Statistical Simulation Study of New Proposed Uniformity Requirement for Bioequivalency Studies

JOHN D. HAYNES

Received February 19, 1980, from the *Medical Research Division, Lederle Laboratories, American Cyanamid Company, Pearl River, NY 10965.* Accepted for publication November 19, 1980.

Abstract □ The results of a statistical simulation study of the FDA-proposed 75/75 requirement for some bioequivalency studies are presented. The study used test drug products with the same true average bioavailability as the corresponding reference drug products but with various degrees of uniformity as measured by the coefficient of variation. The performance of the 75/75 requirement indicated that the probability of the test product passing was greater if both products had identical small coefficients of variation than if both had identical large coefficients of variation. Moreover, a test product compared to a reference product with equal variability had less probability of acceptance than an equally variable test product for another drug for which the reference product had less variability. Both results indicate that the 75/75 requirement should be withdrawn. An alternative uniformly most powerful, unbiased test, the Pitman–Morgan F test, is presented.

Keyphrases □ Bioequivalency studies—statistical simulation of the proposed 75/75 requirement □ Uniformity—statistical simulation of the proposed 75/75 requirement □ Bioavailability—statistical simulation of the proposed 75/75 requirement

The proposed rules for bioequivalency studies of tricyclic antidepressants (1), carbonic anhydrase inhibitors (2), probenecid (3), phenothiazine (4), and quinidine (5) all have a requirement dealing with the blood level results for individual subjects in addition to the usual requirement dealing with the average rate and extent of bioavailability. The new requirement essentially is that the relative bioavailability of the test product (compared to the study results for the clinically proven reference product for each subject) must be ≥75% for 75% of the subjects (or 70% for 70% of the subjects in Ref. 4).

The performance of this new 75/75 specification was found in the present study to have very undesirable characteristics for test and reference products that have equal true averages. Examination of bioequivalency studies of six hypothetical drugs indicated that the performance of the 75/75 specification sometimes can be irrelevant and inverse in practice. A different test is proposed here, consisting of the ratio of standard deviations with an adjustment for correlation.

BACKGROUND

Until recently, the only required proof of bioequivalence of two drug products has consisted of showing that any differences in the calculated averages for rate and/or extent of bioavailability were medically and/or statistically insignificant. The study must have a number of subjects sufficient to ensure a large probability, commonly 0.80, of detecting a true difference that would be large enough to be clinically important.

The proposed rules (1–5) for bioequivalency studies of five drug types have an additional requirement on the relative bioavailability for the subjects. This requirement pertains to both the average and the range, the latter being a characteristic mentioned only recently in such guidelines. Specifically, the relative bioavailability of the test product must be ≥75% for at least 75% of the subjects (or 70/70 in Ref. 4).

This specification is similar in some respects to the content uniformity specifications on individual capsules and tablets in Ref. 6, which require almost all assay values to fall within 15% of the stated potency. Among the differences, one is very important; in place of the fixed stated potency (usually the labeled amount), the 75/75 specification requires the use of an estimate of the reference drug product potency (*e.g.*, the area under the serum drug level curve) in each subject.

While other specifications deal with the average test product results, this proposed specification places a requirement on the range, uniformity, or homogeneity of individual results. (For some drugs, but not all, this requirement may be reasonable.) There are two obvious ways to compare the test product and the reference product in this respect; one is the 75/75 approach and the other is the ratio of the standard deviations for the two products. Both methods were examined in the present study.

The performance characteristics of the 75/75 specification in the special case in which the true averages are practically equal (as they are for most well-formulated products) were examined. This study examined the success of the 75/75 specification in the detection of poor and superior uniformity of blood drug levels following administration of the test product. Its performance was considered in studies involving six hypothetical drugs with some important parameter, such as the area under the serum curve, for the test products and reference products with various intersubject coefficients of variation (Table I). In all studies, the design was a two-way crossover. The intrasubject coefficients of variation in Table I were taken as 30%, a value often obtained from the error mean square of the crossover analysis of variance, and in Table II were taken as 20%, also a common value.

