An Oral Contraceptive Drug Interaction Study

Thomas E. Bradstreet and Deborah L. Panebianco
Merck Research Laboratories

Journal of Statistics Education Volume 12, Number 1 (2004),
ww2.amstat.org/publications/jse/v12n1/datasets.bradstreet.html

Copyright © 2004 by Thomas E. Bradstreet and Deborah L. Panebianco, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Bioequivalence; Crossover; Graphics; Phase I Clinical Trial.

Abstract

This article focuses on a two treatment, two period, two treatment sequence crossover drug interaction study of a new drug and a standard oral contraceptive therapy. Both normal theory and distribution-free statistical analyses are provided along with a notable amount of graphical insight into the dataset. For one of the variables, the decision on the presence or absence of a drug interaction is reversed depending on whether the normal theory or the distribution-free analysis is favored. The data also contain statistically significant period effects, statistically significant but clinically unimportant treatment effects, some modest degree of structural nonnormality; and modest to more extreme outliers. This and 28 other pedagogically useful datasets can be found at www.math.iup.edu/~tshort/Bradstreet.

1. Introduction

In the pharmaceutical industry, following the evaluation of new drugs in animals, testing in humans begins in Phase I clinical pharmacology studies which typically are conducted in healthy subjects. For a few diseases (such as cancer and AIDS) patients may be used initially. It is important to evaluate to what degree, if any, that a new drug interacts with food (Bradstreet 2000), alcohol, or other drugs. In the study presented, the interest was in determining whether a new asthma drug, Drug D, altered the presence in the blood of an established oral contraceptive therapy when the two therapies were taken together.

This dataset can be used in relatively sophisticated statistics courses which emphasize either biostatistics, statistical methods in clinical trials, data analysis, or experimental design. The subject matter is of interest and is accessible to faculty, and both statistics and nonstatistics students. The level of study detail and statistical analyses which should be transferred to the classroom is dependent upon the students’ needs, for example, more detail and more sophisticated statistical analyses and diagnostic work for graduate level biostatistics or statistics students, and less detail and less sophisticated statistical analyses and diagnostic work for undergraduate service course students. This and 28 other pedagogically useful datasets can be found at www.math.iup.edu/~tshort/Bradstreet or see Bradstreet (1991, 1992,
Section 2 reviews bioavailability and "proving" similarity through an average bioequivalence evaluation as applied to interaction studies. Section 3 presents the motivation for the study and how the study was conducted. Section 4 illustrates the content of the dataset contained in the file OCDRUG.dat.txt. Section 5 presents a normal theory statistical analysis of the two treatment, two period, two treatment sequence (2,2,2) crossover study, the corresponding distribution-free methods, insightful graphical procedures specific to this study design, and guidance for software. Section 6 describes how to obtain OCDRUG.dat.txt.

2. Bioavailability, Interaction Studies, and Bioequivalence

2.1 Bioavailability

Bioavailability is defined as "the rate and extent to which the active ingredient or active moiety [portion] is absorbed from a drug product and becomes available at the site of action" in the human body (Food and Drug Administration 2000, p. 3). The bioavailability of a drug is characterized by summarizing its time course in the blood. It is assumed that the concentrations in the blood are representative of the concentrations at the site of action. After a drug is given to a subject, plasma samples are taken at previously determined time points. The plasma samples are then assayed individually for drug concentration levels, and a plot of the plasma concentration (y-axis) vs. time (x-axis) curve is constructed. Usually, the plasma concentration vs. time curve is summarized by three variables: (i) Area Under the plasma Concentration vs. time curve (AUC), a measure of total absorption; (ii) maximum plasma concentration (Cmax), a measure of the extent of absorption; and (iii) time to maximum plasma concentration (Tmax), a measure of the rate of absorption. Most often, AUC and Cmax are the primary summary measures of bioavailability. AUC is estimated from zero hours to the last measured time point (for example, 24 hours after dosing) using the linear trapezoidal rule (Gibaldi and Perrier 1982; pp. 445-447), and if necessary, extrapolated from the last measured time point to infinity by incorporating the elimination rate constant which describes the rate of drug removal from the body (Gibaldi and Perrier 1982; pp. 447-448). Cmax and Tmax are simply observed from the plasma concentration vs. time curve.

2.2 Interaction Studies

It should be emphasized that the primary objective of an interaction study is to "prove" the absence of interaction. This focus is quite different statistically from the frequentist hypothesis testing paradigm which many statisticians are initially trained to consider. Typically, we start with the null hypothesis of "no difference" and then we proceed to work quite hard either to provide statistical evidence against the null case (Fisher's significance testing), or we attempt to choose between the null and alternative hypotheses with maximal power for a certain size test procedure (Neyman-Pearson hypothesis testing). However, in interaction studies, it is assumed initially that some degree of interaction exists (a nonzero difference), and it is the objective of the experimenters to show that the size of the interaction is not of clinical importance, possibly being zero. In the current drug interaction study, the interest was in demonstrating that the AUC and Cmax of the oral contraceptive are relatively unchanged, within reasonable clinical allowance, when Drug D is given concomitantly. If this is indeed the case, then giving Drug D with the oral contraceptive (OCD) is said to be bioequivalent to giving the oral contraceptive alone (OC).

