Journal of Pharmacokinetics and Biopharmaceutics, Vol. 18, No. 3, 1990

Consideration of Individual Bioequivalence

Sharon Anderson[1,2] and Walter W. Hauck[2,3]

Received May 16, 1989—Final January 30, 1990

Current procedures for assessing the bioequivalence of two formulations are based on the concept of average bioequivalence. That is, they assess whether the average responses between individuals on the two formulations are similar. Average bioequivalence, however, is not sufficient to guarantee that an individual patient could be expected to respond similarly to the two formulations. To have reasonable assurance that an individual patient could be switched from a therapeutically successful formulation to a different formulation (e.g., a generic substitute) requires a different notion of bioequivalence, which we refer to as individual (or within-subject) bioequivalence. We propose a simple, valid statistical procedure for assessing individual bioequivalence. The decision rule, TIER (Test of Individual Equivalence Ratios), requires the specification of the minimum proportion of subjects in the applicable population for which the two formulations being tested must be bioequivalent (a regulatory decision). The TIER rule is summarized in terms of the minimum number of subjects with bioavailability ratios falling within the specified equivalence interval necessary to be able to claim bioequivalence for given sample size and Type I (α) error. We recommend that the corresponding lower bounds (one-sided confidence intervals) for the proportion of bioequivalent subjects be calculated. TIER is partly motivated by the U.S. FDA's 75/75 Rule (at least 75% of the individual subject bioavailability ratios must be within 75–125%). TIER retains the sensible idea of considering the individual ratios but, unlike the 75/75 rule, is a statistically valid procedure.

KEY WORDS: bioequivalence; binomial tests; 75/75 rule.

INTRODUCTION

If one were to ask the "man (or woman) on the street" what they believe is meant by approval of a generic drug as equivalent to a name-brand product, the answer is likely to be something like "it doesn't matter which I take." While this may be a somewhat naive notion of what we refer to technically as bioequivalent formulations, it is still a very sensible notion.

[1]Syntex Research, 3401 Hillview Avenue, P.O. Box 10850, Palo Alto, California 94303.
[2]Department of Epidemiology and Biostatistics, University of California, San Francisco, California 94143-0560.
[3]To whom correspondence should be addressed.

0090-466X/90/0600-0259S06.00/0

However, it is not what is guaranteed by current procedures for assessing comparative bioavailability; current procedures assess equivalence in terms of average bioavailabilities without directly addressing within-subject equivalence.

In this paper, we first introduce a statistical model that allows us to define more precisely the two notions of bioequivalence: individual (within-subject) and average bioequivalence. It is our premise that we ought to address individual bioequivalence. The first question is whether procedures for addressing average bioequivalence are sufficient for concluding individual bioequivalence. The answer is "maybe," under certain conditions. However, the average bioequivalence procedures themselves do not provide sufficient information to tell whether or not the necessary conditions hold. We thus introduce TIER, Test of Individual Equivalence Ratios, as a proper statistical procedure for assessing individual bioequivalence and provide a simple table for applying TIER. We then contrast our rule to the U.S. FDA's 75/75 Rule for the cases considered, to further demonstrate the lack of validity of the 75/75 rule. After giving an example, we return, in the Discussion, to the conceptual issues, namely, the distinction between individual bioequivalence and average or, more generally, population bioequivalence, and finally to consideration of some wider issues raised by this work.

In what follows, we have imposed some restrictions on the nature of the problem we consider. First, we assume that interest is in a single measure of bioavailability, presumably, but not necessarily, the area under the time-concentration curve (AUC). We do not deal with multiple comparison problems that are common to bioequivalence assessment, regardless of which approach one is following. Second, we assume two formulations, a new (N) and a standard (S). Third, we assume that the design of the comparative bioavailability study is a two-period crossover with no period or differential carry-over effects.

We wish to acknowledge that some of the ideas that are the basis for this work may be found in various articles of Westlake. We note particularly his 1979 and 1988 papers (1,2).

