Design of Experiments and Statistical Analysis

of

Biological Studies

by

Virgil L. Anderson and Robert A. McLean

Technical Report #82-13

Department of Statistics
Purdue University

May 1982

Design of Experiments and Statistical Analysis of Biological Studies

by

Virgil L. Anderson and Robert A. McLean

I. Background

One approach to designing experiments in the last decade considers three essential ingredients of a well designed experiment expressed in the book by Anderson and McLean (1974). These three ingredients, in order of importance, are:

A. Inference Space

B. Randomization

C. Replication.

Inference space is a phrase to replace the term "population", usually used by statisticians. It has been our experience that the term "inference space" demands more attention from the research worker.

The phrase "inference space" means the limits to which the investigator may use the results of the experiment. Common practice at present is for the research worker to indicate how extensively he wishes the results to apply before the experiment is set up. This requires that he define the experimental units (such as "white rats of a certain size" to represent all rats of this type) he is to use in his research and that will be the basis for the inferences. He must also define the time interval and the geographical extent to which he wishes the results to apply, and then decide which levels of all factors he wants controlled

in the experiment. Ordinarily more time should be spent on this phase of designing the experiment than either of the other two (randomization and replication) because without the inference space clearly defined the best so called "designed experiment" may be worthless to the investigator. We define this ingredient (inference space) as a part of the designed experiment. Hence there can be no "best designed experiment" without a carefully defined inference space.

Randomization is the next most important ingredient in designing experiments. It must be present in the experiment for probability statements to be made. Fisher (1960), p. 17, expressed the idea that it is the physical basis of the validity of the test. It is also the basis of validity of confidence intervals.

Included in the ingredient, randomization, is another concept, "restriction" on randomization. To understand restriction on randomization let us first explain "completely randomized" which means no restriction on randomization. This concept can be seen by considering a factor, t, with five levels and three experimental units treated with each of the five levels of factor t completely at random.

Assume there are 15 randomly drawn experimental units from the inference space to be used for the entire experiment. One way to obtain a completely randomized design is to select a random number between 1 and 5, say 2. Then the first experimental unit must receive treatment 2 or the second level of factor t. Select another number between 1 and 5, say 5, then the second experimental unit must receive treatment 5. Continue sampling or drawing random numbers in this manner until the 15 experimental units have all been "treated". This sampling procedure

requires that each level of factor t is represented three times in the experiment and the design of this experiment is called "completely randomized".

The mathematical model for analyzing the data from such an experiment is

$$y_{ij} = \mu + T_i + \epsilon_{(i)j} : i = 1,2,...,5; j = 1,2,3 \quad (1)$$

where

y_{ij} = the response from experimental unit j treated with level i of factor t,

μ = overall mean

T_i = effect of the i^{th} level of factor t,

$\epsilon_{(i)j}$ = the experimental error caused by the j^{th} experimental unit nested in the i^{th} level of factor 1.

The assumptions for the analysis of the data in a model such as this are:

a) y_{ij} is a random variable,

b) the variances of the responses within levels of factor t are equal,

c) The model is additive,

d) The experimental error is NID$(0,\sigma^2)$, normal and independently distributed with mean zero and variance, σ^2.

This complete randomization assures the experimenter that the experimental units from the inference space have three (3) opportunities for selection for each treatment. These three experimental units for each treatment then allow a measure of the variation within the treatments.

It follows, then, that the test of significance on treatment effects (T_i in equation (1)) must be based on the excessive amount of variation among the means of the treatments over the amount of variation obtained from within the treatments where the variation within the treatments is accounted for by the variance due to $\epsilon_{(i)j}$ in equation (1). Hence in equation (1) we can indicate that $\epsilon_{(i)j}$ is the correct error or evaluating T_i.

If, however, the sampling procedure was such that the first random draw, level 2, was used on the first three experimental units; then the second draw, level 5, was used on the next three experimental units and so on, we would have a "restriction" on randomization because the randomization procedures were allowed only five times, not 15 as is required for complete randomization. If, as is frequently the case, there tends to be similarity between adjacent units in space or tiem, the variation within the group of three is smaller than the variation between the groups of three. In this design then, the variation due to treatments is not separable from the variation between groups of three units (treatments confounded with groups). If then there is a souce of error variation between groups, it will not be possibly to test for treatments, and the following model depicts this

$$y_{ij} = \mu + T_i + \delta_{(i)} + \epsilon'_{(i)j} \qquad (2)$$

where: $\delta_{(i)}$ is the "restriction" error or that random component due to the i^{th} group of units (note how the subscript is identical to the subscript to T_i indicating complete confounding), and

$\epsilon'_{(i)j}$ is NID$(0,\sigma^2_{\epsilon'})$,

the error due the j^{th} unit in the i^{th} group. It is the variation due

to the $\epsilon'_{(i)j}$ in equation (2) represents only a small portion across the inference space. Hence one needs an estimate of σ^2_δ to test for the effect of treatments. Of course $\delta_{(i)}$ has no degrees of freedom, which indicates this is a poor design and should not be used.

Using the algorithm for deriving the expected mean squares described in Chapter 2 of Anderson and McLean (1974), we can show that the correct error term for testing treatments is $\delta_{(i)}$. However, there is no estimate of this error in equation (2) unless the whole experiment is repeated. Hence the "restriction" on the randomization has caused $\delta_{(i)}$ which, in turn, tells the experimenter that this sampling procedure is not a good one even before the first observation is made. This, then, allows the experimenter to change his design early, before he has taken data that will not give a good analysis.

Before explaining the third ingredient of designed experiments, namely replication, a few words should be stated about special cases where randomization may not be required. If an investigator can show by actual experimentation that the results are the same whether randomization takes place or not it may not be necessary to randomize. This will happen on extremely well controlled experiments only. We are familiar with an example on a one cylinder engine gasoline consumption laboratory experiment in which it did not matter whether speeds were randomized or taken in order. The reason for this was that the controls on speed were so precise and the set-ups so repeatable that the errors were identical within the capabilities of the recording equipment. We have not experienced such control, however, in biological experiments.