2.3 Bioequivalence

Bioequivalence is defined as "the absence of a significant difference in the rate and extent to which the
active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study" (Food and Drug Administration 2000; p. 4). There are three categories of criteria by which bioequivalence could be evaluated: average bioequivalence, population bioequivalence, and individual bioequivalence. The former is the currently favored regulatory criterion for most drugs and formulations (Food and Drug Administration 2000; p. 11); the latter two are criteria which are under development and serious consideration by several groups of scientists (Food and Drug Administration, 1999). This article follows the average bioequivalence evaluation paradigm.

In many countries, the current regulatory criterion for establishing average bioequivalence is by way of the so-called "20% rule." Two treatment regimens are considered bioequivalent if their true mean AUCs are within 20% of each other. In our example, the true mean AUC for the oral contraceptive when Drug D is administered concomitantly, , should be within 20% of the true mean AUC when the oral contraceptive is given alone, , i.e., their ratio, , falls in the interval [0.80, 1.25] (Food and Drug Administration 1999; pp. 2, 3). The lower end of this average bioequivalence interval addresses equivalent efficacy; the upper end reflects on the concern for equivalent safety. The upper end of the interval is extended to 1.25 so that the interval on the log scale is symmetric about zero.

Currently, the most popular approach for establishing average bioequivalence is a two one-sided hypothesis testing procedure (Food and Drug Administration 1999; pp. 2, 3) which is equivalent to a confidence interval approach (Food and Drug Administration 1999; p. 11). Regulatory agencies typically require that bioequivalence be concluded with a reasonable degree of assurance. For example, in the United States a 90% confidence interval is required. Two treatment conditions are declared as bioequivalent if the calculated 90% confidence interval for the true ratio of means falls between 0.80 and 1.25 (Food and Drug Administration 1999; pp. 3, 5, 6, 11). AUC and Cmax are natural log transformed prior to statistical analysis (Food and Drug Administration 1999; pp. 3, 4, 10, 11, D-1, D-2) for many reasons, among them are: multiplicative pharmacokinetic models, the bioequivalence criterion compares the treatment conditions proportionally, and generally the log-transformed data appear to more closely follow a normal distribution.

A reader who is interested in more examples and exploring further the issues and methods used in establishing bioequivalence should see the manuscripts by Bradstreet (1993) and Bradstreet and Dobbins (1996), and the book by Chow and Liu (2000). Also see the special bioequivalence issues of the Drug Information Journal (1995), the Journal of Biopharmaceutical Statistics (1997), and Statistics in Medicine (2000), and the two guidance documents for the U.S. Food and Drug Administration (1999, 2000).

3. The Study

3.1 Motivation for the Study

There have been reports that some drugs can enhance the metabolism of oral contraceptives and result in pregnancy in some women. A study was planned to evaluate the impact of the oral dosing of 125 milligrams (mg) of a new drug for asthma, Drug D, given twice daily, on the bioavailability of the components of a standard oral contraceptive combination of 35 micrograms (mcg) of ethinyl estradiol (EE) and 1 milligram (mg) of norethindrone (NET). It was conjectured that Drug D given concomitantly with the EE and NET combination would not have any clinically significant effect on either the AUC (in picograms times hours per milliliter; pg*hr/ml) or Cmax (in picograms per milliliter; pg/ml) of either EE or NET. Specifically, it was predetermined that the coadministration of the oral contraceptive and Drug D (OCD) would be considered clinically similar to the oral contraceptive alone (OC) if the experimental
evidence suggested that there is a 20% or less difference in the true mean values, OCD vs. OC, for each of EE AUC, EE Cmax, NET AUC, and NET Cmax.

3.2 Study Conduct

Twenty-two female subjects were allocated randomly to one of two treatment sequences in a 2,2,2 crossover design. The study was conducted over two consecutive menstrual cycles for each subject. The first treatment period corresponded with the first menstrual cycle; the second treatment period corresponded with the second menstrual cycle.

