Subject-by-Formulation Interaction in Bioequivalence: Conceptual and Statistical Issues

Walter W. Hauck,[1,7] Terry Hyslop,[1] Mei-Ling Chen,[2] Rabindra Patnaik,[3] Roger L. Williams,[4,5] and the FDA Population/ Individual Bioequivalence Working Group[6]

Received December 9, 1999; accepted January 3, 2000

Purpose. The FDA has proposed replacing the current average bioequivalence criterion with population and individual bioequivalence criteria that consider variances in addition to the difference of averages. One of these variances in the individual bioequivalence criterion measures subject-by-formulation interaction, the extent to which the test-reference difference varies from person to person. This paper discusses conceptual and statistical issues raised in various publications and presentations with respect to the presence and estimation of such an interaction.

Methods. We focus on the importance of subject-by-formulation interaction, an understanding of what is a large interaction, and the assessment of the magnitude of this interaction in bioequivalence studies. Simulation studies, examples from the literature, and data from FDA files are used to demonstrate the magnitude of the interaction and its distribution under various conditions.

Results. The concept of a large interaction is tied to the concept of a large mean difference. We suggest that an interaction greater than 0.15 is a conservative criterion for a large interaction. Magnitudes of estimated interaction are affected by variability, sample size, and the selection of data sets that pass average bioequivalence.

Conclusions. Examples of substantial interactions are beginning to appear. More data is needed before reaching definitive conclusions regarding the frequency and importance of observed interactions.

KEY WORDS: individual bioequivalence; within-subject variability; subject-by-formulation interaction; subgroups.

INTRODUCTION

Current regulatory practice in the US and elsewhere for establishing bioequivalence is to compare the mean of bioavailability measures, such as AUC and C_{MAX}, for the test (T) and reference (R) products. This is an average equivalence criterion (1). Following a publication by Anderson and Hauck (2), several alternative criteria have been proposed that were designed to establish bioequivalence in a manner that assesses within-subject similarity of the test and reference drug products. Hauck and Anderson (3) used the term switchability for the public health objective of ensuring that a patient currently taking a reference formulation could switch to the test formulation and expect essentially the same safety and efficacy outcomes. Bioequivalence criteria that focus on switchability are termed individual bioequivalence criteria. A specific example of an individual bioequivalence criterion has been discussed in various publications (4–6) and presented publicly in a preliminary and later draft FDA guidance for industry (7,8).

Most of the proposed individual bioequivalence criteria (9–13) depend implicitly or explicitly on three components of the T and R comparison: 1) comparison of means; 2) comparison of within-subject variances; and 3) subject-to-subject similarity of the difference between T and R products. The third component refers to a standard deviation that assesses subject-by-formulation interaction. A letter by Hwang et al. (14) is often credited as the first call for consideration of subject-by-product interaction. The FDA had a bioequivalence criterion, the 75/75 rule, intended to address this concern that was later dropped due to statistical considerations (15). The initial statistical work on the interaction was by Ekbohm and Melander (16), which proposed the interaction as a basis for assessing interchangeability of drug products.

Several published articles (17–19) have commented on the difficulties of estimating the subject-by-formulation interaction and have further suggested that it is unlikely to be present to any clinically important degree. Some of these papers (18–19) have also noted that large estimates of the subject-by-formulation interaction become more common with larger values of the within-subject standard deviation. The purpose of this paper is to discuss the subject-by-formulation interaction and to comment on some of these published statements.

CONCEPTUAL ISSUES

Definition

Average bioequivalence is based on the ratio of (geometric) mean bioavailability measures of the two formulations over all individuals in the study population. However, each individual has his or her own ratio of means for the two formulations. The average bioequivalence approach does not address whether the individual mean ratios differ from individual to individual. If they do differ, that is, if individuals vary in their ratios of average responses to the two formulations, a subject-by-formulation interaction is present.