EXPERIMENTAL

Studies with Drug D1—Both drug products, *i.e.*, reference product and test product, have a true intersubject coefficient of variation of 60% (a common value in bioavailability studies), and the true averages are equal. Since both products are tested in the same subjects, there is usually a degree of correlation; *i.e.*, for a given subject, if the area under the serum drug curve on the reference product is above its 24-subject average, then on the test product the subject's area under the serum drug curve also is likely to be above its average.

The computerized, stochastic simulation procedure described in the *Appendix* was used to generate 1000 studies of 24 subjects (*i.e.*, 24,000 subjects) with a true correlation of 0.00 to serve as a baseline. Of these studies, 140 met the 75/75 criterion. The same procedure was followed for a range of correlation coefficients, ρ = 0.30–0.90 (Table I). The proportion increased toward 1.00 (the fair, just value for these two products) as rho increased.

Studies with Drug D2—Both of the products have an intersubject coefficient of variation of 40% (also a common value in bioavailability studies), and the true averages are equal. Table I shows that the proportion of D2 studies meeting the 75/75 criterion at any given ρ value was notably greater than the corresponding one for Drug D1. Although for both drugs the test products are just as good as the corresponding reference product in this respect, Drug D1 test product is at a disadvantage simply because the intersubject coefficient of variation is large (although equal) for both products. The probability of accepting the test product is greater for D2 than for D1, although both test products really are equally acceptable as substitutes for their respective reference products.

Studies with Drug D3—The intersubject coefficient of variation is 60% for the test product and 40% for the reference product, and the true averages are equal. The proportions in Table I are intermediate to those for D1 and D2 at any given ρ value except 0.80 and 0.90, where it is less than the D1 proportion. By comparing the results obtained with D1 and D3, it is seen that, according to the 75/75 specification, a test product with a range of AUC values that is 50% greater than the range of the reference product AUC values (D3: 60% for test product and 40% for reference product) usually has a greater chance of being declared bioequivalent, in this respect at least, than does a test product with the same range as the reference product (D1: 60% test product and 60% reference product).

Other Drugs—The determination of how many subjects should

0022-3549/81/0600-0673$01.00/0

Table I—Partial Description of Simulated Studies with Three Hypothetical Drugs

Drug	Intersubject CV, % Test Product	Intersubject CV, % Reference Product	Proportion of 1000 Studies (Intrasubject CV = 30%, 24 Subjects in Each) Meeting 75/75 Criterion[a] ρ = 0.00	ρ = 0.30	ρ = 0.40	ρ = 0.50	ρ = 0.60	ρ = 0.70	ρ = 0.80	ρ = 0.90
D1	60	60	0.15	0.19	0.22	0.26	0.30	0.35	0.43	0.52
D2	40	40	0.72	0.78	0.79	0.81	0.84	0.86	0.88	0.90
D3	60	40	0.29	0.32	0.34	0.36	0.38	0.39	0.41	0.43

[a] Rho (ρ) is the true correlation coefficient; the *Appendix* contains the assumptions and an outline of the method of calculating the proportions.

Table II—Partial Description of Simulated Studies with Three Different Hypothetical Drugs

Drug	Intersubject CV[a], % Test Product	Intersubject CV[a], % Reference Product	Proportion of 1000 Studies (Intrasubject CV = 20%, 12 Subjects in Each) Meeting 75/75 Criterion ρ = 0.00	ρ = 0.30	ρ = 0.40	ρ = 0.50	ρ = 0.60	ρ = 0.70	ρ = 0.80	ρ = 0.90
D4	60	60	0.24	0.29	0.32	0.35	0.40	0.47	0.54	0.66
D5	40	40	0.55	0.63	0.66	0.69	0.73	0.79	0.82	0.88
D6	60	40	0.38	0.44	0.45	0.47	0.51	0.53	0.56	0.59

[a] The values in these two columns are the same as for Drugs D1–D3.

participate in a bioequivalency study was made with the conventional equation involving Type I and II error rates, intrasubject coefficient of variation, and the size of the difference, *e.g.*, in average area under the curve, which is considered clinically important, often 20%. When a study is being planned, generally the variation of only the reference product is known. Therefore, that is what is used in the equation.