TWO TYPES OF BIOEQUIVALENCE

To formally introduce the two notions of bioequivalence, we first need some notation and a statistical model. The statistical model is one that permits identification of the averages that are assessed by current standard procedures for assessing bioequivalence while allowing individuals to differ in their responses to the two formulations. Let X_{ij} be the measured bioavailability of the jth formulation ($j = \mathrm{N}, \mathrm{S}$) in the ith subject ($i = 1, \ldots, n$). We

assume that the distributions of the X_{ij} are such that

$$\text{mean}\,[\ln(X_{ij})] = \mu_{ij}$$
$$\text{and} \qquad \text{variance}\,[\ln(X_{ij})] = \sigma_W^2 \qquad (1)$$

where we use the subscript W to indicate the within-subject variance. We further assume that, conditional on the ith subject, X_{iN} and X_{iS} are independent. We could allow σ_W^2 to depend on j, but have not done so to simplify the presentation. The key notion in Eq. (1) is that each individual has their own average responses, μ_{iN} and μ_{iS}, to the two formulations. The next step is to relate the individual averages. That is done by introducing a distribution for the individual averages. Specifically, we assume

$$\text{mean}\,(\mu_{ij}) = \mu_j$$
$$\text{variance}\,(\mu_{ij}) = \sigma_B^2 \qquad (2)$$
$$\text{and} \qquad \text{correlation}\,(\mu_{iN}, \mu_{iS}) = \rho$$

μ_N and μ_S are the population average (log) bioavailabilities. As with σ_W^2, we could let σ_B^2, the between-subject variance, depend on j but have not done so for simplicity of presentation. The correlation, ρ, represents the degree to which subjects' expected responses to the two formulations are related. Finally, let τ's represent the means in the original scale, where τ would be subscripted as μ. We can now present the different notions of bioequivalence. For each, we first give a definition in words and then express it as a statistical hypothesis.

Average Bioequivalence

Two formulations are average bioequivalent when the bioavailability of the new formulation, averaged over some appropriate population, is "sufficiently close" to the average of the standard. The corresponding statistical hypothesis is

$$1 - R_A \leq \frac{\tau_N}{\tau_S} \leq 1 + R_A \qquad (3)$$

This is the type of bioequivalence that is currently assessed. Typically, $R_A = 0.2$, corresponding to the common bioequivalence interval of (80%, 120%). Average bioequivalence is a special case of population bioequivalence; that is, bioequivalence defined in terms of similarity of the distributions in the population of responses to the two formulations.

Individual Bioequivalence

Two formulations are individual bioequivalent if the bioavailability of the new formulation is "sufficiently close" to that of the standard in "most"

individuals. This is a more formal statement of the naive notion of bioequivalence mentioned in the Introduction. The essential idea is that most individuals will be expected to have similar bioavailabilities on the two formulations in order to call them bioequivalent. To state the corresponding statistical hypothesis requires another quantity, P_E, the proportion of the population that is bioequivalent for the two formulations of interest:

$$P_E = Pr\left[1 - R_I \leq \frac{\tau_{iN}}{\tau_{iS}} \leq 1 + R_I\right] \tag{4}$$

R_I may but need not be the same as R_A, the criterion for average bioequivalence. Now, the statistical hypothesis corresponding to individual bioequivalence is: $P_E \geq MINP$. *MINP* is the minimum proportion of the population in which the two formulations must be bioequivalent in order to call the two formulations (individual) bioequivalent. *MINP* would normally be expected to be a regulatory concern. For example, the regulatory agency could specify that for at least 80% of the population, their average bioavailability on the new formulation, τ_{iN}, must be within 10% of that of the standard, τ_{iS}. That would correspond to $R_I = 0.1$ and $MINP = 0.8$.

The two notions of bioequivalence correspond to two distinct clinical contexts. In the first context, a patient is started on a new drug. In this context, average bioequivalence, or some other form of population bioequivalence, may be appropriate—since the clinician has no information on that individual's response, the population averages are an appropriate basis for clinical decision making. (Though similarity of population distributions, not just similar averages, is actually what would be ideal here.) In the second context, the patient is switched to a new formulation, perhaps by the clinician but maybe by the pharmacist in areas permitting generic substitution. In this context, when switching a patient, particularly one who has been titrated, one should want reasonable assurance that that patient will get the same efficacy from the new formulation, thus individual bioequivalence is required.