Plots of hypothetical and actual data from a bioequivalence study assist in understanding subject-by-formulation interactions. In Figures 1–3, each line ("stick") connects an individual subject's mean bioavailability data for the test (T) and reference (R) products. Figure 1 displays a stick plot of a hypothetical bioequivalence study where the data show an increase in average bioavailability for the test product. There is no subject-by-formulation interaction in Fig. 1, since the increase is the same for all subjects. Figure 2 shows a stick plot for actual data from

[1] Biostatistics Section, Division of Clinical Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania.
[2] Office of Clinical Pharmacology and Biopharmaceutics, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.
[3] Division of Bioequivalence, Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.
[4] Office of Pharmaceutical Science, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.
[5] Present address: U.S. Pharmacopeia, Rockville, Maryland.
[6] Current membership: Mei-Ling Chen (Co-Chair), Rabindra Patnaik (Co-Chair), Dale Conner, Lawrence J. Lesko, Stella Machado, Donald Schuirmann, and Roger L. Williams.
[7] To whom correspondence should be addressed. (e-mail: w_hauck@lac.jci.tju.edu)

0724-8741/00/0400-0375$18.00/0

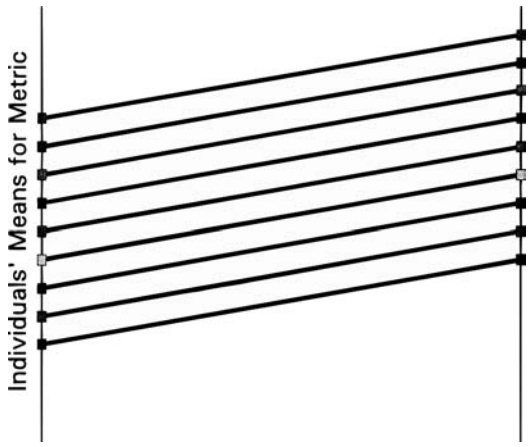

Fig. 1. Each line connects the geometric mean for the test to that for the reference in each individual. The vertical axis is in log scale, so parallel lines correspond to equal percentage changes between the two formulations in the original scale of the bioavailability measure (AUC or C_{MAX}).

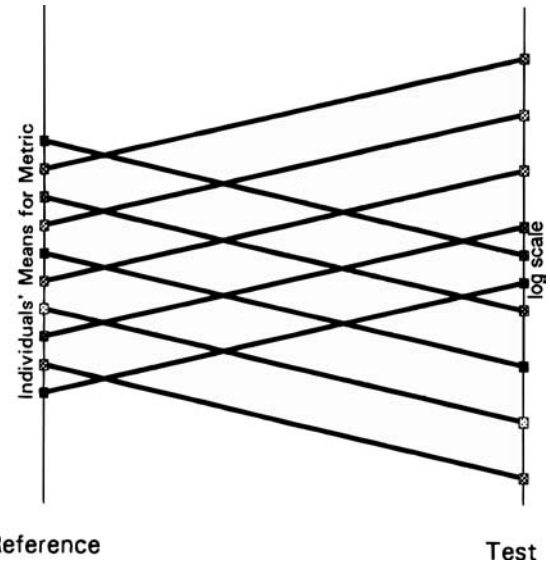

Fig. 3. The format for this figure is the same as Figure 1.

a four-period, two-formulation bioequivalence study for a beta-blocker that exhibits a large subject-by-formulation interaction. In Fig. 2, the four thicker lines (the lines with the greatest slopes) correspond to subjects who had average reference values about double those of the test or average test values about double those of the reference. These are the subjects manifesting a large subject-by-formulation interaction.

One way a subject-by-formulation interaction can arise is when a subgroup in the study population responds differently to either the T or R formulations than does the balance of the population. Consider hypothetical data where T is 50% more bioavailable than R in about half of the subjects and R is 50% more bioavailable than T in the remaining subjects (Fig. 3). These data indicate a subject-by-formulation interaction, because the T—R comparison is not the same across individuals. In this extreme situation, any individual taking one formulation would likely not expect the same safety and efficacy outcomes when switched to the other. However, averaging over subjects, the two formulations are similar, and so the test formulation could be declared average bioequivalent to the reference formulation. The average bioequivalence criterion does not require consideration of the individual T—R comparisons, though regulatory agencies have sometimes evaluated individual responses in the past (for example, the FDA's 75/75 rule used in the 1980's).