Table II shows the results of the simulation procedure for drugs with a smaller intrasubject coefficient of variation of 20%. The statistical power of detection of differences in average blood level *AUC* values in the studies is substantially the same for each of the three drugs and the same as that for the first three drugs.

Studies with Drug D5—The D5 proportions have been brought slightly closer to those for D4 than for D1 and D3, but the conclusions are unchanged. The probability of accepting the test product is greater for D5 than for D4, although both test products really are equally acceptable as substitutes for their respective reference products.

Studies with Drug D6—The D6 proportions also have been reduced but exceed the D4 proportions for ρ of ≤0.80. The picture is less unreasonable than for D3 but still is not reasonable. The conclusion given relating to D3 must be changed slightly for Drug D6; a test product with a range of *AUC* values that is 50% greater than the range of the reference product *AUC* values usually has a greater chance of being declared bioequivalent, in this respect at least, than does a test product with the same range as the reference product.

Independent Confirmation—The logic of the calculation procedure and the numerical results reported in Tables I and II were checked using the distribution theory provided by Marsaglia (7). The numbers agreed exactly to two decimals or were only 1, 2, or 3 low in the second decimal place.

Results in "Equal Means" Situations—For test products that give blood levels just as uniform as those obtained with the corresponding reference products (*e.g.*, D1 and D2), the probability of meeting the 75/75 specification is inversely proportional to the intersubject coefficient of variation for the reference product. Thus, the specification is unfair to those particular test products that must be compared to reference products with large intersubject coefficients of variation.

For test products giving blood levels that differ significantly in uniformity from those obtained with the corresponding reference products (*e.g.*, D3), the 75/75 specification is irregular; the inferior test product, that for D3, is either more likely or just as likely to be accepted as the equivalent test product, that for D1.

DISCUSSION

Generalization—The present study dealt with cases where the true means for the two products were equal, as they are for most well-formulated products. When the true means were unequal, the performance characteristics of the 75/75 specification were not investigated here; in those cases, the performance of the 75/75 requirement involves the differences in averages as well as ranges. The findings here are pertinent also in cases where the true means are almost equal. The lack of a broader generalization does not preclude a conclusion about the 75/75 specification. The results indicate that in some common situations, *i.e.*, where the means for the test and reference products are practically equal, the 75/75 requirement is unfair to some particular test products and is irregular in its disposition of test products of various qualities in other studies. Therefore, the 75/75 requirement should be dropped.

The performance (*i.e.*, the probability of accepting the test product) of the 75/75 specification when the test and reference products were truly practically identical (with only random variation operating) depended on the magnitude of the variation, an irrelevant factor. The performance was the inverse of what it should be when the test product's variability is large; with a given, fixed degree of variation for a test product, the smaller the reference product's variation (*i.e.*, the more divergent the two products), the greater the probability of acceptance of the test product.

Pitman–Morgan *F* Specification Procedure—A procedure for comparing the uniformity, range, or standard deviation of blood levels of the test product to the uniformity of those of the reference product is the use of intersubject standard deviations for the two products. However, since the standard deviations are determined in the same subjects (*i.e.*, in a crossover study), the usual Fisher F ratio of variances must be adjusted. This adjusted F, which is one form used to represent the Pitman–Morgan results, was shown (8) to be a uniformly most powerful, unbiased test.