During the first treatment period, each female subject received either 125 mg of Drug D or a matching placebo tablet two times daily for eight days (16 doses) starting with either Day 1, 2, 3, or 4 of her menstrual cycle as was convenient for the subject. The oral contraceptive administration always started on Day 1 of her menstrual cycle. On the morning of the eighth day of taking either Drug D or placebo, the female subject took her dose of either Drug D or placebo and also the oral contraceptive. Blood samples were drawn immediately prior to dosing and at 0.5, 1, 1.5, 2, 3, 4, 6, 8, 10, 12, and 24 hours after dosing. The quantities of EE and NET present at each time point were assayed from the plasma samples, and these assay values were used to construct separate plasma concentration vs. time curves for EE and NET for each subject. From the two curves, AUC (pg*hr/ml) and Cmax (pg/ml) values were calculated. During the second treatment period, this process was repeated with each female subject receiving whichever treatment, either 125 mg of Drug D or placebo, that she had not received during the first treatment period. The oral contraceptive was again started on Day 1 of her menstrual cycle; Drug D or placebo was started on the same relative day (1, 2, 3, or 4) as during her previous menstrual cycle. As an example, the four plasma concentration vs. time curves constructed for Subject 8 are shown in Figure 1.
Eleven of the female subjects (numbers 2, 4, 5, 8, 9, 12, 14, 16, 18, 20, 21) received Drug D concomitantly with the oral contraceptive combination of EE and NET in the first study period, followed by matching placebo and the oral contraceptive in the second study period (treatment sequence 1). The other eleven (numbers 1, 3, 6, 7, 10, 11, 13, 15, 17, 19, 22) received matching placebo and the oral contraceptive during the first study period followed by Drug D given concomitantly with the oral contraceptive in the second study period (treatment sequence 2).

4. The Data

The dataset contains several pieces of information about the study. For each female subject, there are two lines of data which display the AUC (pg*hr/ml) and Cmax (pg/ml) values which were calculated and observed from the plasma concentration vs. time curves for both EE and NET. The details of which treatments (Drug D or placebo) were taken concomitantly with the oral contraceptive in each treatment period (first or second) are also provided. For example, female Subject 8 received Drug D in the first treatment period and placebo in the second treatment period (treatment sequence 1). The AUC values calculated for EE and NET were 3,328.1 and 118,340.0 pg*hr/ml when she received Drug D during the first treatment period; the corresponding values in the second period following treatment with placebo.
were 2,941.4 and 148,220.0 pg*hr/ml, respectively. The observed Cmax values were 431 and 18,100 pg/ml for Drug D; they were 302 and 24,200 pg/ml for placebo.

5. Data Analyses

5.1 Initial Exploration of the Data

An informative first look at the data can be achieved by constructing three groups of graphics: hanging dot plots of individual subject ratios, individual subject profile (sometimes called "spaghetti") plots, and scatter plots.

Hanging Dot Plots of Individual Subject Ratios

presents individual within-subject natural log transformed ratios, log OCD/OC, of the AUC and Cmax data for EE and NET. The y-axis is labeled with antilog values. The target bioequivalence (interaction) limits are indicated by the horizontal dotted lines located at 0.8 and 1.25 on the y-axis. If the two treatments, OCD and OC, are bioequivalent, then it is reasonable to expect the individual ratios to cluster somewhere between 0.8 and 1.25. Figure 2 suggests that when Drug D is given concomitantly with the oral contraceptive, EE AUC is increased for most individuals, often by more than 20%; EE Cmax is increased for many individuals, sometimes to a notable degree (Subjects 2 and 20, for example); and NET Cmax is decreased for most individuals often by more than 20%, sometimes to a notable degree (Subject 9, for example). It also appears that NET AUC generally decreases in the subjects in treatment sequence 1 (OCD then OC), but generally increases among the subjects in treatment sequence 2 (OC then OCD). But, as will be shown in the next section, this result is due to an unequal effect of the treatment periods and not due to a true sequence or subject group effect.
Individual Within-Subject Natural Log Transformed Ratios, log OCD/OC. The $y$-axis is labeled with antilog values. The target bioequivalence (interaction) limits are indicated by the horizontal dotted lines located at 0.8 and 1.25 on the $y$-axis.

**Individual Subject Profile ("Spaghetti") Plots**

In Figure 3 each line represents the response of one subject. The line connects the log AUC or log Cmax value which was observed when Drug D was given concomitantly with the oral contraceptive to the corresponding log AUC or log Cmax value which was observed when the oral contraceptive was given alone. The $y$-axis is labeled with antilog values. The $y$-axes are all different. In addition to providing within-subject information, these spaghetti plots provide important between-subject information such as the range of log AUC or log Cmax values observed, and indications of marginal location and marginal variances. For example, Figure 3c suggests for each of Subjects 18, 22, and 20, that there is little difference between their log NET AUC value either with or without Drug D. But, the log NET AUC values observed for Subjects 18 and 22 are many fold apart from those observed for Subject 20. Figure 3a suggests that log EE AUC is marginally greater with Drug D than without Drug D since most of the lines are decreasing from left to right, but the marginal variances appear quite similar.
**Figure 3.** Individual Subject Profile ("Spaghetti") Plots Ordered by Treatment. The y-axes are labeled with antilog values. The y-axes are all different. (a) log EE AUC; (b) log EE Cmax; (c) log NET AUC; (d) log NET Cmax.