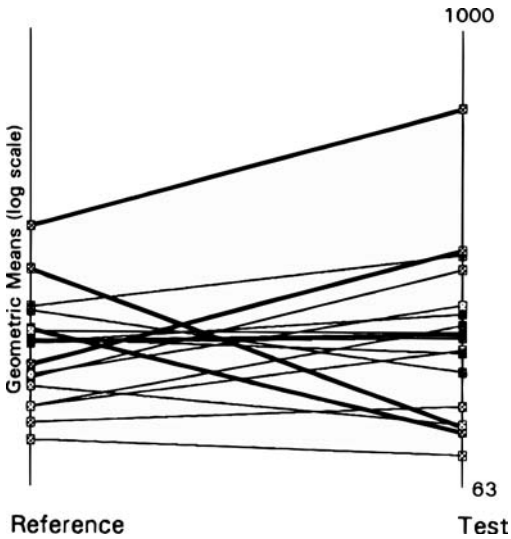

Fig. 2. The format for Fig. 2 is the same as Fig. 1. The data are for C_{MAX} from FDA data set #8 (a beta blocker) in data made publicly available by the FDA on its web site (http://www.fda.gov/cder/bioequiv/index.htm). The sample size is 19 and the estimated subject-by-formulation is 0.312.

In the analysis of standard two-period crossover trials, any subject-by-formulation interaction is included in the mean square for error from the analysis of variance (ANOVA). It would thus inflate the error variance over what would be true of just within-subject error, and thus would require more subjects to have the same power to pass average bioequivalence as would be required with no interaction. For example, if the test and reference each had a within-subject standard deviation (natural log scale) of 0.20 (corresponding to approximately a 20% within-subject coefficient of variation), a two-period design for the 80%/125% average bioequivalence criterion requires 19 subjects (or 20 for balanced assignment to sequence) to meet 80% power when the actual ratio of geometric means was 1.05. If the subject-by-formulation interaction (σ_D) was 0.25, then the apparent within-subject standard deviation, as estimated by the analysis of variance, would increase to 0.27 and the required sample size to retain 80% power would increase to 33 (34 for a balanced design). The standard two-period design used for assessing average bioequivalence cannot separate the subject-by-formulation interaction from the within-subject variability. A large subject-by-formulation interaction just makes a product appear more highly variable. For this reason, the presence of an interaction does make it more difficult but does not prevent a study from passing the average bioequivalence criterion, regardless of the magnitude of the interaction [unless subject selection adversely affects the mean comparison (20)].

In contrast, a sufficiently large interaction will make it impossible to pass an individual bioequivalence criterion, regardless of the sample size.

Magnitude of Subject-By-Formulation Interaction

The subject-by-formulation interaction is measured by σ_D, the standard deviation of the individual mean formulation differences in the log scale (see Note 1 in the Appendix). If all individuals have the same mean difference (as in Fig. 1), no subject-by-formulation interaction is present ($\sigma_D = 0$). If individuals vary considerably in their mean differences, the subject-by-formulation interaction is large. Within these general boundaries, we consider two approaches for interpreting the magnitude of a subject-by-formulation interaction.

The first approach relates to the subject-by-formulation interaction when the formulations differ in a subgroup but not in the remaining subjects of the population. This is accomplished by calculating the standard deviation of the formulation differences in individuals based on a specified proportion in the subgroup and specified ratio of test and reference means in the subgroup (Table I). As an example, if a subgroup represents 20% of the population and the test-reference ratio of means is 1.4 for individuals in the subgroup but 1.0 for everyone else, the value of the subject-by-formulation interaction is 0.135. The estimated subject-by-formulation interaction for the data in Fig. 2 is 0.312. This may be seen to be 'large' relative to many of the values in Table I. Examples of interactions with subgroups are beginning to appear. Carter et al. (21) reported an interaction with age for verapamil. One of the generic products had average AUC and C_{MAX} values 43% and 77% higher in the elderly than in the young subjects, while the reference and another generic had similar values in the elderly and young. Chen et al. (22) reported AUC and C_{MAX} values for 47 analytes in 26 bioequivalence studies that included both males and females. Thirteen percent of the 94 analyte-measure combinations showed some indication of gender-based subject-by-formulation interaction as defined by a greater than 20% difference between males and females in their test-reference mean difference.