Is average bioequivalence sufficient to provide that reasonable assurance? The answer is "maybe," as shown in Table I, where we show values of P_E, the proportion of the population that is equivalent on the two formulations, corresponding to various cases of average bioequivalence. To compute values for P_E, we need to specify the actual distribution in Eq. (2). For Table I, we assumed normal distributions for the μ_{ij}.

Clearly, there are situations when average bioequivalence does imply individual bioequivalence. Those situations are when the correlation, ρ, between subjects' expected responses to the two formulations is sufficiently high or if between-subject variation in responses to the formulations, σ_B^2, is sufficiently low. While this may be somewhat reassuring regarding average

Table I. P_E For Various Cases of Average Equivalence[a]

Coefficient of variation (%) between-subjects	ρ	$\tau_N/\tau_S = 1.0$ $100(1-R_I)$		$\tau_N/\tau_S = 1.1$ $100(1-R_I)$	
		75%	80%	75%	80%
5	0.9	1.00	1.00	1.00	1.00
	0.7	1.00	1.00	1.00	0.99
	0.5	1.00	1.00	0.99	0.96
15	0.9	1.00	1.00	0.97	0.90
	0.7	0.97	0.92	0.86	0.77
	0.5	0.91	0.82	0.80	0.70
30	0.9	0.94	0.87	0.83	0.74
	0.7	0.73	0.63	0.67	0.57
	0.5	0.61	0.51	0.57	0.48

[a]Calculations are based on the assumptions of normal distributions for the μ_{ij}, as well as conditions Eq. (2). The between-subjects coefficient of variation is expressed in the raw (τ_{ij}) scale.

bioequivalence, there are two problems. First, how high is sufficiently high correlation and how low is sufficiently low between-subject variation depends on the magnitudes of both these variables. Second, standard procedures for average bioequivalence would need to be supplemented in some way in order to determine where on the table a given pair of formulations fell. The entire Table I corresponds to average bioequivalence by the usual (80%, 120%) criteria and the $\tau_N/\tau_j = 1$ columns correspond to identical population distributions, yet many cases, the boxed portion of Table I, correspond to less than 75% of the population demonstrating individual bioequivalence. As a final note, the values for τ_N/τ_S and between-subject coefficients of variation shown in Table I are not extreme ones; they correspond to value considered reasonable by Cabana (3) in his response to Haynes (4).

A METHOD FOR ASSESSING INDIVIDUAL BIOEQUIVALENCE

How might one assess individual equivalence? One approach would be based on modeling. One would specify particular distributions in Eqs. (1) and (2), express P_E in terms of the parameters of those distributions, and then, at least in principle, draw inferences on ρ and/or P_E. Essentially, this would be equivalent to determining where on Table I the average equivalence case fell. However, there is a major problem with this approach, namely, that inference procedures for ρ and P_E will (almost certainly) depend on asymptotic theory and hence may not be sensible choices for the typical sample sizes used in comparative bioavailability studies.

As an alternative for assessing individual bioequivalence, we propose the TIER. The basic idea behind the TIER is a statistical test on the single binomial probability P_E. To review, P_E is the proportion of the population of subjects for whom the two formulations being compared are indeed equivalent. If this proportion, P_E, is sufficiently high (at least *MINP*) in the population of subjects who may be treated, then we want to declare the new and standard formulations to be bioequivalent.

We wish to test the statistical hypotheses: (i) Null hypothesis (H_0): $P_E < MINP$; (ii) Alternative hypothesis (H_A): $P_E \geq MINP$. For testing hypotheses about binomial probabilities, the most commonly used method is based on the normal approximation to the binomial. This approach can be risky in comparative bioavailability studies, however, since the sample n, in these studies is usually not large. Consequently, we use exact binomial probability calculations in what follows.

The steps of TIER are:

1. Define subject i to be bioequivalent if

$$1 - ECRIT \leq \frac{X_{iN}}{X_{iS}} \leq 1 + ECRIT \tag{5}$$

   where *ECRIT* is the equivalence criterion in the sample. In what follows, we take $ECRIT = R_I$, where R_I is the equivalence criterion in the population.
2. Count the number of subjects, Y, meeting the equivalence criterion [Eq. (5)].
3. Calculate the p value as p value $= Pr[Y$ or more bioequivalent subjects$\,|\,P_E = MINP]$ and conclude individual bioequivalence if the p value is less than α, the selected significance level of the test.