**Figure 4** represents an alternative presentation of the spaghetti plot for log NET AUC where each subject’s profile is constructed ordered by study period rather than by treatment. This reflects the order in which the data were collected in the study and can be insightful for evaluating period effects, and sometimes for discriminating between carryover effects, sequence effects, and treatment-by-period interaction (see Jones and Kenward 1989, pp. 20-22, 39-51 and Pikounis, Bradstreet, and Millard 2001).
clearly suggests a period effect since the log NET AUC value observed for each subject in the second study period was generally larger than the corresponding value in the first treatment period, regardless of treatment, OCD or OC. This explains the individual ratio pattern observed in Figure 2. The period effect manifested itself as smaller individual ratios in treatment sequence 1 (OCD then OC) and larger individual ratios in treatment sequence 2 (OC then OCD).

Scatter Plots

presents square scatter plots of the log AUC and log Cmax data. The x- and y-axes are labeled with antilog values. The x-axes are all different; the y-axes are all different. If the two treatments are bioequivalent, then we expect to see many of the points close to the diagonal line. However, most of the points in Figure 5a lie above the diagonal line suggesting an increase in log EE AUC when Drug D is given. Similarly, in Figure 5d, most of the points lie below the diagonal line suggesting a decrease in log NET Cmax when Drug D is given.
Figure 5. Scatter Plots. The $x$- and $y$-axes are labeled with antilog values. The $x$-axes are all different; the $y$-axes are all different. (a) log EE AUC; (b) log EE Cmax; (c) log NET AUC; (d) log NET Cmax.

In addition to assessing the location of the bivariate point cloud vs. the diagonal line, it is important clinically to identify both concordant and discordant outliers. In the bioequivalence framework, concordant outliers are those bivariate points close to the diagonal line but are notably distant from the point cloud, situated near either end of the diagonal line. These points represent those subjects whose paired responses are somewhat similar for that subject, but are notably different (smaller - lower end of the diagonal line; larger - upper end of the diagonal line) than the magnitudes of the response of the other subjects. Subject 12 in Figure 5b and Subjects 20, 22, and 18 in Figure 5c are concordant outliers. Discordant outliers are those bivariate points which stray from the point cloud and stray from the diagonal line. These belong to subjects whose paired responses are not similar to each other for that subject, and are also notably different from the responses of the other subjects. For example, Subjects 2 and 20 in Figure 5b, and Subject 9 in Figure 5d, are discordant outliers.

5.2 The Model and the Normal Theory ANOVA

The two treatment, two period, two treatment sequence (2,2,2) crossover design can be modeled as:

$$Y_{yk} = \mu + \varepsilon_{yk} + \pi_j + \mu_j + \lambda_{d[k,j]} + \alpha_{d[k,j]-i} + \sigma_{yk}$$
where

\[ \mu \] = the overall mean effect

\[ \epsilon_{ik} \] = the effect of subject \( k \) in group \( i \), \( i = 1,2; k = 1, \ldots, n_i \)

\[ \pi_j \] = the effect of period \( j \), \( j = 1,2 \)

\[ \tau_{d_{(i,j)}} \] = the direct effect of the treatment administered in period \( j \) of group \( i \)

\[ \lambda_{d_{(i,j-1)}} \] = the effect of carry-over of the treatment administered in period \( j - 1 \) of group \( i \), where \( \lambda_{d_{(i,0)}} = 0 \)

\[ \varepsilon_{ikj} \] = random error for subject \( k \) in period \( j \) in group \( i \).

The subject effects, \( \epsilon_{ik} \), are assumed as random independent and identically distributed Normal(0, \( \sigma^2_{\epsilon} \)); the random errors, \( \varepsilon_{ikj} \), are assumed as Normal(0, \( \sigma^2 \)); other effects are fixed (Jones and Kenward 1989; pp. 9, 10, 22, 23).

The normal theory mixed effects ANOVA for a 2,2,2 crossover study has three single degree of freedom contrasts: one for unequal carryover effects; one for unequal period effects; and one for unequal treatment effects. The contrast for unequal carryover effects is assessed against a between-subject error term; the other two are compared to within-subject variability (Jones and Kenward 1989; pp. 30-33). The ANOVA results for our study are shown in Table 1.

---

**Table 1.** Normal theory mixed effects ANOVA results.
An insightful presentation of the ANOVA includes a discussion of confounding. In the 2,2,2 crossover study design, the unequal carryover, treatment-by-period interaction, and subject group (or sequence) effects are confounded in the single degree of freedom sum of squares which is tested against the between-subject error term (Jones and Kenward 1989; pp. 39-51). Some authors recommend against formally testing for the unequal carryover effect (Senn 1993; pp. 10-15, Senn 1997; pp. 239-247). And, whether or not to treat the subject effect as either a fixed or random effect is a topic of debate.