With this non-random possibility in mind and knowing an experiment

cannot be designed well without knowing the inference space, we rank inference space above randomization in importance when considering ingredients of a well designed experiment.

Finally the third ingredient, replication, is quite often required for an estimate of an error term, or to provide the basis for making decisions on the importance of factors contributing to the response variables. In addition, as the number of observations increases on a given treatment the more precise the estimate of the effect of the treatment becomes, or the smaller its variance becomes. For example in estimating the mean of the response from a treatment, the variance of the estimated mean, $\bar{y}$, is

$$s^2_{\bar{y}} = \frac{s^2}{n}$$

where

s^2 = variance estimated from the sample

n = number of observations randomly obtained from the entire inference space for $\bar{y}$.

If, however, previous experimentation has shown certain information is available, e.g. the variance is known or that higher order terms in model are zero, it may not be necessary to replicate the entire experiment. In fact there are many good experiments run with fewer than the total number of combination of levels of the factors in a "factorial" experiment. These experiments are called "fractional replicated factorials" which will be described later.

With the various well-designed experiments without complete replications in mind, the ingredient "replication" is placed in third position

behind inference space and randomization.

Before going into designs and analyses it should be understood that the readers of this material should understand (1) basic statistical concepts, (2) distributions such as the mean $\bar{y}$, t, χ^2, F and (3) analysis of variance [ANOVA including expected mean squares (EMS)] and regression models.

II. Designs

A definition of a designed experiment is an arrangement of the experimental material, including randomization of experimental units to the treatments so that statistical tests of significance (and confidence intervals) on the effects and interactions of the factors being studied can be made. In order to accomplish this, care must be taken to set up the arrangement efficiently (keep the cost reasonably low) and, at the same time, cover the inference space. For coverage of designs, major headings are used to indicate designs encountered by us in bio-availability studes.

A. Block Designs

1. Importance of Blocking (Handling Extraneous Variables)

Many authors of books on design of experiments express the importance of "blocking", placing all treatments or all combinations of the levels of all factors of interest in a homogeneous group (thereby removing some of the effect of an extraneous variable from the experimental error) and repeating this group or block in time and/or space with different experimental units. To show this concept we use mathematical equations which are to be used as the basis for analyses of the data from the designed experiment.

Returning to the concepts of setting up equations (1) and (2) it follows that another design of the experiment could be to arrange three blocks of five treatments each, where the five experimental units were randomized onto the five treatments per block. Pictorially, one could have the following arrangement:

	Blocks	
Treatments	Treatments	Treatments
5	2	3
2	4	1
3	1	5
1	5	4
4	3	2

It follows that the equation to be used as the basis for the analysis is:

$$y_{ij} = \mu + B_i + \delta_{(i)} + T_j + \epsilon_{ij} \qquad (3)$$

where: y_{ij} and μ have the same meaning as they did for equation (2)

B_i = effect of block i

$\delta_{(i)}$ is similar to $\delta_{(i)}$ in equation (2)

T_j = effect of the j^{th} treatment

ϵ_{ij} = error due to the j^{th} treatment in block i (assuming there is is no interaction).

Since the experimenter is interested in testing for treatment effects only, this is an excellent design because ϵ_{ij} is the basis for the test of T_i since the variation across treatments is compared to the variation due to ϵ_{ij}'s. If it should turn out that the effects of B_i and $\delta_{(i)}$ are zero, B_i, $\delta_{(i)}$ and ϵ_{ij} may be pooled and equation (3) becomes equation (1). This completes the demonstration that, in general,

blocking is always worthwhile in experiments and should be used whenever possible.

To show the effectiveness of blocking, the following simple example is provided:

a. Block Effect Large

Blocks					
1		2		3	
Treatment	y	Treatment	y	Treatment	y
5	1	2	9	3	2
2	7	4	3	1	8
3	5	1	8	5	5
1	4	5	8	4	3
4	1	3	10	2	10
	18		38		28

Treatment:

1: 4 + 8 + 8 = 20

2: 7 + 9 + 10 = 26

3: 5 + 10 + 2 = 17

4: 1 + 3 + 3 = 7

5: 1 + 8 + 5 = 14

ANOVA

[Reference Table 5.1.3 p.129 of Anderson and McLean (1974)]

Source	df	SS	MS	F	EMS
Blocks	2	40.00	20.00		$\sigma^2+5\sigma^2_\delta+5\sigma^2_B$
δ	0	--	--		$\sigma^2+5\sigma^2_\delta$
Treatments	4	66.27	16.57	3.7 N.S.	$\sigma^2+3\phi(T)$
Error	8	35.33	4.42		σ^2

$F_{4,8}(.05) = 3.8$

N.S. means not significant at α=.05.

Analyzed as a completely randomized design (ignoring block effects).

Source	df	SS	MS	F	EMS
Treatments	4	66.27	16.57	2.20 N.S.	$\sigma^2+3\phi(T)$
Error	10	75.33	7.53		σ^2

b. Block Effect Zero

Blocks

1		2		3	
Treatment	y	Treatment	y	Treatment	y
5	3	2	7	3	2
2	9	4	1	1	8
3	7	1	6	5	5
1	6	5	6	4	3
4	3	3	8	2	10
	28		28		28

Treatment 1: 6 + 6 + 8 = 20

2: 9 + 7 + 10 = 26

3: 7 + 8 + 2 = 17

4: 3 + 1 + 3 = 7

5: 3 + 6 + 5 = 14

ANOVA

Source	df	SS	MS	F
Blocks	2	0	0	
δ	0	--	--	
Treatments	4	66.27	16.57	3.7 N.S.
Error	8	35.33	4.42	

$F_{4,8}(.05) \cong 3.8$

N.S. means not significant at α=.05.