Let:

$$F = (s_2/s_1)^2 \qquad \text{(Eq. 1)}$$

where the standard deviations s_1 and s_2 are for the reference product and test product, respectively. (If F is <1.00, no further calculations are necessary because the uniformity of blood levels of the test product is estimated to be as good as that of the reference product or better.)

Then:

$$\text{adjusted } F = [(F-1)^2(n-2)]/[4F(1-R^2)] \qquad \text{(Eq. 2)}$$

where R is the correlation coefficient for the pairs of values for each subject. If this correlation is zero or negative, then the unadjusted F test is appropriate with $N-1$ and $N-1$ degrees of freedom. The adjusted F has the same distribution as the tabulated F with 1 and $N-2$ degrees of freedom. The 0.05 probability level seems appropriate.

The results of one study[1] serve to illustrate the use of these equations for the area under the serum drug concentration–time curve for 18 subjects. The standard deviation of the area was 129.8 for the reference product and 156.3 for the test product. The intersubject correlation coefficient between the area values was 0.90. The Fisher F ratio was 1.45, leading to a Pitman–Morgan F value of 2.94 as calculated from $[(1.45-1)^2(18-2)]/[4(1.45)(1.0-0.90^2)]$. This value has 1 and 16 degrees of freedom and the associated probability is 0.11. It was concluded that the two standard deviations do not differ significantly.

The same Monte Carlo simulation data used for Drugs D1–D6 were used for the Pitman–Morgan F test with the critical F at the 0.05 probability level. The proportions of acceptable studies were as follows: D1 and D2, 0.992 at all ρ values; D4 and D5, 0.986 at all ρ values; D3 (N = 24), 0.743–0.749; and D6 (N = 12), 0.869–0.874. These results are reasonable.

Statistical Power—If the true (population) correlation coefficient were 0.80 (chosen here because it is tabulated in Ref. 9), if the Pitman–

[1] Unpublished data.

Morgan F test result in a study would be declared significant at the unilateral probability level of 0.01, if the study used 18 subjects, and if the power of detection of a difference in test and reference product standard deviations were set at 0.80, then the statistically detectable ratio would be (by rough extrapolation of the data in Ref. 9) >2 or 3. In other words, a test product standard deviation that is truly more than double or triple that of the reference product is likely to be detected in a study under these conditions. Any increase in sample size to improve the detection of differences in the standard deviations is believed to be unwarranted since, in view of other sources of inherent variation known in posology but not considered further here and in view of the desirability of minimizing the use of human subjects in bioequivalency studies, the power seems ample.

APPENDIX

It was assumed that the test product values and the reference product values followed a bivariate normal distribution with both means equal to 100, with certain chosen standard deviations and with correlation coefficients (Table I) modified to allow for the intrasubject variability which, among other things, reduces the estimated value of ρ (10).

The assumption that all actual bioequivalency data follow the bivariate normal distribution may be questioned; therefore, it is noted that very large intersubject coefficients of variation for *AUC*, *e.g.*, 150%, are symptomatic of drug products with either a very skewed distribution or some outliers. In such cases, the mean and standard deviation are not sufficient to describe the bivariate nonnormal distribution. However, if differences in only means are tested, the analysis of variance has been found to be a very robust procedure that is practically undisturbed by such things. Remedies after the fact include transformation of the data (*e.g.*, logarithmic) or identification of assignable causes for outliers; remedies *a priori* consist of controlling factors that have been found to cause aberrant results for the particular drug. In cases where the log transformation is used, the 75/75 rule also must be transformed. The performance of that transformed rule would be investigated differently than here, but intuitively the deficiencies of the rule would be substantially the same. The findings in the present study are appropriate for bioavailability parameters that are normally distributed.

The Monte Carlo simulation proceeded according to a FORTRAN program[2], which used a multivariate normal random deviate generator subroutine, GGNRM[3], and a local subroutine, CALC, which applied the 75/75 rule to each of the 1000 studies in turn. The main program then tallied up the number of studies that "passed." One run was made to obtain each cell in Tables I and II. The Pitman–Morgan F (9) was calculated for the identical sets of data.