The second approach to assess the magnitude of subject-by-formulation interaction is to determine what proportion of subjects have individual T/R mean ratios outside some predetermined interval. This approach ties the concept of a large interaction to that of a large mean difference. Table II shows the proportion of individuals outside an interval of 80–125% for various values of the subject-by-formulation interaction. For example, if the overall T/R mean ratio is 1.0, but there is a subject-by-formulation interaction of 0.15, almost 14% of the subjects would have their individual T/R ratios outside 80–125%. The 80%/125% limits are commonly used for average bioequivalence and thus serve as a commonly accepted notion of a large mean difference. It seems reasonable to say that a large interaction is one that corresponds to a substantial proportion of individuals with large individual mean differences; that is, in this case, falling outside 80%–125%. What constitutes "substantial" is an item for discussion and judgment. Our judgment was that if the proportion of individuals outside 80%–125% reached about 15%, that would constitute a large proportion.

Table II. Proportion Of Individual Test-Reference Ratios Outside 80/125%

Subject-by formulation interaction	Proportion of individuals outside 80/125%
0.050	0.000
0.075	0.003
0.100	0.026
0.125	0.074
0.150	0.137
0.175	0.202
0.200	0.265
0.225	0.321
0.250	0.372
0.275	0.417
0.300	0.457
0.325	0.492
0.350	0.524
0.375	0.552
0.400	0.577

Note: These calculations assume that the ratio of overall test-reference means is 1.0 and assume a bivariate normal distribution for the between-subject distribution (as described in Note 1).

Table I. Values of Subject-By-Formulation Interaction from Subgroups

Mean ratio in subgroup	Proportion in subgroup							
	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
1.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.10	0.021	0.029	0.034	0.038	0.041	0.044	0.047	0.048
1.20	0.040	0.055	0.065	0.073	0.079	0.084	0.089	0.091
1.30	0.057	0.079	0.094	0.105	0.114	0.120	0.129	0.131
1.40	0.073	0.101	0.120	0.135	0.146	0.154	0.165	0.168
1.50	0.088	0.122	0.145	0.162	0.176	0.186	0.199	0.203
1.75	0.122	0.168	0.200	0.224	0.242	0.256	0.274	0.280
2.00	0.151	0.208	0.248	0.277	0.300	0.318	0.340	0.347

Note: These calculations assume the mean ratio in individuals not in the subgroup is 1.0. If the proportion of individuals in the subgroup is p, the ratio of (geometric) means in the subgroup is R, and the ratio in the balance of subjects is 1.0, then, $\sigma_D = \sqrt{p(\ln R - p \ln R)^2 + (1 - p)(p \ln R)^2}$.

Accordingly, in public discussions, our group suggested that any interaction larger than 0.15 be considered large. From Table I, however, the identification of 0.15 as "large" seems conservative since important subgroup interactions correspond to values of the interaction smaller than 0.15. Considering only the parent compounds in the 25 three- and four-period datasets available to the FDA, for 17% of AUC_t's and 25% of C_{MAX}'s the estimated σ_D exceeds 0.15.

We recognize that if one's notion of a large mean difference changes, then the notion of a large interaction would change accordingly. One possibility for changing what is a large mean difference is to scale the difference to the within-subject standard deviation of the reference product. As the within-subject variability increased, so would the limits of what is a large mean difference. If the mean difference was scaled, then it would be sensible to scale the criterion for a large interaction in the same manner. In public meetings, however, there have been calls for constraining the allowed mean difference when scaling. With any constraint, the notions of large mean difference and large σ_D would change again to be consistent.