The assessment of drug interaction is made by way of the average bioequivalence criterion using the single degree of freedom contrast for unequal treatment effects. A 90% confidence interval is constructed based upon the observed estimate of the difference, OCD-OC, in the average log transformed values. The geometric mean ratio (OCD/OC) estimate and the corresponding 90% confidence interval are calculated by exponentiation of the log scale results. A summary of the interaction analysis is shown in Table 2 and Figure 6. In Figure 6, the y-axis is labeled with antilog values. The target bioequivalence (interaction) limits are indicated by the horizontal dotted lines located at 0.8 and 1.25 on the y-axis.

| Variable | Source          | df | SS   | MS   | F     | p    |
|----------|-----------------|----|------|------|-------|------|
| Log EE AUC | Between-Subject | 1  | 0.0472 | 0.0472 | 0.43  | .52  |
|          | Carryover       |    | 2.2002 | 0.1100 |       |      |
|          | B-S Error       | 20 | 0.2546 | 0.2546 | 26.44 | <.0001|
|          | Within-Subject  |    | 2.0460 | 0.0246 | 2.35  | .13  |
|          | Treatment       | 1  | 0.0246 | 0.0246 |       |      |
|          | Period          | 1  | 0.1226 | 0.0096 |       |      |
|          | W-S Error       | 20 | 0.1226 | 0.0096 |       |      |
|          | Total           | 43 | 2.7191 |       |       |      |
| Log EE Cmax | Between-Subject | 1  | 0.1956 | 0.1956 | 1.60  | .22  |
|          | Carryover       |    | 2.4393 | 0.1220 |       |      |
|          | B-S Error       | 20 | 0.0318 | 0.0918 | 3.97  | .060 |
|          | Within-Subject  |    | 0.0400 | 0.0400 | 1.73  | .20  |
|          | Treatment       | 1  | 0.4820 | 0.0231 |       |      |
|          | Period          | 1  | 0.1956 | 0.1956 | 1.60  | .22  |
|          | W-S Error       | 20 | 0.1956 | 0.1956 |       |      |
|          | Total           | 43 | 3.2288 |       |       |      |
| Log NET AUC | Between-Subject | 1  | 0.0442 | 0.0442 | 0.36  | .56  |
|          | Carryover       |    | 2.4381 | 0.1234 |       |      |
|          | B-S Error       | 20 | 0.0004 | 0.0004 | 0.10  | .75  |
|          | Within-Subject  |    | 1.0147 | 1.0147 | 26.50 | <.0001|
|          | Treatment       | 1  | 0.0791 | 0.0040 |       |      |
|          | Period          | 1  | 0.1047 | 1.0147 |       |      |
|          | W-S Error       | 20 | 0.0791 | 0.0040 |       |      |
|          | Total           | 43 | 2.6985 |       |       |      |
| Log NET Cmax | Between-Subject | 1  | 0.1780 | 0.1780 | 1.19  | .29  |
|          | Carryover       |    | 2.0389 | 0.1490 |       |      |
|          | B-S Error       | 20 | 0.2396 | 0.2396 | 8.57  | .019 |
|          | Within-Subject  |    | 0.0173 | 0.0173 | 0.47  | .50  |
|          | Treatment       | 1  | 0.7293 | 0.0386 |       |      |
|          | Period          | 1  | 0.0173 | 0.0173 | 0.47  | .50  |
|          | W-S Error       | 20 | 0.7293 | 0.0386 |       |      |
|          | Total           | 43 | 4.1631 |       |       |      |
Table 2. Summary of normal theory interaction analyses.

| Variable      | Geometric Means | Geometric Mean Ratio (OCD/OC) | 90% CI for True Ratio |
|---------------|-----------------|------------------------------|-----------------------|
| In EE AUC     | 3.393.8         | 1.16                         | (1.11, 1.23)          |
| In EE Cmax    | 339.5           | 1.10                         | (1.01, 1.19)          |
| In NET AUC    | 169.672.0       | 1.01                         | (0.97, 1.04)          |
| In NET Cmax   | 20.779.3        | 0.86                         | (0.78, 0.95)          |

Figure 6. Point Estimates and 90% Confidence Intervals – Normal Theory. The y-axis is labeled with antilog values. The target bioequivalence (interaction) limits are indicated by the horizontal dotted lines located at 0.8 and 1.25 on the y-axis.

Table 1, Table 2, and Figure 6 indicate that there is no clinically meaningful interaction for either log EE AUC, log EE Cmax, or log NET AUC, as the corresponding 90% confidence intervals are contained in the target bioequivalence interval of [0.8, 1.25]. However, there is some evidence that log EE AUC ($p < 0.0001$) and log EE Cmax ($p = 0.060$) are elevated when Drug D is given with the oral contraceptive, as both of the lower limits of the corresponding 90% confidence intervals exceed 1.0. There is a clinically meaningful decrease in log NET Cmax ($p = .019$) as the lower limit of the corresponding 90%
confidence interval is less than 0.8 which is outside the target bioequivalence interval of [0.8, 1.25]. In addition, the upper limit is less than 1.0. Given the outlying log EE Cmax data for Subjects 2 and 20 and the outlying log NET Cmax data for Subject 9 shown previously in Figures 2, 3b, 3d, 5b, and 5d, it is not surprising that the lengths of the 90% confidence intervals for log EE Cmax and log NET Cmax are longer than the corresponding confidence intervals for log EE AUC and log NET AUC. The period effect for log NET AUC which was suggested by Figure 4 is confirmed (p < .0001) in Table 1.

