CENTER FOR DRUG EVALUATION AND RESEARCH

Guidance for Industry

The FDA published Good Guidance Practices in February 1997. This guidance was developed and issued prior to that date.

Additional copies are available from:
Office of Training and Communications
Division of Communications Management
Drug Information Branch, HFD-210
5600 Fishers Lane
Rockville, MD 20857

(Tel) 301-827-4573
(Internet) http://www.fda.gov/cder/guidance/index.htm

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, FOOD AND DRUG ADMINISTRATION

JUL 1 1992

GUIDANCE[1]

STATISTICAL PROCEDURES FOR BIOEQUIVALENCE STUDIES USING A STANDARD TWO-TREATMENT CROSSOVER DESIGN

I. INTRODUCTION

The Division of Bioequivalence in the Office of Generic Drugs usually evaluates bioequivalence by comparing the *in vivo* rate and extent of drug absorption of a test and reference formulation in healthy subjects. In a standard *in vivo* bioequivalence study design, study participants receive test and reference products on separate occasions, in either single or multiple doses, with random assignment to the two possible sequences of product administration. Samples of an accessible biologic fluid such as blood or urine are analyzed for drug and/or metabolite(s) concentrations, and pharmacokinetic parameters (AUC, C_{max} and T_{max}) are obtained from the resulting concentration-time curves. These pharmacokinetic parameters are then analyzed statistically to determine if the test and reference products yield comparable values.

Standard statistical methodology based on the null hypothesis is not appropriate to assess bioequivalence (1). The Division of Bioequivalence has therefore employed a testing procedure termed the two one-sided tests procedure (2) to determine whether average values for pharmacokinetic parameters measured after administration of the test and reference products are comparable. This procedure involves the calculation of a confidence interval (3) for the ratio (or difference) between the test and reference product pharmacokinetic variable averages. The limits of the observed confidence interval must fall within a pre-determined range for the ratio (or difference) of the product averages. The determination of the confidence interval range and the statistical level of

---

[1]This statement, prepared by the Division of Bioequivalence, Office of Generic Drugs (OGD), in consultation with the Division of Biometrics, Office of Epidemiology and Biostatistics, is an informal communication under 21 CFR 10.90 (b)(9) that represents the best judgment of the Division of Bioequivalence and the Office at this time. This statement does not necessarily represent the formal position of the Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA), and does not bind or obligate CDER, FDA, to the views expressed. For information about this guidance, please contact the Division of Bioequivalence (HFD-650), 7500 Standish Place, Metro Park North II, Rockville, MD 20855 (Tel: 301-295-8290; Fax: 301-295-8183).

significance are judgments made by the Division of Bioequivalence.

This Guidance provides information about general pharmacokinetic and statistical analyses of bioequivalence data to be conducted by sponsors of abbreviated new drug and antibiotic applications and addresses three specific aspects of the statistical analysis as follows:

1. Logarithmic transformation of pharmacokinetic data

2. Sequence effect

3. Outlier consideration

This Guidance becomes effective July 1, 1992. Any investigations initiated after that date should generally conform to the recommendations of the Guidance. Sponsors following a different approach are encouraged to discuss the matter in advance with FDA to prevent the expenditure of money and effort on preparing a submission that may later be determined to be unacceptable.

II. GENERAL METHODOLOGY

A. Pharmacokinetic Analysis

1. Single Dose Studies

At a minimum, the following pharmacokinetic parameters for the substance(s) of interest should be measured in a single dose bioequivalence study:

a. Area under the plasma/blood concentration - time curve from time zero to time t (AUC_{0-t}), calculated by the trapezoidal rule, where t is the last measurable time point.

b. Area under the plasma/blood concentration - time curve from time zero to time infinity ($AUC_{0-\infty}$), where $AUC_{0-\infty} = AUC_t + C_t/\lambda_z$, C_t is the last measurable drug concentration and λ_z is the terminal elimination rate constant calculated according to an appropriate method. The terminal or elimination half-life of the drug ($t_{1/2}$) should also be reported.

c. Peak drug concentration (C_{max}) and the time to peak drug concentration (T_{max}), obtained directly from the data without interpolation.

