + \varepsilon$ takes the four values $-0{\cdot}70, 0{\cdot}80, 0{\cdot}90$ and $2{\cdot}40$ with probabilities $0{\cdot}72, 0{\cdot}08, 0{\cdot}18$ and $0{\cdot}02$.

We shall use Δ as a yardstick to measure m, ε and Y. This is equivalent to assuming that all the variables have been divided by Δ, and hence we may assume from now on that $\Delta = 1$ and $P_{\min} = 0{\cdot}8$. The relevant probabilities are $P_0 = \mathrm{pr}(|m| < 1) = 0{\cdot}8$ and $P_W = \mathrm{pr}(|Y| < 1) = 0{\cdot}98$. Note that $P_W > P_0$, which contradicts (A·6), and so the conclusion of Theorem 1 fails. More specifically, we consider $n = 24$ and $x = 23$. Based on the binomial distribution with the incorrect $p = 0{\cdot}8$, the p-value suggested by the test of individual equivalence ratios is $\mathrm{pr}(X \geqslant 23) = 0{\cdot}0331$. On the other hand, the true distribution of X corresponds to $p = 0{\cdot}98$, which leads to the p-value $\mathrm{pr}(X \geqslant 23) = 0{\cdot}917$. Hence the test of individual equivalence ratios underestimates the p-value drastically.

In Example 1, the distributions of m and ε were asymmetrical. However, we could have symmetric distributions which fail Theorem 1, as demonstrated below in Example 2. Obviously, in Theorem 1, the distribution of m cannot be unimodal. The assumption of unimodality therefore seems crucial.

Example 2. Assume that the probability distributions of two independent random variables m and ε are as follows: m takes three values $-0{\cdot}25, 0{\cdot}45$ and $1{\cdot}15$ with probabilities $0{\cdot}2, 0{\cdot}6$ and $0{\cdot}2$; and ε takes the two values $-0{\cdot}3$ and $0{\cdot}3$ with probability $0{\cdot}5$ for each. Hence $Y = m + \varepsilon$ takes the six values $-0{\cdot}55, 0{\cdot}05, 0{\cdot}15, 0{\cdot}75, 0{\cdot}85$ and $1{\cdot}45$ with probabilities $0{\cdot}1, 0{\cdot}1, 0{\cdot}3, 0{\cdot}3, 0{\cdot}1$ and $0{\cdot}1$. Note that $\mathrm{pr}(|m| < 1) = 0{\cdot}8 < \mathrm{pr}(|Y| < 1) = 0{\cdot}9$. For $n = 24$, and $x = 23$, the p-value as suggested by the test of individual equivalence ratios is $0{\cdot}0331$, whereas the correct p-value is $0{\cdot}293$.

Example 3 (*the continuous modification of Example* 1). Assume that m and ε are independent, with the normal mixture densities

$$0{\cdot}76N(x;\ -0{\cdot}55, 0{\cdot}05) + 0{\cdot}24N(x;\ 1{\cdot}05, 0{\cdot}05), \quad 0{\cdot}9N(x;\ -0{\cdot}15, 0{\cdot}05) + 0{\cdot}1N(x;\ 1{\cdot}35, 0{\cdot}05),$$

respectively, where $N(x;\ \mu, \sigma)$ denotes the probability density function of a normal distribution with mean μ and standard deviation σ. Hence Y has the distribution

$$0{\cdot}684N(-0{\cdot}70, 0{\cdot}05\sqrt{2}) + 0{\cdot}076N(0{\cdot}8, 0{\cdot}05\sqrt{2}) + 0{\cdot}216N(0{\cdot}9, 0{\cdot}05\sqrt{2}) + 0{\cdot}024N(2{\cdot}4, 0{\cdot}05\sqrt{2}).$$

The probability density functions of m and Y exhibit clear multimodality, as revealed by comparing the distances between the mixture component means with the component

standard deviation. Some calculations give

$$\text{pr}(|m|<1)=0{\cdot}8 \tag{3·1}$$

and $\text{pr}(|Y|<1)=0{\cdot}959$. Note that Example 3 uses slightly different probabilities from Example 1 in order that (3·1) is satisfied. For $n=24$ and $x=23$ the correct p-value would be 0·74, corresponding to the binomial distribution with $p=0{\cdot}959$, whereas the p-value suggested by the test of individual equivalence ratios is 0·0331, using $p=0{\cdot}8$.