It is also informative to construct a confidence interval for P_E. In this case a one-sided confidence interval, a lower bound for P_E, would correspond to the one-sided hypothesis of interest. Given Y bioequivalent subjects, the $100(1-\alpha)\%$ lower bound for P_E, P_{LB}, is found as the solution to

$$\alpha = Pr[Y \text{ or more bioequivalent subjects}\,|\,P_E = P_{LB}] \tag{6}$$

As for the test, tables of or a computer program for the binomial distribution are required to solve for P_{LB}. See ref. 5, Section 3.1. We also note that the usual correspondence between test and confidence interval holds, namely, that the test of the above one-sided hypothesis will be significant at the α level if and only if the $100(1-\alpha)\%$ lower bound for P_E, P_{LB}, exceeds *MINP*. [On request, we will provide a listing of a short FORTRAN program for calculating exact binomial confidence intervals. Alternatively, the .EXE file for use on IBM-compatible computers will be sent on receipt of a

DOS-formatted floppy ($5\frac{1}{4}$ inch, 2S/2D) in a *stamped* self-addressed floppy disk mailer.]

Table II summarizes the decision rule of the TIER as a function of *MINP* and the number of subjects, *n*. The alpha level is taken to be 0.05. The entries in the first four columns of the table are the critical values of the test, i.e., the minimum number of subjects (of *n*) required to be bioequivalent to be able to conclude (at the α level) that the two formulations are individual bioequivalent in the sense that P_E is at least *MINP*. If no value is given, such as for $n = 12$ and $MINP = 0.8$, then there is no value of *Y* for which it is possible to conclude $P_E \geq MINP$. This binomial test is a stringent rule. If $P_E \geq 0.9$ is desired, $n > 29$ is required for it even to be possible to conclude bioequivalence of the formulations; for common comparative bioavailability study samples of 14 to 21 subjects, all subjects in the study must be bioequivalent in order to conclude $P_E \geq 0.8$.

Table II. Minimum Number of Bioequivalent Subjects Necessary to Conclude $P_E \geq MINP$ at $\alpha = 0.05$ and Lower Confidence Bound Corresponding to 75% Bioequivalent in Sample

	Minimum Number of Bioequivalent Subjects (*MINP*)				Lower Bound
n	0.6667	0.75	0.8	0.9	for P_E
10	10	—	—	—	0.493
11	11	11	—	—	0.530
12	12	12	—	—	0.473
13	12	13	—	—	0.505
14	13	14	14	—	0.534
15	14	15	15	—	0.560
16	15	16	16	—	0.516
17	15	17	17	—	0.539
18	16	17	18	—	0.561
19	17	18	19	—	0.581
20	18	19	20	—	0.544
21	18	20	21	—	0.563
22	19	21	21	—	0.580
23	20	21	22	—	0.596
24	21	22	23	—	0.565
25	21	23	24	—	0.581
26	22	24	25	—	0.595
27	23	25	26	—	0.608
28	24	26	27	—	0.581
29	24	26	28	29	0.594
30	25	27	28	30	0.606
31	26	28	29	31	0.617
32	27	29	30	32	0.594
33	27	30	31	33	0.605
34	28	30	32	34	0.615
35	29	31	33	35	0.625

Table III. Required Sample Sizes for TIER (5% Test)[a]

	Assumed true P_E		
MINP	0.8	0.9	0.95
0.667	77/101	24/29	15/18
0.75	461/621	48/60	25/30
0.80	—	93/119	38/46

[a] Numbers in table are required sample sizes for powers: 80%/90%. Calculations are based on normal approximation, with continuity correction, to the binomial test.