2. Multiple Dose Studies

At a minimum, the following pharmacokinetic parameters for the substance(s) of interest should be measured in a multiple dose bioequivalence study:

a. Area under the plasma/blood concentration - time curve from time zero to time τ over a dosing interval at steady state ($AUC_{0-\tau}$), where τ is the dosing interval.

b. Peak drug concentration (C_{max}) and the time to peak drug concentration (T_{max}), obtained directly from the data without interpolation, after the last dose is administered.

c. Drug concentrations at the end of each dosing interval during steady state (C_{min}).

d. Average drug concentration at steady state (C_{av}), where $C_{av} = AUC_{0-\tau}/\tau$.

e. Degree of fluctuation (DF) at steady state, where $DF = 100\% \times (C_{max} - C_{min})/C_{av}$.

Evidence of attainment of steady state for the test and reference products should be submitted in the bioequivalence study report.

B. Statistical Analysis

Parametric (normal-theory) general linear model procedures are recommended for the analysis of pharmacokinetic data derived from *in vivo* bioequivalence studies. An analysis of variance (ANOVA) should be performed on the pharmacokinetic parameters AUC and C_{max} using General Linear Models (GLM) procedures of SAS (4) or an equivalent program. Appropriate statistical models pertaining to the design of the bioequivalence study should be employed. For example, for a conventional two-treatment, two-period, two-sequence (2 X 2) randomized crossover study design, the statistical model often includes factors accounting for the following sources of variation:

1. Sequence (sometimes called Group or Order)

2. Subjects, nested in sequences

3. Period (or Phase)

4. Treatment (sometimes called Drug or Formulation)

The sequence effect should be tested using the [subject(sequence)] mean square from the ANOVA as an error term. All other main effects should be tested against the residual error (error mean square) from the ANOVA. The LSMEANS statement should be used to calculate least squares means for treatments. The ESTIMATE statement in SAS should be used to obtain estimates for the adjusted differences between treatment means and the standard error associated with these differences.

The two one-sided hypotheses at the $\alpha = 0.05$ level of significance should be tested for AUC and C_{max} by constructing the 90% confidence interval for the ratio between the test and reference averages.

III. LOGARITHMIC TRANSFORMATION OF PHARMACOKINETIC DATA

A. Statistical Assumptions

The assumptions underlying the ANOVA are (5):

1. Randomization of samples

2. Homogeneity of variances

3. Additivity (linearity) of the statistical model

4. Independency and normality of residuals

In bioequivalence studies, these assumptions can be interpreted as follows:

1. The subjects chosen for the study should be randomly assigned to the sequences of the study.

2. The variances associated with the two treatments, as well as between the sequence groups, should be equal or at least comparable.

3. The main effects of the statistical model, such as subject, sequence, period and treatment effect for a standard 2 X 2 crossover study, should be additive. There should be no interactions between these effects.

4. The residuals of the model should be independently and normally distributed. In other words, data from bioequivalence studies should have a normal distribution.

If these assumptions are not met, additional steps should be taken prior to the ANOVA including data transformation to improve the fit of the assumptions or use of a nonparametric statistical test in place of ANOVA. However, the normality and constant variance assumptions in the ANOVA model are known to be relatively robust, i.e., small or moderate departure from each (or both) of these assumptions will not have a significant effect on the final result.

B. Rationale for Log Transformation

1. Clinical Rationale

In a meeting in September 1991, the Generic Drugs Advisory Committee (GDAC) concluded that the primary comparison of interest in a bioequivalence study was the ratio rather than the difference between average parameter data from the test and reference formulations. Using log transformation, the general linear statistical model employed in the analysis of bioequivalence data allows inferences about the difference between the two means on the log scale, which can then be re-transformed into inferences about the ratio of the two averages (means or medians) on the original scale. Log transformation thus achieves the general comparison based on the ratio rather than the difference (6).