ACKNOWLEDGEMENT

The authors wish to thank the associate editor and the editor for their comments which are included in the paper and which help to balance the interpretation of the results in this paper. They also thank Professor Graeme Bailey for editing through the paper.

APPENDIX

Proof of Theorem 1

We shall drop the subscript i in ε_i and m_i. Consider

$$\text{pr}(|m+t\varepsilon|<\Delta), \tag{A·1}$$

where $0\leqslant t\leqslant 1$ is a parameter we introduce. The crucial part of the proof involves showing that the probability decreases in t if $\Delta\geqslant|a|$, which implies

$$\text{pr}(|m|<\Delta)\geqslant\text{pr}(|m+\varepsilon|<\Delta)\quad(\Delta\geqslant|a|). \tag{A·2}$$

When $a=0$, (A·2) requires only the assumption that ε be independent of m, by a theorem in Anderson (1955). However, when $a\neq 0$, (A·2) holds for $\Delta\geqslant|a|$ only, and may fail for $\Delta<|a|$ even under the assumptions of the theorem. Hence the result is different from the assertion that m is more peaked about zero than $m+\varepsilon$. See, for example, Birnbaum (1948) for the definition of peakedness.

LEMMA A·1. *Let f denote the probability density function of m with respect to Lebesgue measure and let G denote the cumulative distribution function of ε. Then*

$$\frac{d}{dt}\text{pr}(|m+t\varepsilon|<\Delta)=-\int_{-\infty}^{\infty}\varepsilon\{f(\Delta-t\varepsilon)-f(-\Delta-t\varepsilon)\}\,dG(\varepsilon). \tag{A·3}$$

Furthermore, if the distribution of ε is symmetric about zero, then (A·3) *becomes*

$$-\int_{0}^{\infty}\varepsilon\{f(\Delta-t\varepsilon)+f(-\Delta+t\varepsilon)-f(\Delta+t\varepsilon)-f(-\Delta-t\varepsilon)\}\,dG(\varepsilon). \tag{A·4}$$

Proof. Write (A·1) as

$$\int\int_{-\Delta-t\varepsilon}^{\Delta-t\varepsilon}f(m)\,dm\,dG(\varepsilon).$$

Its derivative with respect to t equals

$$\int\varepsilon\{-f(\Delta-t\varepsilon)+f(-\Delta-t\varepsilon)\}\,dG(\varepsilon),$$

which establishes (A·3).

To prove (A·4), write the right-hand side of (A·3) as

$$-\left(\int_{0}^{\infty}+\int_{-\infty}^{0}\right)\varepsilon\{f(\Delta-t\varepsilon)-f(\Delta-t\varepsilon)\}\,dG(\varepsilon).$$

Applying a change of variables to the second integral and using the symmetry of the distribution of ε and the fact that the integrand is zero at $\varepsilon = 0$, we establish (A·4). □

LEMMA A·2. *Assume that m has a symmetric unimodal probability density function at a. If*

$$|a| \leqslant \Delta \tag{A·5}$$

then

$$\text{pr}(|m+\varepsilon| < \Delta) \leqslant \text{pr}(|m| < \Delta). \tag{A·6}$$