REFERENCES

(1) *Fed. Regist.*, **43**, 6968 (1978).
(2) *Ibid.*, **45**, 11853 (1980).
(3) *Ibid.*, **45**, 48160 (1980).
(4) *Ibid.*, **45**, 56832 (1980).
(5) *Ibid.*, **45**, 72200 (1980).
(6) "The United States Pharmacopeia," 20th rev., United States Pharmacopeial Convention, Rockville, Md., 1980, p. 955.
(7) G. Marsaglia, *J. Am. Stat. Assoc.*, **60**, 193 (1965).
(8) M. G. Kendall and A. Stewart, "The Advanced Theory of Statistics," 3rd ed., vol. 2, Hafner, New York, N.Y., 1973, p. 531.
(9) W. A. Morgan, *Biometrika*, **31**, 13 (1939). E. J. G. Pitman, *ibid.*, **31**, 9 (1939).
(10) K. A. Brownlee, "Statistical Theory and Methodology in Science and Engineering," 1st ed., Wiley, New York, N.Y., 1960, p. 366.

ACKNOWLEDGMENTS

The author thanks R. Ganesan, S. Heft, R. Kohberger, J. Johnson, A. Weiss, and N. Bauman for helpful comments and Dr. Daniel Weiner for helpful comments and discussion. He also thanks T. Hoffman for the application of Marsaglia's work which provided a different viewpoint.

[2] Available from the author.
[3] International Mathematical and Statistical Libraries, Houston, Tex.

Assignment of Conformation and Configuration to Potassium Permanganate Oxidation Products of Quinidine

YEHUDA YANUKA, SHIMONA YOSSELSON-SUPERSTINE *, ARTOUL GERYES, and EDWARD SUPERSTINE *x

Received August 8, 1980, from the *Department of Pharmacy, School of Pharmacy, Faculty of Medicine, Hebrew University–Hadassah Medical Center, Jerusalem, Israel.* Accepted for publication December 2, 1980. *Present address: College of Pharmacy, University of Utah, Salt Lake City, UT 84112.

Abstract □ Two epimeric aldehydes [(*R*)- and (*S*)-quinidinals] and the corresponding acids [(*R*)- and (*S*)-norhydroquinidinoic acids] were prepared by the oxidation of quinidine. The π–π interactions of the carbonyl group and the aromatic moiety, as reflected in the NMR spectra, were compared with those of quinidine. NMR spectroscopic analyses made it possible to assign both the stable conformation and their configuration at C-3 to these molecules. The free hydroxyl group at C-9 must be present for the chemical shift values to be concentration dependent. These findings provide more information on association in the parent molecules.

Keyphrases □ Quinidine—oxidation products prepared, conformation and configuration assigned, NMR analyses □ Oxidation products—quinidine, conformation and configuration assigned to oxidation products, NMR analyses □ NMR spectroscopy—analyses, quinidine and oxidation products, conformation and configuration assigned

Previous work showed that the NMR spectra of quinidine (I) and hydroquinidine (1) differ significantly. The unique features encountered in the I molecule were attributed to π–π interactions. If changes in these π–π interactions are reflected in the NMR spectra of new compounds compared to the parent substances, it must be determined whether there is any correlation between the NMR data obtained and the biological activity of the compound. The antiarrhythmic activity of I and quinine differ greatly because of differences in configuration at C-8 and C-9. Comparison of the NMR data of I and quinine may provide information on the origin of these differences. Intraatomic distances in the molecules play an important role in their respective biological activities. NMR analyses may be valuable in assessing such differences and thus evaluating their possible therapeutic potential.

NMR analysis was used in the present work to measure the extent of π–π interaction by changing the intensity of

0022-3549/81/0600-0675$01.00/0