An interesting teaching point is to demonstrate that each of the single degree of freedom $F$-tests is equivalent to a two sample t-test comparing the subjects in one treatment sequence with the subjects in the other treatment sequence with respect to linear combinations of the two data values collected for each subject (Jones and Kenward 1989; pp. 22-28). For example a difference between the two groups of subjects, sequence 1 (OCD then OC) minus sequence 2 (OC then OCD) in our study, in the mean within-subject sum of log AUC (or Cmax) values indicates an unequal carryover effect. To test for unequal period effects, compare the mean within subject treatment difference, OCD-OC, of the log AUC (or Cmax) values. To test for unequal treatment effects, compare the mean within subject period difference, period 1 - period 2 in our study, of the log AUC (or Cmax) values. This equivalence can be used to visually evaluate the appropriateness of the normal theory ANOVA (Section 5.2) and to motivate and develop the distribution-free analysis of the 2,2,2 crossover design (Section 5.3).

**Graphics for Visualizing the Data in the ANOVA Contrasts**

To visualize and assess nonnormality, heteroscedasticity, and outliers for the data which were input into the two sample t-test procedure outlined above, plot the two sample distributions of sums or differences side-by-side. A difference in location of the two distributions suggests an effect. Further details can be found in Pikounis, Bradstreet, and Millard (2001).

For example, Figure 7a portrays the single degree-of-freedom $F$-test evaluating the treatment effect comparison (p < .0001) for log EE AUC. The period effect (p < .0001) observed for log NET AUC is shown in Figure 7b. Figure 7c displays the lack of carryover effect (p = .52) for log EE AUC.
**Figure 7.** Sum and Difference Plots. The y-axes are labeled with log values. The y-axes are all different. The horizontal bars represent sample means. (a) Treatment effects; (b) Period effects; (c) Carryover effects.

In **Figure 8a**, an outlier (Subject 9) is evident in the treatment effect evaluation ($p = .019$) for log NET Cmax. And some degree of heteroscedasticity may be suggested by **Figure 8b** which shows the treatment effect evaluation ($p = .060$) for log EE Cmax.
Figure 8. Sum and Difference Plots. The y-axis is labeled with log values. The y-axes are both different. The horizontal bars represent sample means. (a) Outlier; (b) Possible heteroscedasticity.

To further describe the three single degree of freedom contrasts, plot the mean log response of each treatment by study period. Connect the mean log responses across periods with a line either by like treatments or by treatment sequence, and label the y-axis with anti-log values. Then evaluate three characteristics of the plot: 1) the relative ordering of the two treatment means within each study period; 2) the magnitude of the difference between the two treatment means within each period; and 3) the magnitude of any difference in means between periods in each treatment group. For example, Figure 9a represents the treatment effect ($p < .0001$) observed for log EE AUC, and Figure 9b represents the period effect ($p < .0001$) observed for log NET AUC.
Figure 9. Mean Log Response Plots. The y-axis is labeled with antilog values. The y-axes are both different.
(a) Treatment effects; (b) Period effects.

Diagnostic Plots: Normal Probability, Model Fit, Residuals

Three types of diagnostic plots assess the appropriateness of the normal theory model for the 2,2,2 crossover design: normal probability plots, scatter plots of observed vs. fitted values, and scatter plots of studentized residuals vs. fitted values. These diagnostic plots are somewhat routine in concept, but there is an interesting twist in their construction for the 2,2,2 crossover design. Because the residuals within a subject sum to zero, the residuals from the two treatment periods have the same magnitudes and opposite signs. No additional information is gained from the second set of residuals. Therefore, plot the residuals from only one of the two treatment periods. We used information from the first treatment period.

Normal Probability Plots: Figure 10 presents normal probability plots for the log scaled AUC and Cmax variables for EE and NET. Given a sample size of twenty-two and the results on random sampling from a normal distribution presented by Daniel and Wood (1980; pp. 33-43), a normal theory model for log EE AUC (Figure 10a) and log NET AUC (Figure 10c) is not unreasonable. The convex curvature and
"shelving" of the internally studentized residuals for log EE Cmax in Figure 10b could be of some concern. The outlier (Subject 9) in Figure 10d suggests that a normal theory analysis is probably not best for log NET Cmax.

Figure 10. Normal Probability Plots. The x- and y-axes are labeled with log values. The x-axes are all different. The y-axes are all different. (a) log EE AUC; (b) log EE Cmax; (c) log NET AUC; (d) log NET Cmax.

Observed vs. Fitted Values: Figure 11 presents plots displaying the observed and corresponding fitted values on the log scale. These figures suggest that the normal theory model is a reasonable approach except for log NET Cmax (Figure 11d) where some convex curvature and an outlier (Subject 9) suggest otherwise.
Figure 11. Observed vs. Fitted Values. The $x$- and $y$-axes are labeled with log values. The $x$-axes are all different. The $y$-axes are all different. (a) log EE AUC; (b) log EE Cmax; (c) log NET AUC; (d) log NET Cmax.