Proof. Let f_0 denote the probability density function of $m-a$ and hence $f(t) = f_0(t-a)$. Note that $f_0(t)$ is a nonincreasing function of $|t|$. To determine the sign in (A·4), note that

$$f(\Delta - t\varepsilon) + f(-\Delta + t\varepsilon) - f(\Delta + t\varepsilon) - f(-\Delta - t\varepsilon) = f_0(\Delta - t\varepsilon - a) - f_0(\Delta + t\varepsilon - a) + f_0(-\Delta + t\varepsilon - a) - f_0(-\Delta - t\varepsilon - a). \tag{A·7}$$

By (A·5), for $\varepsilon > 0$ and $t > 0$,

$$|\Delta - t\varepsilon - a| \leqslant |\Delta + t\varepsilon - a|, \quad |-\Delta + t\varepsilon - a| \leqslant |-\Delta - t\varepsilon - a|.$$

Thus (A·7) is nonnegative and hence (A·4) and (A·3) are nonpositive, which, in turn, implies the conclusion of this lemma. □

Now we return to the proof of Theorem 1. All we need to show is that (2·2) holds. To do so, consider two cases: $|a| \leqslant \Delta$ and $|a| > \Delta$. For the first case, Lemma A·2 gives (2·2). For the second case, where $|a| > \Delta$, we shall show that H_0^W holds, which completes the proof. Since $P_{\min} \geqslant \frac{1}{2}$, we see that $\text{pr}(|m+\varepsilon| < \Delta) \leqslant \frac{1}{2}$, which implies H_0^W. This last inequality is obvious, since $m+\varepsilon$ has a distribution symmetric with respect to a, which is outside the interval $[-\Delta, \Delta]$.

Proof of Corollary 1. Liu & Chow's test is based on the difference

$$Y_{iT} - Y_{iR} = F_T - F_R + \varepsilon_{iT} - \varepsilon_{iR}. \tag{A·8}$$

The true model we are considering is model (1·1), which gives

$$Y_{iT} - Y_{iR} = F_T - F_R + U_{iT} - U_{iR} + e_{iT} - e_{iR}. \tag{A·9}$$

Note that Liu & Chow's test is valid for testing $P_{\text{LC}} \leqslant P_{\min}$ and hence it should be valid for testing the same hypothesis where $Y_{iT} - Y_{iR}$ is replaced by (A·9). This is because we may redefine $U_{iT} - U_{iR} + e_{iT} - e_{iR}$ in (A·9) as $\varepsilon_{iT} - \varepsilon_{iR}$ in (A·8). Now, under model (A·9) or model (1·1), we consider the hypothesis (1·3). By Theorem 1, Liu & Chow's test is valid.

REFERENCES

ANDERSON, S. (1993). Measuring switchability and prescribability: When is average bioequivalence sufficient? *J. Pharmacokin. Biopharm.* **22**, 551–64.

ANDERSON, S. & HAUCK, W. W. (1990). Consideration of individual bioequivalence. *J. Pharmacokin. Biopharm.* **18**, 259–73.

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Am. Math. Soc.* **6**, 170–6.

BIRNBAUM, Z. W. (1948). On random variables with comparable peakedness *Ann. Math. Statist.* **19**, 76–81.

FDA (1992). *FDA Guidance on Statistical Procedures for Bioequivalence Studies using a Standard Two-Treatment Crossover Design.* Rockville, MD: Division of Bioequivalence, Office of Generic Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration.

LIU, J. P. & CHOW, S. C. (1997). A two one-sided tests procedure for assessment of individual bioequivalence. *Statist. Med.* To appear.

SCHALL, R. & LUUS, H. G. (1993). On population and individual bioequivalence. *Statist. Med.* **12**, 1109–24.

WELLEK, S. (1989). Vorschläge zur Reformulierung der Statistischen Definition von Bioäquivalenz. In *Medizinische Informatik und Statistik*, Ed. G. Giani and R. Repges, pp. 95–9. Berlin, Heidelberg: Springer-Verlag.

WELLEK, S. (1993). Basing the analysis of comparative bioavailability trials on an individualized statistical definition of equivalence. *Biom. J.* **35**, 47–55.

[*Received January* 1996. *Revised April* 1997}