As with any statistical hypothesis test, detection of small differences requires large samples, which implies that the relatively small samples typically used in comparative bioavailability studies are sufficient to detect only relatively large differences. For TIER, the difference is between P_E and *MINP*; to conclude bioequivalence when P_E is close to *MINP* is difficult. This is demonstrated in Table III. The samples in Table III are the minimum number of subjects required to have 80 or 90% power to conclude bioequivalence when P_E is the proportion of bioequivalent subjects and assuming negligible intraindividual error (see Discussion). The boxed section corresponds to typical samples sizes used in comparative bioavailability samples. For example, for $MINP = 0.75$, 25 subjects are required to have 80% power (30 to have 90% power) for declaring the two formulations to be equivalent (individual bioequivalence) when 95% of subjects in the population are bioequivalent ($P_E = 0.95$).

THE 75/75 RULE

In the early 1980s, the U.S. Food and Drug Administration (FDA) sought to design a rule for declaring two formulations to be bioequivalent if both the average bioavailability and the variability of the bioavailabilities were similar. In particular, the FDA suggested the "75/75 rule," where similarity was defined to be that at least 75% of the subjects would have bioavailability of the new formulation to be within at least 75% of that of the standard, i.e., at least 75% of the within-subject bioavailability ratios falling within 0.75 to 1.25.

A motivation for use of such a rule is to take into account the variability of bioavailability in addition to the average. The standard methods for assessing bioequivalence are comparisons of average bioavailabilities and do not address variability. While one can include a test of the null hypothesis of equal variances, such as the Pittman-Morgan adjusted F test as suggested by Haynes (4), this test, like the analysis of variance test for bioequivalence,

is for the wrong hypothesis (see refs. 6, and 7). In addition, this practice also means that we are doing two hypothesis tets that ought to be combined in some way into a joint decision rule with a joint Type I error.

The literature contains several other complaints about the 75/75 rule. For example, Metzler and Huang (8) noted that "it is apparent that it [the 75/75 rule] is not based on statistical principles" (p. 124). Metzler and Huang, Haynes (4), and Thiyagarajan and Dibbons (9) have complained about the poor performance characteristics of the rule. Responding to the very valid criticisms of the 75/75 rule, an FDA Task Force (10) has recommended that use of the rule be discontinued. We agree that the 75/75 rule as stated should never be used.

It will be apparent however that TIER is partially motivated by the 75/75 rule. Our premise is that the FDA's proposal of the 75/75 rule was based on a sound idea, namely, direct consideration of the individual bioavailability ratios; the problem was in the implementation. TIER starts with the idea behind the 75/75 rule and implements it as a proper statistical procedure. However, TIER should *not* be viewed as a defense of the 75/75 rule.

Returning now to the FDA's original proposal, the last column in Table II summarizes what could be concluded from the 75/75 Rule. The values given are the 95% lower bounds (P_{LB}) for P_E assuming that 75% of the sample subjects were determined to be bioequivalent. (The lack of monotonicity as one goes down the column is due to the discreteness of the problem. For example, for $n = 16$ to 19, "at least 75%" corresponds to $X \geq 12$ to $X \geq 15$, respectively, but for $n = 20$, at least 75% still corresponds to $X \geq 15$.) The 75/75 rule can be seen to allow conclusion of bioequivalence for potentially low values of P_E, namely, in the 0.5–0.6 range. For example, in a study of 20 subjects, 15 of whom satisfy the equivalence criterion, thus satisfying the 75% rule, all one can say is that the proportion of subjects in the populations who would demonstrate equivalence is at least 0.54. Our basis for claiming that the 75/75 rule was a bad decision rule was that there was no control over what could actually be said about the underlying P_E. In effect, the 75/75 rule fixed the decision process by fixing the rule for concluding bioequivalence at 0.75 n, and let the statistical hypotheses vary with n and α. This is the reverse of standard statistical procedure that fixes α, n, and the hypotheses, and then determines the rule for concluding in favor of the alternative hypothesis.