2. Pharmacokinetic Rationale

Westlake (7,8) observed that a multiplicative model is postulated for pharmacokinetic parameters in bioavailability/bioequivalence studies, i.e., AUC and C_{max} (but not T_{max}). Assuming that elimination of the drug is first order and only occurs from the central compartment, the following equation holds after an extravascular route of administration:

$$AUC_{0-\infty} = FD/CL$$

$$= FD/(VK_e)$$

where F is the fraction absorbed, D is the administered dose, and FD is the amount of drug

absorbed. CL is the clearance of a given subject which is the product of the apparent volume of distribution (V) and the elimination rate constant (K_e).[2]

The use of AUC as a measure of the amount of drug absorbed thus involves a multiplicative term (CL) which might be regarded as a function of the subject. For this reason, Westlake contends that the subject effect is not additive if the data is analyzed on the original scale of measurement.

Logarithmic transformation of the AUC data will bring the CL (VK_e) term into the equation in an additive fashion.

$$\ln AUC_{0-\infty} = \ln F + \ln D - \ln V - \ln K_e$$

Similar arguments were given for C_{max}. The following equation applies for a drug exhibiting one compartmental characteristics:

$$C_{max} = (FD/V) \times e^{-Ke \times Tmax}$$

where again F, D and V are introduced into the model in a multiplicative manner. However, after logarithmic transformation, the equation becomes

$$\ln C_{max} = \ln F + \ln D - \ln V - K_e T_{max}$$

Log transformation of the C_{max} data also results in the additive treatment of the V term.

3. Statistical Rationale

Logarithmic transformation of the data from bioequivalence studies can be used to circumvent the use of estimates of the reference product average for computation of the confidence interval for the ratio of product averages. This is an

---

[2]Note that a more general equation can be written for any multi-compartmental model as

$$AUC_{0-\infty} = FD/(V_{d\beta}\lambda_z)$$

where $V_{d\beta}$ is the volume of distribution relating drug concentration in plasma or blood to the amount of drug in the body during the terminal exponential phase, and λ_z is the terminal slope of the concentration-time curve.

advantage for the cases where a least squares estimate for the reference product mean is not well defined. Standard parametric methods are ill-suited to making inferences about the ratio of two averages, though some valid methods do exist (9). Log transformation changes the problem to one of making inferences about the difference (on the log scale) of two averages, for which the standard methods are well suited.

Many biological data correspond more closely to a log-normal distribution than to a normal distribution. The plasma concentration data including the derived parameters AUC and C_{max} tend to be skewed, and their variances tend to increase with the means. Log transformation is likely to remedy this situation and make the variances independent of the mean. In addition, frequency distributions skewed to the left (with a long tail to the right) are often made more symmetrical by log transformation.

This argument is actually less persuasive than the argument based on the additivity of the statistical model because it is based largely on the between-subject distribution of AUC and C_{max} values. For crossover studies, it is largely the within-subject distribution of values that determines the validity and efficiency of the standard parametric methods of analysis.

Despite the arguments regarding the effect of log transformation on normality of bioequivalence data, the Division of Bioequivalence recognizes that the limited sample size (20-30 subjects) in a bioequivalence study precludes a reliable determination of the underlying normal distribution of the data set either with or without log transformation.

C. General Procedures

Based on the arguments in the preceding section, the Division of Bioequivalence recommends that the pharmacokinetic parameters AUC and C_{max} be log transformed. Firms are not encouraged to test for normality of data distribution after log transformation, nor should they employ normality of data distribution as a justification for carrying out the statistical analysis on the original scale. Robustness of a balanced study to

nonnormality of the data plus use of log transformation will be adequate in most cases.

If a firm believes that the data of a particular bioequivalence study should be statistically analyzed on the original scale rather than the log scale, justification based upon a sound scientific rationale, as well as the statistical methods to be used, ought to be submitted to and reviewed by the Division of Bioequivalence.