Analyzed as a completely randomized design (pooling block and error effects)

Source	df	SS	MS	F
Treatments	4	66.27	16.57	4.7*
Error	10	35.33	3.53	

$F_{4,10}(.05) = 3.5$

*significant at α=.05

c. Conclusions

(i) The effectiveness of blocking is shown in case a. by observing the differences in the error mean square when the data are analyzed as a randomized complete block design as compared to a completely randomized design (4.42 versus 7.53). Note that the mean square for treatments is the same in both tables in case a. and that treatments are very close to being significant when analyzed properly, i.e., as a randomized complete block design.

(ii) The initial model used for case b. is

$$y_{ij} = \mu + B_i + \delta_{(i)} + T_j + \epsilon_{ij}$$

It is seen, however, that since $B_i + \delta_{(i)} = 0$ that one can now write

the model as

$$y_{ij} = \mu + T_j + \epsilon_{ij}$$

which gives the second ANOVA table where the F for treatments is now 4.7 as compared to 3.7 for the initial model.

(iii)In both of the above examples the data were analyzed using the models for the randomized complete block design and the completely randomized design. In the first case the effect of blocking is clear and illustrates the power of blocking. In the second case blocking was used to no avail but the indication of no block effects was seen in the ANOVA table and then by pooling block and error effects the power of the significance test was again increased.

The point of all of this is to demonstrate that one does not lose by blocking in the original experiment because if the block effects are very small one can pool to obtain an error term that would have been the same as if complete randomization had occurred.

2. Incorrect Use of Blocks (Treatments Used as Blocks)

In the previous example, the effect of $\delta_{(i)}$, the restriction error caused by blocks, did not decrease the importance of blocking because the experimenter was interested in the effect of blocking plus the restriction error in reducing the estimated experimental error, ϵ_{ij}. In that case, ϵ_{ij} was the basis for testing treatment effects only.

Now, one must consider the case in which the blocking concept is used incorrectly. Consider an example in manufacturing drugs in tablet form. The interest here was in studying disintegration of the tablets. The experimenter wanted to use five drugs compressed with three different pressures in the experiment.

The suggested design of the experiment was to use 15 batches of tablet mixture, one batch (or block) for each combination of the five

drugs and three pressures. After the tablets were compressed from each batch of about 500 tablets, 5 were to be pulled at random from each 500 tablet batch and the time to disintegration was to be measured in seconds. This procedure would be carried out for all 15 batches, making 75 tablets to be tested for disintegration time.

The analysis of the data may be summarized with the following degrees of freedom (df) and model:

df: $75 = 1 + 4 + 2 + 8 + 0 + 60$

model: $y_{ijk} = \mu + D_i + P_j + DP_{ij} + \delta_{(ij)} + \epsilon_{(ij)k}$

$i=1,2,3,4,5;\ j=1,2,3;\ k=1,2,3,4,5.$ (4)

where:

y_{ijk} = disintegration time of the k^{th} tablet with the j^{th} pressure using drug i,

μ = overall mean

D_i = the effect of the i^{th} drug,

P_j = the effect of the j^{th} pressure,

DP_{ij} = the effect of the interaction of drug i, at pressure j

$\delta_{(ij)}$ = restriction error caused by all of the tablets in the $(i,j)^{th}$ drug, pressure combination being manufactured under the same conditions. This is the only error appropriate for testing D, P and DP

and

$\epsilon_{(ij)k}$ = the experimental error caused by the k^{th} tablet in the $(i,j)^{th}$ cell.

Model (4) points out (by using $\delta_{(ij)}$) that one should <u>not</u> run the

experiment in this manner because there will be no degrees of freedom for the appropriate error to test drugs, pressures and their interaction. That error, of course, is the $\delta_{(ij)}$. This is a demonstration of using blocks incorrectly, that is, the blocks (batches) are completely confounded with treatments (drugs and pressures) of interest. One could not test for blocks separately in the previous example but one was not interested in blocks, per se, there. When the experimenter is interested in testing for treatments (drugs and pressures), the individual should not use treatments as blocks.

The correct way to handle this problem is to run at least two batches for each of the 15 combinations of drugs and pressures. Then take one tablet per batch (some experimenters may demand to take two but for this demonstration assume only one was used) and analyze the results of these tablets. The two batches provide the correct error for between "batch treatments" (drugs and pressures) and the inference space is sampled much more effectively because there are two samples (batches) from each bomvination rather than only one. The equation for the analysis of the data from this design is:

$$y_{ijk} = \mu + D_i + P_j + DP_{ij} + B_{(ij)k} + \varepsilon_{(ijk)} \qquad (5)$$

where: $i=1,\ldots,5$; $j=1,2,3$; $k=1,2$.

The analysis of variance is:

Source	df	f 5 i	f 3 j	r 2 k	EMS
D_i	4	0	3	2	$\sigma^2+\sigma_B^2+6\phi(D)$
P_j	2	5	0	2	$\sigma^2+\sigma_B^2+10\phi(P)$
DP_{ij}	8	0	0	2	$\sigma^2+\sigma_B^2+2\phi(DP)$
$B_{(ij)k}$	15	1	1	1	$\sigma^2+\sigma_B^2$
$\in_{(ijk)}$	0	1	1	1	σ^2

where:

the arrows (↖) indicate the F-tests, $D_i, P_j, DP_{ij}, \in_{(ijk)}$ are defined as for equation (4) and $B_{(ij)k}$ is the error caused by the k^{th} batch within treatment combination (ij), the correct error for testing D, P, and DP.

B. Designs with Treatments on Same Experimental Units

1. Repeated Measures

The following examples use time as the treatments on the same experimental units but this is not the only type of treatment used in bioavailability studies:

	Drugs(Fixed)			
	1	2	3	4
	People (Random)	People (Random)	People (Random)	People (Random)
Time (Fixed)	1 2	3 4	5 6	7 8
1				
2				
3				
4				
5				
6				

Quite frequently in this type experiment the research worker wants to compare the efficacy of only those four drugs used in the experiment on a population of people. To establish the experimental units in the inference space, the people may be defined in terms of age, sex, race and so on with the sample for the experiment assuemd to be random from the indicated population. The efficacy is then investigated over six periods of time.