EXAMPLE

As an example we consider data from two erythromycin formulations in a study published by Clayton and Leslie (11) and considered previously

by Hauck and Anderson (7) and Metzler and Huang (8). Eighteen subjects were studied to compare the stearate (experimental) to the base (standard). In Fig. 1 we show P_{LB}, the 95% lower bounds for the proportion, P_E, of the population that would be bioequivalent on the two formulations, where an individual is called bioequivalent in the sample if his ratio of new to standard is between $1-ECRIT$ and $1+ECRIT$ for all $ECRIT$ between 0.0 and 0.5. For example, for any choice of $ECRIT$ such that $1-ECRIT$ is between 0.64 and 0.82, 5 of the 18 subjects had their ratios in the appropriate equivalence range [for $ECRIT = 0.18$, that equivalence range is (0.82, 1.18)]. P_{LB} for that X is then found from Eq. (6) as that value of P_E such that

$$0.05 = Pr(\geq 5 \text{ equivalent subjects} \mid P_E = P_{LB})$$

The solution is $P_{LB} = 0.116$, indicating that all we can conclude from these data is that at least 12% of the population will have equivalence ratios between 0.82 and 1.18. All the values for P_{LB} in Fig. 1 are quite small, further supporting the conclusions of Metzler and Huang (8), and Hauck and Anderson (7) that these data do not support Clayton and Leslie's (11) claim of bioequivalence.

Figure 1 highlights an advantage of confidence intervals over hypothesis testing for individual bioequivalence, namely, that neither *MINP* nor *ECRIT* need be prespecified. It would then make sense to present results as in Fig. 1 or as the corresponding table such as Table IV, an "individual bioequivalence profile." While this approach may not be consistent with standard regulatory practice, it does nicely summarize results. It is the sort of data that the physicians who make the treatment decisions perhaps ought

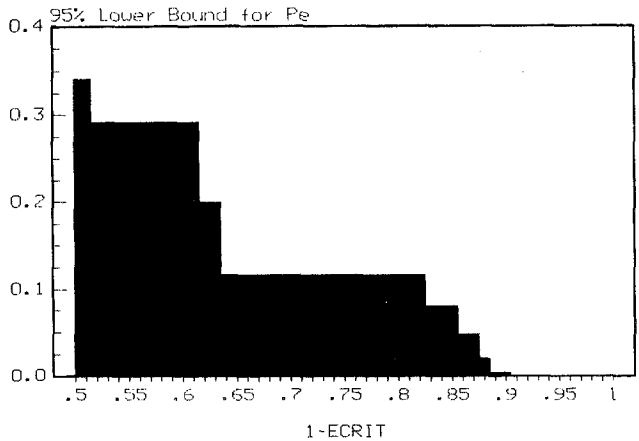

Fig. 1. Individual bioequivalence profile for data from Clayton and Leslie (11).

Table IV. Table Corresponding to Figure 1—Individual Bioequivalence Profile

P_{LB}	Intrasubject equivalence range	No. of subjects demonstrating bioequivalence
0.341	(0.50, 1.50)-(0.51, 1.49)	10
0.291	(0.52, 1.48)-(0.61, 1.39)	9
0.199	(0.62, 1.38)-(0.63, 1.37)	7
0.116	(0.64, 1.36)-(0.82, 1.18)	5
0.080	(0.83, 1.17)-(0.85, 1.15)	4
0.047	(0.86, 1.14)-(0.87, 1.13)	3
0.020	(0.88, 1.12)	2
0.003	(0.89, 1.11)-(0.90, 1.10)	1

to see. Since, in practice, comparative bioavailability studies are not commonly published, we urge that the within-subject bioequivalence profile be considered in the regulatory process.

DISCUSSION

The U.S. FDA's 75/75 rule was based on a sensible concept of bioequivalence, namely, that the response of each patient should be similar on the two formulations. While this is likely a common perception of what it means to claim that two formulations are bioequivalent, it is in fact very different from the concept underlying procedures based on a comparison of average bioavailabilities or, more generally, on any comparison of the population distributions. The comparison of means is based on between-patient considerations and seeks to show that the two formulations perform similarly when averaged over some population; it does not necessarily say anything about how an individual patient responds to the two formulations. In contrast, individual bioequivalence is based on within-subject considerations. Individual bioequivalence is generally a more stringent criterion. We showed in Table I that if between-subject variability is large, then the probability of within-patient equivalence can be low even if the population distributions of the two formulations are identical. While it is possible to construct cases corresponding to individual but not average bioequivalence, it seems likely that in common practice, as in the cases we have studied, within-subject equivalence will imply average equivalence.