D. Presentation of Data

The drug concentration in biological fluid at each sampling time point should be furnished on the original scale for all the subjects who participated in the study. The derived pharmacokinetic parameters should also be furnished on the original scale. The mean, standard deviation, and coefficient of variation for each variable should be computed and tabulated in the final report.

To facilitate bioequivalence comparisons, pharmacokinetic parameters for each individual should be displayed in parallel for the formulations tested. In particular, for AUC and C_{max}, the difference (T-R), ratio (T/R), and log of ratio (log T/R or ln T/R) between the test and reference values should be tabulated side by side for all the subjects. For each subject, the summary tables should indicate in which sequence (test, reference or reference, test) the subject received the product. Firms are encouraged to include histograms showing the frequency distribution of the difference and ln ratio (or log ratio) for the major pharmacokinetic parameters (AUC and C_{max}) in the submission.

In addition to the arithmetic mean for the test and reference products, the geometric means (antilog of the means of the logs), means of the logs, and standard deviations of the logs should be calculated for AUC and C_{max}. All means, including arithmetic mean, geometric mean, and means of the logs, standard deviations, and coefficients of variation are to be included in the report.

It is acceptable to use logarithms to the base 10 rather than natural logarithms. The report must state unambiguously which logarithms were used, and the use must be consistent throughout.

E. Equivalence Criteria

For a broad range of drugs, the Division of Bioequivalence has used a range of 80 to 120% for the ratio of the product averages as the standard equivalence criterion when the study data are analyzed on the original scale. This corresponds to a range of ± 20% for the relative difference between the product averages.

When log-transformed data are used in the analysis of AUC and C_{max}, using a range of 80 to 125% for the ratio of averages has an advantage over the 80 to 120% criterion in that for the analysis of log-transformed data, the probability of concluding equivalence is at a maximum if the ratio of averages is in fact 1.0, i.e., exact equality. For the analysis of log-transformed data with a criterion of 80 to 120%, the maximum probability of concluding equivalence occurs when the ratio of product averages equals approximately 0.98. For this reason, the Division of Bioequivalence has decided to use an equivalence criterion of 80 to 125% for the ratio of the product averages.

The 90% confidence interval for the difference in the means of the log transformed data should be calculated using methods appropriate to the experimental design. The antilogs of the confidence limits so obtained constitute the 90% confidence interval for the ratio of the test and reference product averages.

IV. SEQUENCE EFFECT

A major limitation of a conventional two-treatment, two-period, two-sequence crossover design is the confounding between (i) a true sequence or group effect, (ii) unequal residual or carryover effects, and (iii) treatment-by-period interactions. A true sequence effect (i.e., a difference between the average response for sequence group one and sequence group two) would not bias the determination of bioequivalence. Unequal residual effects, however, would bias this estimate. A treatment-by-period interaction based on an underlying physical basis (i.e., if there were actually something about the periods that caused the difference between the product averages to differ from one period to another), would lead to difficulties in interpreting the estimate of the ratio (difference) in the pharmacokinetic parameters between the test and reference formulations.

Even if there were no true sequence effect, no unequal residual effects, and no period-by-treatment statistical

interaction, approximately ten out of every one hundred standard two-treatment crossover studies would be likely to show an apparent sequence effect, if the testing is carried out at the ten percent level of significance.

If the ANOVA test for the presence of a sequence effect results in statistical significance, the actual cause cannot be determined from the data alone. In some cases, plausible causes might be evaluated by examining demographic or physiological subject data, but this examination is seldom conclusive.