In this portion of the paper, we discussed the interaction parameter, σ_D, which is, of course, unknown and must be estimated from data. Related issues as to judging large *estimates* are discussed in the next section.

STATISTICAL ISSUES

Statistical Significance of the Subject-By-Formulation Interaction

Many of the data sets examined by CDER have estimated subject-by-formulation interactions that are large when considered in the light of the earlier discussion. A common comment in public discussions concerns the lack of statistical significance of the interactions in these data. Although large, these estimated interactions may not be statistically significant because of small sample size and the consequent limited power to obtain statistical significance of the interaction. Given that the studies examined by FDA were not designed to assess subject-by-formulation interaction, it is perhaps not surprising that many do not show statistical significance, though some do. More importantly, to ask for statistical significance is to ask the wrong question. Lack of statistical significance is not sufficient to claim that there is no important interaction (see Note 2). To support a claim of no important subject-by-formulation interaction, a confidence interval for the interaction should be calculated and compared to a range determined to be clinically unimportant. If the entire confidence interval for the subject-by-formulation interaction lies within this range, a claim of no important interaction is concluded.

Though not presented here, we have analyzed all the three- and four-period bioequivalence data sets available to CDER. In some of the FDA data sets, the observed confidence intervals for the subject-by-formulation interaction are narrow and include only small values. These data thus do support claims of no important subject-by-formulation interaction. However, the confidence intervals for many of the interactions in the FDA data sets are wide, including larger values of the interaction, and, thus, do not support a claim of no important interaction. As mentioned, for the statistical significance criticism, this is partly a sample size issue. These studies were not designed to demonstrate either the presence or absence of an interaction.

Endrenyi and colleagues (18–19) have noted that large magnitudes of *estimated* subject-by-formulation interaction are seen more often with increasing within-subject variability for the reference product (σ_{WR}) in the absence of actual interaction. We have observed this as well, and the same is true when comparing the means for average bioequivalence. The issue is variability in estimation. The larger the underlying variability and the smaller the sample size, the less well determined is the interaction. Figure 4 shows the results from simulated data for two identical formulations (equal overall means, equal within-subject variances, and no subject-by-formulation interaction) in four-period studies. The distribution of the estimated interactions clearly increases with the within-subject standard deviation of the reference product (first three columns) and decreases with sample size (last three columns). For example, with within-subject standard deviation of 0.30, 20% of the estimated interactions are greater than 0.15, our conservative suggested limit for a large interaction. This is in contrast to 36% greater than 0.15 for a within-subject standard deviation of 0.50 and less than 1% for a within-subject standard deviation of 0.15. In