The notion of population bioequivalence, the demonstrated similarity of the distributions of response to the two formulations, is appropriate in some situations. For a patient just beginning on a medication or maybe one who has started but is not nearly stabilized as to dose, there does not appear

to be any rationale to requiring individual bioequivalence. Simple average bioequivalence, however, is inadequate as it provides no control over variability—a new formulation that is more variable than the standard could still be close enough on average to be called (average) bioequivalent.

For a patient who has attained a steady dose level on a medication, however, we see no alternative to individual bioequivalence; Table I shows that even identical population distributions are not necessarily sufficient to conclude individual bioequivalence. This suggests that new formulations should be evaluated on the basis of both population and individual bioequivalence.

If one accepts that consideration of individual bioequivalence should be part of the review process, there is the question of how to assess individual bioequivalence. The primary argument in favor of the statistical test we propose, TIER, is that it addresses the notion of individual bioequivalence and it is the only procedure currently proposed for this purpose that is statistically valid. Although TIER is certainly a reasonable procedure, we hope it is not the final word on the subject. We particularly note that the very general requirement on specifying whether a subject demonstrates bioequivalence is both a strength and a weakness of TIER. The strength is that we make few assumptions for what is a nonparametric procedure; in particular, *no* distributional assumption on the X_{ij} need be made nor are homogeneous variances required. TIER, consequently, should also be more robust to outliers. The only assumption we have made is that the washout period is sufficiently long so that there are no carry-over or period effects. The corresponding weakness is that we may discard much information by converting the bioavailability measurements on each individual into a simple dichotomy. Some of the required sample sizes in Table III are certainly not reasonable for this type of study. What we do not know, however, is what are reasonable values for R_I, P_E, and *MINP*. This is a regulatory issue. As we mentioned earlier, an alternative is to model the distributions in Eqs. (1) and (2) and express P_E in terms of the parameters of those distributions. This would make more use of the data, assuming correctness of the distributional assumptions. However, the use of large-sample approximations in such approaches would need to be verified.

A second drawback to TIER is that it is inherently more conservative than we have presented it to be because, with $ECRIT = R_I$ as taken here, the probability of demonstrating equivalence in the sample is less than P_E. The probability in the sample is reduced because of the within-subject error which increases the variability of the equivalence ratio.

$$\text{Var}\,[\ln\,(X_{iN}/X_{iS})] = 2[\sigma_W^2 + (1-\rho)\sigma_B^2]$$

The conservative influence of the within-subject variation can be reduced

by employing extended-period crossover designs. Alternatively, this suggests that TIER could be improved (made more powerful for concluding individual bioequivalence) by letting the sample criterion for bioequivalence be somewhat looser than the population criterion, i.e., $ECRIT > R_I$.

We close by looking at the larger picture. In writing this paper and in talking to people about it, many issues related to bioequivalence studies become clearer. We are now going beyond our original intent, which was to formulate the concept of individual bioequivalence within the current context of bioequivalence testing. It is hard, however, to avoid the many issues once one starts looking critically at this area.

In this paper, we question the current practice in this area. As we have stated, the simple assessment of average bioequivalence seems largely inappropriate. In the context where bioequivalence studies are conducted for the purpose of concluding (near) equality of therapeutic efficacy without conducting additional clinical trials of efficacy, it seems most sensible to begin by asking that the test and standard formulations at least have sufficiently similar *distributions* (not just averages) of an appropriate measure of drug exposure (such as AUC) in the relevant populations, i.e., population bioequivalence. Then, if equal distributions is the relevant question, average bioequivalence is only part of the answer. At minimum what is missing is mention of variances.

The notion of similar distributions would argue that near equality of variances as well as near equality of means be assessed by methods for bioequivalence testing. However, what about a new drug with a similar average bioavailability to the standard formulation but smaller variance? Might not such a formulation, being more consistent, be preferable to the original formulation? If the variance of the new formulation is different in this way, should that formulation be called "bioequivalent" to the original formulation and thereby avoid the necessity for safety and efficacy studies? While a smaller variance might seem obviously preferable, what if the reported efficacy of the standard is due in part to its large variance and thus greater proportion of larger available dosages? On the other hand, if the variance of the new formulation is larger, the safety profile of the standard formulation may not be applicable.