Studentized Residuals vs. Fitted Values: Figure 12 presents plots displaying the studentized residuals and corresponding fitted values on the log scale. These figures do not suggest any easily recognizable variance pattern. Perhaps, Figure 12b suggests an unknown covariate which was not accounted for in the analysis, as the pattern of points appears to move downward to-the-right. However, the signal for such an unknown covariate does not appear to be overwhelming.
5.3 Distribution-Free Analysis of a 2,2,2 Crossover Study

Given the suggestion of a modest degree of nonnormality for log EE Cmax, a sample size of only 22, and the presence of a few discordant outliers, most notably for log NET Cmax and log EE Cmax, a distribution-free statistical analysis of these data may be prudent.

A distribution-free statistical analysis of a 2,2,2 crossover study consists of a series of three Wilcoxon rank sum tests analogous to the three two sample t-tests outlined in Section 5.2. For example, unequal treatment effects are evaluated using the Wilcoxon rank sum test to compare the two treatment sequence groups with respect to the within-subject period difference in the log AUC (or Cmax) values. Drug interaction is assessed by way of the average bioequivalence criteria using the two sample Hodges-Lehmann point estimator and the two sample Moses confidence interval (see Hollander and Wolfe 1973, pp. 68-82; Koch 1972; Jones and Kenward 1989, pp. 51-59). Note that the results from the calculation of the Hodges-Lehmann estimator and Moses confidence interval on the log scale must be divided by two before exponentiation to arrive at the correct estimate and confidence interval. This is because the distribution-free comparison, sequence 1 (OCD then OC) minus sequence 2 (OC then OCD) in our study, between the two groups of within-subject paired differences (period 1 - period 2 in our study) of
log AUC (or Cmax) values essentially counts the treatment difference twice. That is, the distribution-
free calculations give (OCD - OC) - (OC - OCD) = 2(OCD - OC), but we are interested in only OCD -
OC. No similar issue arises for the contrast using the Wilcoxon Rank Sum Test. A few timely remarks
about the scale invariance (Mood, Graybill, and Boes 1974, p. 336) of the Wilcoxon Rank Sum Test and
the lack of scale invariance of both the Hodges-Lehmann point estimator and Moses confidence interval
would be valuable here. And it is useful to point out that the Moses confidence interval will not
necessarily be symmetric about the Hodges-Lehmann point estimate.

Table 3 and Figure 13 present a summary of the distribution-free analyses. In Figure 13, the results of
the normal theory analysis are shown as a reference set. The y-axis is labeled with anti-log values.

| Variable      | Hodges-Lehmann Point Estimate (OCD/OC) | (At Least) 80% Moses CI for True Ratio | Wilcoxon Rank Sum Test p-Value |
|---------------|----------------------------------------|---------------------------------------|-------------------------------|
| In EE AUC     | 1.16                                   | (1.11, 1.24)                          | <.0001                        |
| In EE Cmax    | 1.08                                   | (0.99, 1.20)                          | .15                           |
| In NET AUC    | 1.00                                   | (0.97, 1.04)                          | .05                           |
| In NET Cmax   | 0.90                                   | (0.82, 0.96)                          | .032                          |

Table 3. Summary of distribution-free interaction analyses.
The conclusions drawn from the normal theory analyses of the absence of a clinically meaningful drug interaction on log EE AUC, log EE Cmax, and log NET AUC, are also suggested by the distribution-free analyses. Again it is clear that log EE AUC is elevated ($p < .0001$) to some degree when Drug D is given. The evidence that log EE Cmax is elevated ($p = .15$) is less impressive, as the lower limit of the 90% confidence interval, 0.99, is now just below 1.0. This difference between the normal theory and distribution-free analysis may simply be attributable to the modest degree of nonnormality shown in the normal probability plot for log EE Cmax (Figure 10b). The period effect for log NET AUC is confirmed again ($p = .0002$).

The most notable difference between the distribution-free and normal theory analyses is observed for log NET Cmax, where Subject 9 is a rather extreme outlier. The point estimate for the true ratio is now greater, 0.90 vs. 0.86, and the confidence interval, (0.82, 0.96), is now somewhat shorter (length = 0.14 as compared to length = 0.17 for the normal theory analysis) and it is contained completely within the target bioequivalence interval of [0.80, 1.25]. So, the original conclusion of a clinically meaningful reduction in log NET Cmax arrived at through the normal theory analysis is now reversed based upon the distribution-free analysis to a conclusion of no clinically important interaction. However, it is still clear that log NET Cmax is reduced ($p = .0032$) to some degree when Drug D is given concomitantly as the upper confidence limit, 0.96, is still below 1.0. And one must wonder, what proportion of the population does Subject 9 represent?

5.4 Software
For guidance on computing the statistical analyses of the 2,2,2 crossover design and the accompanying complement of graphics in S-PLUS, see Pikounis, Bradstreet, and Millard (2001). SAS users should consult Morris and Bradstreet (2000).