One of the statistical dangers in this type designed experiment is that the assumption of no correlation between responses may not be met for the same person over the time periods. This has been recognized by statisticians and methods to handle the problem are discussed by various authors including Greenhouse and Geisser (1959), Cole and Grizzle (1966), Winer, p. 522-524 (1971), Anderson and McLean p. 166 and 167 (1974) and Elashoff (1981).

The model appropriate to analyze data from the drug experiment above if all assumptions are met is:

$$y_{ijk} = \mu + D_i + P_{(i)j} + \delta_{(ij)} + T_k + DT_{ij} + PT_{(i)jk} + \epsilon_{(ijk)} \qquad (6)$$

where:

y_{ijk} = response of the j^{th} person in the k^{th} time period using the i^{th} drug,

μ = overall mean

D_i = the effect of the i^{th} drug (fixed),

$P_{(i)j}$ = the effect of the j^{th} person using the i^{th} drug (random),

$\delta_{(ij)}$ = the restriction error caused by the same person responding over time, $NID(0,\sigma^2_\delta)$,

T_k = the effect of the i^{th} time period (fixed),

DT_{ik} = the effect of the interaction of the i^{th} drug in time period k,

$PT_{(i)jk}$ = the effect of the interaction of the j^{th} person using drug i in time period k,

and $\epsilon_{(ijk)}$ = the random error caused by the j^{th} person using drug i in time period k, $NID(0,\sigma^2)$.

The ANOVA for this model is:

Source	df	SS	MS	EMS
Drugs (D)	3	72.25	24.08	$\sigma^2+6\sigma^2_\delta+6\sigma^2_p+12\phi(D)$
People in D(P)	4	7.00	1.75	$\sigma^2+6\sigma^2_\delta+6\sigma^2_p$
δ	0	none		$\sigma^2+6\sigma^2_\delta$
*Time (T)	5	105.42	21.08	$\sigma^2+\sigma^2_{PT}+8\phi(T)$
*DT	15	38.25	2.55	$\sigma^2+\sigma^2_{PT}+2\phi(DT)$
*PT	20	15.00	0.75	$\sigma^2+\sigma^2_{PT}$
ϵ	0	none		σ^2

*Anderson and McLean p.166 (1974) show that if there is high correlation between observations from time to time within people, the df for testing Time and DT should be 1 and 3 respectively for the numerator of the F and 4 for the denominator. The actual mean squares calculated from the date should be used, however. Hence for testing hypotheses.

a) H_0: $\phi(T) = 0$

Use $F_{1,4} = \frac{\text{MS Time}}{\text{MSPT}}$,

and for

b) H_0: $\phi(DT) = 0$

Use $F_{3,4} = \frac{\text{MSDT}}{\text{MSPT}}$.

Let us consider a small experiment (for demonstration purposes only) of this type and show the calculations one usually follows to analyze the data.

	DRUGS								
	1		2		3		4		
	People		People		People		People		
Time	1	2	3	4	5	6	7	8	Totals
1	2	3	4	2	6	5	7	5	34
2	4	4	4	4	5	5	5	4	35
3	5	7	4	3	5	6	6	5	41
4	3	5	5	3	8	10	9	10	53
5	7	7	8	5	9	9	10	11	66
6	6	6	6	7	7	8	8	9	57
Totals	27	32	31	24	40	43	45	44	286
	59		55		83		89		

The first calculation should be to run a homogeneity of variance test on all the data unless the experimenter knows the distribution of the errors for the variable to be analyzed. In this case the investigator may immediately transform the data appropriately. For example, if it is known that the variable, y, is exponential in nature the log y may be run without bothering to investigate the homogeneity of the variances or the distributional properties of y.

For these data let us assume that the distributional peculiarities are not known. To recognize the two errors in the nested factorial design, one should run two different tests for homogeneity of variances as follows:

a) Between people within drugs

(In this case there are six time periods for each person. Hence there is a divisor of 6.)

For the four drugs

$s_1^2 = \frac{(27-32)^2}{6 \cdot 2 = 12} \cong 2.08$, $s_2^2 = \frac{(31-34)^2}{12} \cong 4.08$,

$s_3^2 = \frac{(40-43)^2}{12} = .75$, $s_4^2 = \frac{(45-44)^2}{12} = .08$

$$q_{(1,4)} = \frac{4.3264 + 16.6464 + 5624 + .0064}{(2.08 + 4.08 + .75 + .08)^2} \cong .44$$

$q_{(1,4)}(.01) = .92$ $\therefore$ accept homogeneity of the variances between people for the four drugs.

b) Interaction of People x Time for the four drugs:

Example: Drug 1

Time	People 1	2	Total
1	2	3	5
2	4	4	8
3	5	7	12
4	3	5	8
5	7	7	14
6	6	6	12
Total	27	32	59

Drug 1:

SS Time: $\frac{5^2+8^2+\ldots+12^2}{2} - \frac{59^2}{12} = \underline{23.42}$

SS People: $\frac{(27-32)^2}{12} = 2.08$

SS Total: $2^2 + \ldots + 6^2 - \frac{59^2}{12}$

$= 323 - 290.08 = 32.92$

SS Int^n: $32.92 - 25.50 = \underline{7.42}$

Drug 2:

SS Time: $\frac{6^2 + 8^2 + \ldots + 13^2}{2} - \frac{55^2}{12} = \frac{551}{2} - 252.08$

$= 22.42$

SS People: 4.08

SS Total: $285 - 252.08 = 32.92$

SS Int^n: $32.92 - 26.50 = \underline{6.42}$

Drug 3:

SS Times: $\frac{11^2 + \ldots + 15^2}{2} - \frac{83^2}{12} = 33.42$

SS People: .75

SS Total: $6^2 + \ldots + 8^2 - 574.08 = 36.92$

SS Int^n: $36.92 - .75 - 33.42 = \underline{2.75}$

Drug 4:

SS Time: $\frac{12^2 + \ldots + 17^2}{2} - \frac{(89)^2}{12} = 58.42$

SS People: .08

SS Total: $7^2 + \ldots + 9^2 - 660.08 = 62.92$

SS Int^n: $62.92 - .08 - 58.42 = \underline{4.42}$

$q_{(5,4)} = \frac{55.06 + 41.22 + 7.56 + 19.54}{(7.42 + 6.42 + 2.75 + 4.42)^2} = \frac{123.38}{(21.01)^2} \cong \underline{.280}$

$q_{(5,4)}(.01) = .498$ $\therefore$ Accept the homogeneity of the interaction variances:

Next if the number of people per drug were 5 or greater, one could run a Shapiro-Wilk W-test, Anderson and McLean (p.26, 1974), to test for normality. Here there are only 2 people per drug so no test for normality can be made.