Large within-subject variance is another problem: If that variance is large enough in the standard formulation, what does it mean to be bioequivalent? In such situations, it would be difficult to demonstrate bioequivalence of the standard to itself. Methods for average bioequivalence could demonstrate bioequivalence in that sense regardless of the variability by using a sufficiently large sample. Is there some level of inconsistency (variability) of bioavailability beyond which we should not be trying to find bioequivalent formulations?

Finally, we note two other issues that have arisen in discussions. First, are "healthy, male volunteers" the "relevant population" for conducting bioequivalence studies? The obvious answer is "no," and that use of such individuals requires justification in each particular instance. The concern here, of course, is whether patient's bioavailabilities are the same as those of healthy individuals. In the case of a generic alternative to an already approved drug, we do not see why patients cannot ethically be used, and wonder whether it is ethical to use healthy volunteers at all.

The second issue is one of sample size. In the United States, the regulatory agency (FDA) seems to believe that about 12–24 subjects is the right number for most bioequivalence studies. Given the history of inappropriate statistical methods for bioequivalence studies, we wonder about the provenance of this guideline. We also find the idea that the sample size might not vary considerably with experimental conditions such as variability to be contrary to the rest of statistical practice. We may be a bit sensitive about this, as the first objection we hear about TIER is that is requires too many subjects. We are not at all sympathetic to that objection; the sample size should be determined last after the appropriate hypotheses or, more generally, goals of the studies have been determined and the appropriate statistical methods selected. We expect that any method that uses the statistically correct assessment of average bioequivalence or goes beyond assessment of the simple average and demands reasonable variability as well will require larger studies than are now common practice.

To change the practice of comparative bioavailability studies requires regulatory changes. We hope this paper helps the process of encouraging new thinking and approaches.

ACKNOWLEDGMENTS

The views expressed in this paper are solely those of the authors. This paper has benefited considerably from the comments of many people. We particularly acknowledge Lewis Sheiner and K.-F. Yee and the other participants of the meeting on Statistical Methods in Biopharmacy held September 1989, in Paris.

REFERENCES

1. W. J. Westlake. Statistical aspects of comparative bioavailability trials. *Biometrics* **35**:273–280 (1979).
2. W. J. Westlake. Bioavailability and bioequivalence of pharmaceutical formulations. In K. E. Peace (ed.), *Pharmaceutical Statistics for Drug Development*, Marcel Dekker, New York, 1988, pp. 329–352.
3. B. E. Cabana. Assessment of 75/75 rule: FDA viewpoint. *J. Pharm. Sci.* **72**:98–99 (1982).

4. J. D. Haynes. Statistical simulation study of new proposed uniformity requirement for bioequivalency studies. *J. Pharm. Sci.* **70:**673-675 (1981).
5. W. J. Conover. *Practical Nonparametric Statistics*, 2nd ed., Wiley, New York, 1980.
6. S. Anderson and W. W. Hauck. A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Commun. Statist.* **A12:**2663-2692 (1983).
7. W. W. Hauck and S. Anderson. A new statistical procedure for testing equivalence in two-graoup comparative bioavailability trials. *J. Pharmacokin. Biopharm.* **12:**83-91 (1984).
8. C. M. Metzler and D. C. Huang. Statistical methods for bioavailability and bioequivalence. *Clin. Res. Practices & Drug Res. Affairs* **59:**109-132 (1983).
9. B. Thiyagarajan and T. W. Dobbins. An assessment of the 75/75 rule in bioequivalence. 1987 Proceedings of the Biopharmaceutical Section, American Statistical Association, Alexandria, VA, 1988, pp. 143-148.
10. Food and Drug Administration. Report by the Bioequivalence Task Force on recommendations from the bioequivalence hearing conducted by the Food and Drug Administration September 29-October 1, 1986 (1988).
11. D. Clayton and A. Leslie. The bioavailability of erythromycin stearate versus enteric-coated erythromycin base when taken immediately before and after food. *J. Int. Med. Res.* **9:**470-477 (1980).