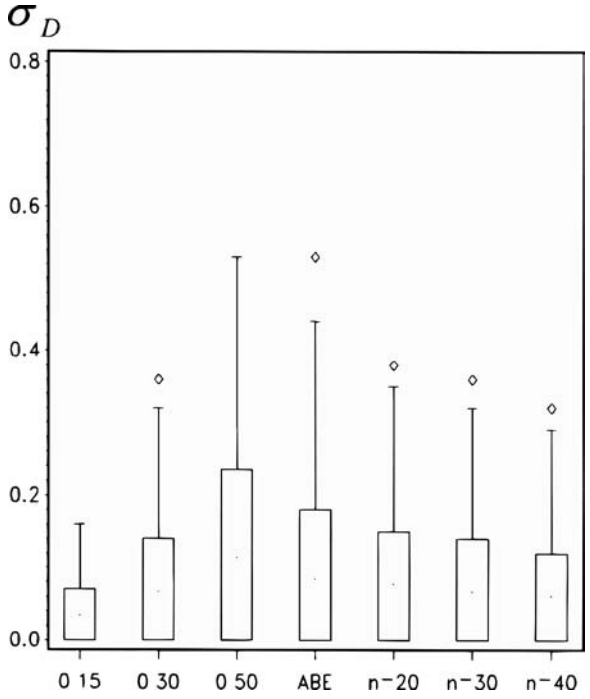

Fig. 4. Distributions of estimates of σ_D as a function of within-subject variability of the reference product (σ_{WR}) and of sample size. The distributions are each based on 1000 simulations of 2-sequence, 4-period crossover trials, assuming normality in the log scale and equal overall means, equal within-subject variances, and no subject-by-formulation interaction. For the first 3 bars, the trials have 30 subjects per trial (N) and σ_{WR} as denoted on each bar. The middle bar ("ABE") demonstrates the effect of restricting to datasets that pass the 80%/125% average bioequivalence criterion and is for the case N = 30 and σ_{WR} = 0.50. The last 3 bars represent σ_{WR} of 0.30 for varying sample sizes as noted on each bar. Analysis by the method of moments provides the estimates of σ_D. The distributions are shown as box plots. The center box shows the middle 50% of the distribution. The median is the center line, not seen here since the median is 0.0.

terms of sample size, 24% of estimated interactions are greater than 0.15 for a sample size of 20, 20% for sample size of 30, and 16% for sample size of 40 subjects. The sample size relationship was apparently missed by previous comments (18,19), but is important in considering the data available to the FDA, as the studies are larger on average than is typical. For the 22 four-period studies currently available to CDER, the sample sizes ranged from 19 to 67, with a mean of 33 and a median of 31.

We have seen above that the distribution of estimates of the interaction depends on the variability and sample sizes of the individual studies. It is also affected by selection, particularly for more variable studies; almost all studies submitted to the FDA pass average bioequivalence. The middle column of Fig. 4 shows what happens to the distribution in the third column when the simulated datasets that fail average bioequivalence are removed.

The types of distributions seen in Figure 4 are a possible means to assess the importance, in aggregate, of observed interactions from studies that were not designed to assess the interaction. We can compare the distribution of estimates from actual data to a simulated distribution of what would be expected in the absence of any actual interaction. If all or nearly all studies actually have no interaction between the two formulations, then the estimated interactions should have a distribution that looks like the simulated distributions. As more and more data sets have real interactions, the distribution from the overall data sets will be shifted higher than the simulated distributions. This is an idea that waits for more data.

Estimation of the Subject-By-Formulation Interaction

Two approaches have been considered for estimating the subject-by-formulation interaction: 1) maximum likelihood and restricted maximum likelihood (REML); and 2) method of moments. In its initial efforts (7), the FDA used a REML approach that forced the estimate of σ_D^2 to be nonnegative (see Note 3). One consequence of this approach is that it introduced a small positive bias in the estimate. Concern had been expressed that this bias could make it more difficult to pass an individual bioequivalence criterion. More recently (8), FDA has adopted a method of moments approach based on the work of Chinchilli and Esinhart (13, 23). This approach eliminates bias in the estimate arising from a requirement that the estimate of σ_D^2 be nonnegative. While avoidance of bias is important, the difference between the two approaches is not major. In our experience with simulated and actual data, where the data suggest more than a trivial interaction, the two methods result in similar values. The two methods yield different values at small values of the interaction, where the REML approach provides a small positive estimate while the method of moments provides a small estimate of either sign. A separate and more important advantage of the method of moments approach is that it is a much simpler method to describe and requires considerably less computational effort. REML, on the other hand, may be more efficient (have greater power) if data are missing, and it may be easier to incorporate carryover terms into the model when needed.

SUMMARY

The proposed FDA individual bioequivalence criterion (see Note 1) includes three components to compare test and reference formulations: means, within-subject variances, and subject-by-formulation interaction. The comparison of means, but not the variance terms, is considered by the current average bioequivalence criterion. Though an important component of the individual bioequivalence criterion, comparison of within-subject variances has not been controversial. Perhaps this is due to earlier work (24), unrelated to the topic of individual bioequivalence, calling for a comparison of variances. The most controversial component of individual bioequivalence criteria has been the subject-by-formulation interaction. This paper has attempted to provide background to appreciate what the interaction is measuring and some of the controversy surrounding the term. An important topic not addressed in this report is the mechanistic basis for a subject-by-formulation interaction. This understanding is being developed in further work at FDA and may be considered in a separate publication.