Next one may calculate the sums of squares using the original data because no transformation is indicated from the preliminary test on the homogeneity of variance and there is no thoery to tell the experimenter that a transformation should be made.

Using equation (6) where i=1,2,3,4; j=1,2,; k=1,2,...,6, one can calculate the sums of squares for the sources as:

$$SSD_i = \sum_{i=1}^{4} \left[\frac{\left(\sum_{j,k}^{2,6} y_{ijk} \right)^2}{2 \cdot 6} \right] - \frac{\left(\sum_{i,j,k}^{4,2,6} y_{ijk} \right)^2}{4 \cdot 2 \cdot 6}$$

$$= \frac{59^2 + 55^2 + 83^2 + 89^2}{12} - \frac{286^2}{48}$$

$$= 1776.33 - 1704.08$$

$$= 72.25$$

$$SST_k = \sum_{k=1}^{6} \frac{\left(\sum_{i,j}^{4,2} y_{ijk} \right)^2}{4 \cdot 2} - \frac{\left(\sum_{i,j,k}^{4,2,6} y_{ijk} \right)^2}{48}$$

$$= \frac{34^2 + 35^2 + \ldots + 57^2}{8} - \frac{286^2}{48}$$

$$= 105.42$$

$$SSP_{(i)j} = \sum_{i,j}^{4,2} \left[\frac{\left(\sum_{k}^{6} y_{ijk} \right)^2}{6} \right] - \sum_{i=1}^{4} \left[\frac{\left(\sum_{j,k}^{2,6} y_{ijk} \right)^2}{2 \cdot 6} \right]$$

$$= \frac{27^2 + 32^2 + \ldots + 44^2}{6} - \frac{59^2 + 55^2 + 83^2 + 89^2}{12}$$

$$= 1783.33 = 1776.33$$

$$= 7.00$$

$$SSDT_{ij} = \left[\sum_{i,k}^{4,6} \frac{\left(\sum_{j=1}^{2} y_{ijk} \right)^2}{2} - CT \right] - SSV_i - SSP_{(i)j}$$

$$= \frac{5^2 + 6^2 + 11^2 + \ldots + 17^2}{2} - \frac{286^2}{48} - SSV_i - SS\ P_k$$

$$= 1920.00 - 1704.08 - 72.25 - 105.42$$

$$= 38.25$$

$$SSPT_{(i)jk} = \sum_{i} (SSPT(D_i))$$

where

$$SSPT(D_1) = \left[\sum_{i=1}^{4} y_{ijk}^2 - \frac{\left(\sum_{j,k}^{2,6} y_{ijk} \right)^2}{12} \right] - \left[\sum_{k=1}^{2} \frac{\left(\sum_{j=1}^{2} y_{ijk} \right)^2}{2} - \frac{\left(\sum_{j,k}^{2,6} y_{ijk} \right)^2}{12} \right]$$

$$- \left[\sum_{j=1}^{2} \frac{\left(\sum_{k=1}^{6} y_{ijk} \right)^2}{6} - \frac{\left(\sum_{j,k}^{2,6} y_{ijk} \right)^2}{12} \right]$$

$$= [2^2 + 3^2 + \ldots + 6^2 - \frac{59^2}{12}] - [\frac{5^2 + 8^2 + \ldots + 12^2}{2} - \frac{59^2}{12}]$$

$$- [\frac{27^2 + 32^2}{6} - \frac{59^2}{12}]$$

$$= 2.42$$

$$SSPT(D_2) = [4^2 + 2^2 + \ldots + 7^2 - \frac{55^2}{12}] - [\frac{6^2 + 8^2 + \ldots + 13^2}{2} - \frac{55^2}{12}]$$

$$- [\frac{31^2 + 24^2}{6} - \frac{55^2}{12}]$$

$$= 5.42$$

$$SSPT(D_3) = [6^2 + 5^2 + \ldots + 8^2 - \frac{83^2}{12}] - [\frac{11^2 + 10^2 + \ldots 15^2}{2} - \frac{83^2}{12}] - [\frac{40^2 + 43^2}{6} - \frac{83^2}{12}]$$

$$= 2.75$$

$$SSPT(D_4) = [7^2 + 5^2 + \ldots 9^2 - \frac{89^2}{12}] - [\frac{12^2 + 9^2 + \ldots + 17^2}{2} - \frac{89^2}{12}] - [\frac{45^2 + 44^2}{6} - \frac{89^2}{12}]$$

$$= 4.41$$

Therefore

$$SSPT_{(i)jk} = \sum_{i=1}^{4} SSPT(D_i) = 15.00$$

and

$$SS\ Total = \sum_{i,j,k}^{4,2,6} y_{ijk}^2 - \frac{\left(\sum_{i,j,k} y_{ijk}\right)^2}{48}$$

$$= 323 + 285 + 611 + 723 - 1704.08 \text{ (from our work above)}$$

$$= 237.92.$$

This completes the discussion on repeated measures type designs here. It must be understood that time is not the only treatment on the same experimental unit that an experimenter may encounter.

2. Cross Over Designs

Another design frequently used by bioavailability research people for noncurative drugs is the cross over design [Grizzle (1965)]. The basic structure of the design comes from a 2x2 square called a Latin Square because the treatments are represented by Latin letters inside the designated row and column combinations. A reference for the Latin Square design is Chapter 8 of Box, Hunter and Hunter (1978).