On the basis of these considerations, the Division of Bioequivalence has determined that an *in vivo* standard two-treatment, two-period, two-sequence crossover bioequivalence study showing a statistically significant sequence effect may be acceptable provided:

1. It is a single dose study;

2. It includes only healthy, normal subjects;

3. The drug is not an endogenous entity;

4. More than adequate washout period has been allowed between the two phases of the study, and in the second phase, the predose biological matrix samples do not exhibit any detectable drug level in all subjects; and

5. The study meets all scientific and statistical criteria such as:

    a. It is based upon an acceptable study protocol;

    b. It contains an acceptable validated assay methodology;

    c. The study data are acceptable;

    d. Appropriate statistical analyses of the data are performed; and

    e. Acceptable confidence intervals for the pharmacokinetic parameters are achieved.

Under all other circumstances, the sponsor may be asked to conduct another study. After appropriate review with the Division of Bioequivalence, multiple dose studies and/or studies in patients demonstrating a statistically significant sequence effect may be acceptable provided they meet all other criteria listed above.

V. OUTLIER CONSIDERATION

Subject outliers are defined in bioequivalence studies as subjects having discordant values of one or more pharmacokinetic parameters when compared with other values for the rest of the subjects in a study. Because bioequivalence studies are usually carried out as crossover studies, the most important type of outlier is the within-subject outlier, where one or a few subjects differ notably from the rest of the subjects for the test product response versus the reference product response (e.g., test minus reference difference, test/reference ratio, or the log of the test/reference ratio).

The existence of an outlier could be indicative of the following problems with interchangeability of two formulations:

1. Product failure: a subject obtained an unusually high or low response to one or the other of the products because of a problem with the specific dosage unit(s) administered. Examples include a sustained/modified release dosage form exhibiting dose dumping or a dosage unit whose coating inhibited dissolution.

2. Subpopulation: a subject may be representative of a type of subject, present in the general population in low numbers, for whom the relative bioavailability of the two products is markedly different than it is for the majority of the population, and for whom the two products are not bioequivalent, even though they might be bioequivalent in the majority of the population.

In the case of product failure, it may make a difference whether the unusual response is observed on the test product or the reference product. In the case of a subpopulation, however, even if the unusual response is observed on the reference product, there may still be concern for lack of interchangeability of the two products.

Statistical tests exist for outlier identification. For detection of a single outlier, one important test is based on the absolute value of the Studentized Residual. Out of all the data in the study, the test focuses on the most extreme. Approximate critical values for this test have been published by Lund (10). In principle, however, outliers cannot be dropped from the analysis of the data solely on the basis of a statistical test. Sponsors who have identified one or more outliers should provide scientific evidence or explanations to justify the exclusion of the subject(s) data from statistical analysis.

VI. REFERENCES

1. Hauck WW, Anderson S. A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. J Pharmacokin Biopharm 1984;12:83-91.

2. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokin Biopharm 1987;715:657-680.

3. Westlake WJ. Response to Kirkwood, TBL.: bioequivalence testing - a need to rethink. Biometrics. 1981;37:589-594.

4. SAS/STAT™ User's Guide, Release 6.03 Edition, Cary, NC:SAS Institute Inc., 1988;1028.

5. Sokal RR, Rohlf FJ. Biometry: The principles and practice of statistics in biological research. 2nd ed. New York: WH Freeman and Co., 1981; 400.

6. Schuirmann DJ. Treatment of bioequivalence data: log transformation. In: Proceedings of Bio-International '89 - issues in the evaluation of bioavailability data, Toronto, Canada, October 1-4, 1989;159-161.

7. Westlake WJ. The design and analysis of comparative blood-level trials. In: Swarbrick J. ed. Current concepts in the pharmaceutical sciences, dosage form design and bioavailability, Philadelphia: Lea and Febiger, 1973;149-179.

8. Westlake WJ. Bioavailability and bioequivalence of pharmaceutical formulations. In: Peace KE. ed. Biopharmaceutical statistics for drug development. New York: Marcel Dekker, Inc., 1988;329-352.

9. Locke CS. An exact confidence interval from untransformed data for the ratio of two formulation means. J Pharmacokin Biopharm 1984;12:649-655.

10. Lund RE. Tables for an approximate test for outliers in linear models. Technometrics. 1975;17:473-476.