APPENDIX—TECHNICAL NOTES

1. In the underlying statistical model, a population of individuals has an overall mean in the log scale for the test, μ_T, and reference, μ_R. (Average bioequivalence is a comparison of μ_T to μ_R.) Individual i is assumed to have his or her own mean value for the test, μ_{Ti}, and reference formulation, μ_{Ri}. The statistical model is that the individual means are selected from a distribution with means μ_T and μ_R. Since individuals would be expected to be more similar on test and reference than would two different individuals, we allow for a correlation. In mathematical notation,

$$\begin{pmatrix}\mu_{Ti}\\ \mu_{Ri}\end{pmatrix} \sim \left[\begin{pmatrix}\mu_T\\ \mu_R\end{pmatrix}, \begin{pmatrix}\sigma_{BT}^2 & \rho\sigma_{BT}\sigma_{BR}\\ \rho\sigma_{BT} & \sigma_{BR}\sigma_{BR}^2\end{pmatrix}\right],$$

where σ_{BT} and σ_{BR} are the between-subject standard deviations for the test and reference formulations, respectively. (Note that there is no need for a distributional assumption at this point, though some confidence interval methods may assume normality.) The subject-by-formulation interaction is then

$$\sigma_D = \sqrt{Var(\mu_{Ti} - \mu_{Ri})} = \sqrt{(\sigma_{BT} - \sigma_{BR})^2 + 2(1 - \rho)\sigma_{BT}\sigma_{BR}}$$

A subject-by-formulation interaction may arise either due to less than perfect positive correlation (i.e., $\rho < 1.0$) or to unequal between-subject variances.

The individual bioequivalence criterion recommended by the FDA combines (aggregates) the three components of comparison:

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{\sigma_{WR}^2}$$

The within-subject standard deviations are denoted by σ_{WR} and σ_{WT}.

2. Standard statistical hypothesis testing permits one to reject the null hypothesis in favor of the alternative if the p-value is sufficiently small. Failure to reject the null hypothesis could be due to the correctness of the null hypothesis or due to a sample size insufficient to demonstrate the falseness of the null hypothesis. It is not proper to conclude in favor of the null hypothesis solely on the lack of statistical significance. Confidence intervals are the means to interpret results (25–27). This is the same issue that led to adoption of 90% confidence intervals for average bioequivalence (1,28).

3. Estimation was done using SAS Proc Mixed with a *csh* covariance structure for the between-subject component. The csh structure parametrizes the covariance matrix with a correlation that is constrained to be no more than 1.0 in absolute value. An alternative, the unstructured, or *un*, covariance structure parametrizes the covariance matrix with a covariance that is not constrained. The correlation calculated from a *un* analysis may be greater than 1.0 in absolute value (and often is in these analyses). The situations where *un* leads to a correlation greater than 1.0 correspond to cases where the method of moments leads to a negative estimate of the interaction variance. The bias at issue here is thus one associated with constrained estimation of the covariance matrix, not of REML versus method of moments.

ACKNOWLEDGMENTS

This work was supported in part by a contract from the Food and Drug Administration to Jefferson Medical College, Philadelphia, Pennsylvania.