The overall structure of the cross over design is portrayed in the following layout:

Time of Administrating Drug	Groups of People: 1	Groups of People: 2
1	Drug A	Drug B
2	Drug B	Drug A

where it is assumed that a random group (Group 1) of people were given drug A and blood measures taken then a washout period was allowed before the same group was given Drug B and a final blood measure was taken. Similarly, another group (Group 2) was given Drug B first, measured and allowed a washout period, then given Drug A and measured again. Hence the two drugs (treatments) are used on the same person (experimental unit).

This method allows an evaluation of carryover effects by looking at Time in the analysis. The necessary assumption however, is that there is no interaction of Groups by Time because the effect of Drug A vs. Drug B is completely confounded with that interaction. To demonstrate this consider the following [assume the design is completely randomized with one observation per group so the concept of calculating interaction (sums of squares) may be shown]:

Time	Groups: 1	Groups: 2	Total
1	A = 25	B = 10	35
2	B = 7	A = 23	30
Total	32	33	65

ANOVA

Source	df	SS
Groups (G)	1	.25
Time (T)	1	6.25
Interaction GxT	1	240.25
Total	3	246.75

Groups SS: $(32-33)^2/4 = .25$

Time SS: $(35-30)^2/4 = 6.25$

Total SS: $25^2 + 7^2 + 10^2 + 23^2 - \frac{(65)^2}{4} = 246.75$

GxT SS: $246.75 - .25 - 6.25 = 240.25$

(A vs. B) SS: $(48-17)^2/4 = 240.25$

Since GxT SS = 240.25 and (A vs. B) SS = 240.25 this demonstrates the complete confounding of the interaction of Gruops x Time with (A vs. B) drug effect. Hence one must assume there is no interaction of GxT before he can interpret (A vs. B) drug effect in these designs.

The usual approach to analyze cross over designs utilizing the Latin Square and repeated measures concepts follows:

Example. An illustration of the cross over design occurred in a problem for a drug manufacturing company. Two formulas of a drug were given to 18 human beings selected at random from a population to which the drug would be given. Nine random subjects were given formula one and the other nine were given formula two. The amount of the drug in the blood of each of the 18 subjects was measured eight hours after administration of the drug and that amount was recorded.

After 7 days of "washout" period in which the drug supposedly was

excreted, another blood sample was taken and all 18 subjects had zero amount of the drug in their blood samples. The experimenter then had the group of nine who took formula one take formula two and vice versa for the other group. This part of the experiment is called the "cross over" part of the design.

A model for the analysis of this experiment is:

$$y_{ijk} = \mu + S_i + \delta_{(i)} + T_j + F_k + \epsilon_{(ijk)} \qquad (7)$$

$$i = 1,2,\ldots,18,\ j = 1,2,\ k = 1,2$$

where

y_{ijk} = amount of drug in the blood of the i^{th} subject after 8 hours given order j and formula k.

μ = overall mean

S_i = the effect of subject i

$\delta_{(i)}$ = the restriction error on the i^{th} subject due to all observations on orders and formulas coming from the i^{th} subject

T_j = the effect of the j^{th} time plus any effect due to restricting the randomization at the j^{th} time,

F_k = the effect of the k^{th} formula

$\epsilon_{(ijk)}$ = error, NID$(0,\sigma^2)$, estimated from the remaining effects assuming all interactions are zero.

The data from this experiment are given in the following tabulation:

	Subjects								
Time	1	2	3	4	5	6	7	8	9
				Formula One					(351.7)
1	51.9	35.1	38.6	36.1	34.6	39.7	37.8	38.8	39.1
	Seven Day Washout Period								
	Formula Two								
2	43.5	45.4	35.4	43.7	49.8	39.9	41.4	37.5	39.5
Total	95.4	80.5	74.0	79.8	84.4	79.6	79.2	76.3	78.6

	Subjects								
Time	10	11	12	13	14	15	16	17	18
	Formula Two								(332.0)
1	50.8	41.1	39.1	35.7	33.7	31.2	34.3	31.8	34.3
	Seven Day Washout Period								
	Formula One								(316.6)
2	44.2	33.4	32.7	33.1	33.4	27.1	33.1	39.5	40.1
Total	95.0	74.5	71.8	68.8	67.1	58.3	67.4	71.3	74.4

$$\text{SS Time} = \frac{(683.7)^2 + (692.7)^2}{18} - \frac{(1376.4)^2}{36} = \underline{2.25}$$

$$\text{SS Formula} = \frac{(668.3)^2 + (708.1)^2}{18} - 52{,}624.36 = \underline{44.00}$$

$$\text{SS Subjects} = \frac{(9514)^2 + (80.5)^2 + \ldots + (74.4)^2}{2} - 52{,}624.36 = \underline{717.79}$$

$$\text{SS Total} = (51.9)^2 + (35.1)^2 + \ldots + (40.1)^2 - 52{,}624.36 = \underline{1{,}093.98}$$

SS Error = SS total - SS Time - SS Formulas - SS Subjects

= 1,093.98 - 2.25 - 44.00 - 717.79 = 329.94.

The corresponding ANOVA is tabulated in the following.

Source	df	SS	MS	EMS
Subjects (S_i)	17	717.79	42.22	$\sigma^2+2\sigma^2_\delta+2\sigma^2_S$
Subject restriction error ($\delta_{(i)}0$	0	------	-----	$\sigma^2+2\sigma^2_\delta$
Time (T_j)	1	2.25	2.25	$\sigma^2+18\phi(T)$
Formulas (F_k)	1	44.00	44.00	$\sigma^2+18\phi(F)$
Residual ($\varepsilon_{(ijk)}$)	16	329.94	20.62	σ^2
Total	35	1,093.98		

To find the error with the most degrees of freedom from these data we may test for time

H_0: $\phi(T) = 0$

$F_{1,16} + \frac{2.25}{20.62} < 1$

Hence accept H_0 at $\alpha > 0.25$ and pool mean squares, which actually results in a RCBD where subjects act as blocks and formulas are treatments. The error mean square with 17 df is

$$\text{Error mean square} = \frac{(1)(2.25) + (16)(20.62)}{17} = \underline{19.54}$$

Subjects are random and no interest other than to provide the inference space. The test on formulas is

$$F_{1,17} = \frac{44.00}{19.54} = 2.25$$

Tabled $F_{1,17}$ $\alpha = 0.10 = 3.03$
$\alpha = 0.25 = 1.42$

Hence the effect of formulas is not significant at $\alpha = 0.10$ level.