6. Getting the Data

The file `ocdrug.dat.txt` contains the raw data. The file `ocdrug.txt` is a documentation file containing a brief description of the dataset.

Appendix - Key to Variables in `ocdrug.dat.txt`

| Columns | Description |
|---------|-------------|
| 1-2     | Female Subject Number (1 to 22) |
| 4       | Treatment Sequence (1 = Drug D, placebo; 2 = placebo, Drug D) |
| 6       | Study Period (1, 2) |
| 8       | Treatment (0 = placebo, 1 = Drug D) |
| 10-15   | EE - AUC (pg*hr/ml) |
| 17-19   | EE - Cmax (pg/ml) |
| 21-28   | NET - AUC (pg*hr/ml) |
| 30-34   | NET - Cmax (pg/ml) |

Values are delimited by blanks.

Acknowledgements

The authors would like to thank the referees and the associate editor for their comments, Ms. Laurie Rittle for her careful typing of this article, and Ms. Cindy White for producing the graphics.

References

Bradstreet, T. E. (1991), "Some Favorite Datasets from Early Phase of Drug Research," in Proceedings of the Section on Statistical Education, American Statistical Association, pp. 190-195.

Bradstreet, T. E. (1992), "Favorite Datasets from Early Phases of Drug Research - Part 2," in Proceedings of the Section on Statistical Education, American Statistical Association, pp. 219-223.

Bradstreet, T. E. (1993), "Statistical Applications in the Pharmaceutical Industry - Part III: Phase I and II Clinical Pharmacology Studies," Stats, 10, pp. 20-23.

Bradstreet, T. E. (1994), "Favorite Datasets from Early Phases of Drug Research - Part 3," in Proceedings of the Section on Statistical Education, American Statistical Association, pp. 247-252.

Bradstreet, T. E. (2000), "Food and Drug Interaction: What Role Does Statistics Play?" The College Mathematics Journal, 31, pp. 268-273.

Bradstreet, T. E., and Dobbins, T. W. (1996), "When Are Two Drug Formulations Interchangeable?"
Teaching Statistics, 18, pp. 45-48.

Bradstreet, T. E., and Liss, C. L. (1995), "Favorite Datasets from Early (and Late) Phases of Drug Research - Part 4," in Proceedings of the Section on Statistical Education, American Statistical Association, pp. 335-340.

Bradstreet, T. E., and Short, T. H. (2001), "Favorite Datasets From Early Phases of Drug Research - Part 5," in Proceedings of the Section on Statistical Education, American Statistical Association, CD-ROM.

Chow, S.-C., and Liu, J. P. (2000), Design and Analysis of Bioavailability and Bioequivalence Studies (2nd ed., revised and expanded), New York: Marcel Dekker.

Daniel, C., and Wood, F. S. (1980), Fitting Equations to Data (2nd ed.), New York: John Wiley and Sons.

Drug Information Journal (1995), 29(3).

Food and Drug Administration (1999), "Draft Guidance for Industry: Average, Population, and Individual Approaches to Establishing Bioequivalence" [Online]. (www.fda.gov/cder/guidance/guidance.htm)

Food and Drug Administration (2000), "Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products - General Considerations" [Online]. (www.fda.gov/cder/guidance/guidance.htm)

Gibaldi, M., and Perrier, D. (1982), Pharmacokinetics (2nd ed.), New York: Marcel Dekker.

Hollander, M. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley and Sons.

Jones, B., and Kenward, M. G. (1989), Design and Analysis of Cross-Over Trials, New York: Chapman and Hall.

Journal of Biopharmaceutical Statistics (1997), 7(1).

Koch, G. G. (1972), "The Use of Non-Parametric Methods in the Statistical Analysis of the Two-Period Change-Over Design," Biometrics, 28, pp. 577-584.

Mood, A., Graybill, F. A., and Boes, D. C. (1974), Introduction to the Theory of Statistics (3rd ed.), New York: McGraw-Hill.

Morris, A., and Bradstreet, T. E. (2000), "From Data to Study Report in One Step: A SAS™ Macro for the 2,2,2 Crossover Design," in Proceedings of PharmaSUG 2000, pp. 317-321.

Pikounis, B., Bradstreet, T. E., and Millard, S. P. (2001), "Graphical Insight and Data Analysis for the 2,2,2 Crossover Study," in Applied Statistics in the Pharmaceutical Industry With Case Studies Using S-PLUS, eds. S.P. Millard and A. Krause, New York: Springer-Verlag, pp. 153-179.

Senn, S. (1993), Cross-over Trials in Clinical Research, New York: John Wiley and Sons.
Senn, S. (1997), *Statistical Issues in Drug Development*, New York: John Wiley and Sons.

*Statistics in Medicine* (2000), 19(20).

Thomas E. Bradstreet
Merck Research Laboratories
Blue Bell, PA 19422
USA
thomas Bradstreet@merck.com

Deborah L. Panebianco
Merck Research Laboratories
Blue Bell, PA 19422
USA
deborah Panebianco@merck.com