REFERENCES

1. D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the bioequivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* **15**:657–680 (1987).
2. S. Anderson and W. W. Hauck. Consideration of individual bioequivalence. *J. Pharmacokinet. Biopharm.* **18**:259–273 (1990).
3. W. W. Hauck and S. Anderson. Types of Bioequivalence and related statistical considerations. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **30**:181–187 (1992).
4. R. Schall and R. L. Williams for the FDA Individual Bioequivalence Working Group. Towards a practical strategy for assessing individual bioequivalence. *J. Pharmacokinet. Biopharm.* **24**:133–149 (1996).
5. R. N. Patnaik, L. J. Lesko, M.-L. Chen, R. L. Williams, and the FDA Individual Bioequivalence Working Group. Individual bioequivalence: New concepts in the statistical assessment of bioequivalence metrics. *Clin. Pharmacokinet.* **33**:1–6 (1997).
6. M.-L. Chen, R. N. Patnaik, W. W. Hauck, D. J. Schuirmann, T. Hyslop, R. L. Williams, and the FDA Population/Individual Bioequivalence Working Group. An individual bioequivalence criterion: Regulatory considerations. *Statist. Med.* in press.
7. Food and Drug Administration. In Vivo Bioequivalence Studies Based on Population and Individual Bioequivalence Approaches. Federal Register, December 1997.
8. Food and Drug Administration. Average, Population, and Individual Approaches to Establishing Bioequivalence. Federal Register, September 1999.
9. L. B. Sheiner. Bioequivalence revisited. *Statist Med* **11**:1777–1788 (1992).
10. R. Schall and H. G. Luus. On population and individual bioequivalence. *Statist. Med.* **12**:1109–1124 (1993).
11. D. J. Holder and F. Hsuan. Moment-based criteria for determining bioequivalence. *Biometrika* **80**:835–846 (1993).
12. L. Endrenyi. A method for the evaluation of individual bioequivalence. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **32**:497–508 (1994).
13. V. M. Chinchilli. The assessment of individual and population bioequivalence. *J. Biopharm. Statist.* **6**:1–14 (1996).
14. S. Hwang, P. B. Huber, M. Hesney, and K. C. Kwan. Bioequivalence and interchangeability (letter). *J. Pharm. Sci.* **67**: iv (1978).
15. J. D. Haynes. Statistical simulation study of new proposed uniformity requirement for bioequivalence studies. *J. Pharm. Sci.* **70**:673–675 (1981).
16. G. Ekbohm and H. Melander. The subject-by-formulation interaction as a criterion for interchangeability of drugs. *Biometrics* **45**:1249–1254 (1989).
17. L. Endrenyi, G. L. Amidon, K. K. Midha, and J. P. Skelly. Individual bioequivalence: Attractive in principle, difficult in practice. *Pharm. Res.* **15**:1321–1325 (1998).
18. L. Endrenyi and L. Tothfalusi. Subject-by-formulation interaction in determinations of individual bioequivalence: Bias and prevalence. *Pharm. Res.* **16**:186–188 (1999).
19. L. Endrenyi, N. Taback, and L. Tothfalusi. Properties of the estimated variance component for subject-by-formulation interaction in studies of individual bioequivalence. *Statist. Med.*, in press.
20. N. T. Longford. Selection bias and treatment heterogeneity in clinical trials. *Statist Med* **18**:1467–1474 (1999).
21. B. L. Carter, M. A. Noyes, and R. W. Demmler. Differences in serum concentration of and responses to generic verapamil in the elderly. *Pharmacotherapy* **13**:359–368 (1993).
22. M.-L. Chen, S.-C. Lee, M.-J. Ng, and D. J. Schuirmann. Gender analysis of bioequivalence trials. American Society for Clinical Pharmacology and Therapeutics, Ninety-ninth Annual Meeting, New Orleans, Louisiana, March 30–April 1, 1998.
23. V. M. Chinchilli and J. D. Esinhart. Design and analysis of intrasubject variability in cross-over experiments. *Statist. Med.* **15**:1619–1634 (1996).
24. L. Z. Benet and J. E. Goyan. Bioequivalence and narrow therapeutic index drugs. *Pharmacotherapy* **15**:433–440 (1995).
25. R. Simon. Confidence intervals for reporting results of clinical trials. *Ann. Intern. Med.* **105**:429–435 (1986).
26. L. E. Braitman. Confidence intervals assess both clinical significance and statistical significance. *Ann. Intern. Med.* **114**:515–517 (1991).
27. S. N. Goodman and J. A. Berlin. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann. Intern. Med.* **121**:200–206 (1994).
28. W. W. Hauck and S. Anderson. A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J. Pharmacokinet. Biopharm.* **12**:83–91 (1984).