Take the same example and consider the layout of the experiment to be:

	Groups	
	1	2
	Subjects	Subjects
Time	1 2. . .9	10 . . 18
1	Formula One	Formula Two
2	Formula Two	Formula One

One may think of this design in such a way that he can use the following model:

$$y_{ijk} = \mu + G_i + S_{(i)j} + \delta_{(ij)} + T_k + GT_{ik} + ST_{(i)jk} + \epsilon_{(ijk)} \qquad (8)$$

where

y_{ijk}, μ are the same as for equation (7)

G_i = the effect of the i^{th} group of subjects

$S_{(i)j}$ = the effect of the j^{th} subject in the i^{th} group

$\delta_{(ij)} = \delta_{(i)}$ of equation (7)

$T_k = T_j$ of equation (7)

GT_{ik} = the effect of the interaction of the i^{th} group by the k^{th} time and/or Formulas (confounded with effect of formulas)

or

$= F_k$ of equation (7)

$ST_{(i)jk} = \epsilon_{(ijk)}$ of equation (7) or the effect of the interaction of the i^{th} group by the k^{th} time

$\epsilon_{(ijk)}$ = error, NID(0,σ^2) (not estimable in this model)

The ANOVA for the analysis of the data using equation (8) is:

Source	df
Groups (G)	1
Subjects in Groups ($S_{(i)j}$)	16
$\delta_{(ij)}$	0
Time (T_k)	1
Formulas (GT_{ik})	1
$ST_{(i)jk}$	16
$\epsilon_{(ijk)}$	0

One can see that this ANOVA using Equation (8) explains the error ($ST_{(i)jk}$) to be used to test formulas and the details of the workings

of the design on formulas and time effects much better than the analysis from Equation (7), the traditional approach. The amalgamation of the two designs, Latin Square and repeated measures, becomes apparent in this presentation by showing that the interaction of groups by time is really represented by formulas (Latin Square); and time by subjects within groups is the real error for testing formulas (repeated measures).

Now we are ready to expand on a more recent concept in design of these experiments. A reference for this type design is Albert et al (1979). Consider an experiment in which three drugs A, B, and C are to be compared. All six sequences of A, B, and C may be considered when administering the three drugs to a given person. The minimum number of subjects one should consider in an experiment of this type is 12, two for each sequence.

If one measurement is taken for each subject-drug combination and a sufficient washout period is allowed between drugs, one possibly layout of the experiment is given in Table 1 (where subjects are assigned at random to the sequences):

TABLE 1.

	Sequence											
	1 Subjects		2 Subjects		3 Subjects		4 Subjects		5 Subjects		6 Subjects	
	1	2	3	4	5	6	7	8	9	10	11	12
1.	B	B	A	A	B	B	C	C	C	C	A	A
TIME	WASH OUT PERIOD →											
2.	C	C	B	B	A	A	A	A	B	B	C	C
	WASH OUT PERIOD →											
3.	A	A	C	C	C	C	B	B	A	A	B	B

Another possible layout of the experiment is given in Table 2.

TABLE 2.

Sequence

	1		2		3		4		5		6	
	Subjects		Subjects		Subjects		Subjects		Subjects		Subjects	
Drugs	1	2	3	4	5	6	7	8	9	10	11	12
A	t_3	t_3	t_1	t_1	t_2	t_2	t_2	t_2	t_3	t_3	t_1	t_1
B	t_1	t_1	t_2	t_2	t_1	t_1	t_3	t_3	t_2	t_2	t_3	t_3
C	t_2	t_2	t_3	t_3	t_3	t_3	t_1	t_1	t_1	t_1	t_2	t_2

In order to arrive at a satisfactory method of calculating the sums for the data from such an experiment let us consider the model for the layout given in Table 1 and compare that model with the appropriate model for Table 2.

The model with the corresponding degrees of freedom (df) for analyzing the data from Table 1 is:

df: $36 = 1 + 5 + 6 + 0 + 2 + 10 + 12 + 0$

model: $$y_{ijk} = \mu + Q_i + S_{(i)j} + \delta_{(ij)} + T_k + QT_{ik} + ST_{(i)jk} + \epsilon_{(ijk)} \quad (9)$$

where:

y_{ijk} = response from the k^{th} time, for subject j following sequence i (notice nothing is associated with the drug),

μ = overall mean

Q_i = effect of the i^{th} sequence (fixed)

$S_{(i)j}$ = effect of the j^{th} subject (random) following the i^{th} sequence

$\delta_{(ij)}$ = restriction error due to times and drugs on the same subject,

T_k = effect of the k^{th} time (fixed) plus any effect due to restricting the randomization at the k^{th} time,

QT_{ik} = effect of the interaction of the i^{th} sequence and k^{th} time (includes the drug effect) similar to GT_{ik} of equation (8) except that there are 2df for drugs here witha total of 10df for QT. Hence there are 8df remaining,

$ST_{(i)jk}$ = effect of the interaction the j^{th} subject following the i^{th} sequence by the k^{th} time, and

$\epsilon_{(ijk)}$ = error associated with the observation on the j^{th} subject following the i^{th} sequence at the k^{th} time; NID$(0,\sigma^2)$.

It follows that the corresponding degrees of freedom and model for analyzing the data from an experiment with the layout given in Table 2 is:

df: $36 = 1 + 5 + 6 + 0 + 2 + 10 + 12 + 0$

model: $$y_{ijk} = \mu + Q_i + S_{(i)j} + \delta_{(ij)} + D_k + QD_{ik} + SD_{(i)jk} + \epsilon_{(ijk)} \quad (10)$$

where:

y_{ijk}, μ, Q_i, $S_{(i)j}$, $\delta_{(ij)}$ and $\epsilon_{(ijk)}$ are the same as they are in equation (9),

D_k = effect of the k^{th} drug,

QD_{ik} = effect of the interaction of the i^{th} sequence by the k^{th} drug (imbedded in this component is the effect of time with 2df),

$SD_{(i)jk}$ = effect of the interaction of the j^{th} subject following the i^{th} sequence by the k^{th} drug.

Since the design can be interpreted as two Latin Square designs (Sequences 2, 1, 4 make up one Latin Square and Sequences 6, 3, 5 make up the other one), the terms in equations (9) and (10) may be broken up in a manner that allows one overall model. To show this, in equations (9) and (10) we need look only at the <u>within</u> subject effects because the sequences and subjects are identical in the two equations. From equation (9), $T_k + QT_{ik} = D_k + QD_{ik}$ from equation (10). It follows that $ST_{(i)jk}$ of equation (9) = $SD_{(i)jk}$ of equation (10). Now, if D_k is removed from QT_{ik}, the remainder is exactly equal to the remainder from equation (10) if T_k is removed from QD_{ik}.

As a result of all the above, model (11) and corresponding degrees of freedom (df) can depict a preferred model as follows:

$$\text{df:} \quad 36 = 1 + 5 + 6 + 0 + 2 + 2 + 8 + 12 + 0 \qquad (11)$$

$$\text{model:} \quad y_{ijk\ell} = \mu + Q_i + S_{(i)j} + \delta_{(ij)} + T_k + D_\ell + R_{ik\ell} + n_{ijk\ell} + \epsilon_{(ijk\ell)}$$

where:

y_{ijk}, μ, Q_i, $S_{(i)j}$, $\delta_{(ij)}$, T_k are defined as they are in equation (9),

$D_\ell = D_k$ in equation (10),

$R_{ik\ell}$ = effect of the remainder interaction of Q_i, T_k, and D_ℓ (fixed and testable),

$n_{ijk\ell}$ = error for testing T_k, D_k and $R_{ik\ell}$, composed of parts of interactions of $ST_{(i)jk}$ from equation (9) and/or $SD_{(i)jk}$ from equation (10),

and

$\epsilon_{(ijk\ell)}$ is the same as $\epsilon_{(ijk)}$ from equations (9) and (10).

To show how equation (11) can be evolved from data laid out by equations (9) and (10) the following examples are presented:

A. Layout for equation (9) is given in Table 3:

TABLE 3

Sequences

		1		2		3		4		5		6		
		Subjects		S		S		S		S		S		
		1	2	3	4	5	6	7	8	9	10	11	12	Total
TIME	1	3	B 2	5	A 7	3	B 5	5	C 6	7	C 4	4	A 1	52
	2	6	C 8	4	B 2	6	A 7	2	A 5	3	B 1	5	C 7	56
	3	5	A 4	5	C 8	9	C 6	4	B 3	6	A 4	2	B 5	61
Total		14	14	14	17	18	18	11	14	16	9	11	13	169

where: Letters A, B, C indicate the drugs used.

(1) SS Sequences: $\frac{28^2 + 31^2 + \ldots + 24^2}{6} - \frac{(169)^2}{36} = 17.80$

(2) SS Subjects in Sequences: $\frac{(14-14)^2}{6} + \frac{(14-17)^2}{6} + \ldots + \frac{(11-13)^2}{6} = 11.83$

(3) SS Time: $\frac{52^2 + 56^2 + 61^2}{12} - \frac{(169)^2}{36} = 3.39$

(4) SS Sequences χ time (from Table 3):

Sequences

		1	2	3	4	5	6	Total
TIME	1	5	12	8	11	11	5	52
	2	14	6	13	7	4	12	56
	3	9	13	15	7	10	7	61
Total		28	31	36	25	25	24	169

SS Subtotal: $[\frac{5^2 + 12^2 + \ldots + 7^2}{2} - \frac{(169)^2}{36} = 98.14]$

SS Subtotal - SS Sequences - SS Time

SS Sequences x Time = 98.14 - 17.80 3.39 = 76.95

(5) SS Subjects in Sequences by Time:

i. SS In Sequence 1 (Subjects x time) (from Table 3):

Subjects

		1	2	Total
TIME	1	3	2	5
	2	6	8	14
	3	5	4	9
Total		14	14	28

$$3^2 + 2^2 + 6^2 + 8^2 + 5^2 + 4^2 - \frac{(28)^2}{6} - \frac{(14-14)^2}{6} - \left[\frac{5^2 + 14^2 + 9^2}{2} - \frac{(28)^2}{6}\right] = 3.00$$

ii. SS Sequence 2 (Subjects x Time)

$$5^2 + 7^2 + 4^2 2^2 5^2 + 8^2 - \frac{(31)^2}{6} - \frac{(14-17)^2}{6} - \left[\frac{12^2 + 6^2 + 13^2}{2} - \frac{(31)^2}{6}\right] = 7.00$$

iii. SS Sequence 3 (Subjects c Time) = 7.00

iv. SS Sequence 4 (Subjects c Time = 4.00

v. SS Sequence 5 (Subjects x Time) = 0.333

vi. SS Sequence 6 (Subjects x Time) = 10.333

SS Subjects in Sequence by Time = 3.00 + 7.00 + 7.00 + 4.00 + .333 + 10.333 = 31.67

From all these calculations, one obtains Table 4:

TABLE 4

ANOVA using equation (9)

Source	df	SS	MS
Sequence Q	5	17.80	3.56
Subjects in Sequence (S)	6	11.83	1.97
δ	0	-----	----
Time (T)	2	3.39	1.70
*QxT	10	76.95	----
**Drugs	2	63.39	31.70
Remainder (QxT)	8	13.56	1.70
SxT	12	31.67	2.64
Total	35	141.64	

b. Layout for equation (10) is given in Table 5 which is the data table from the design given in Table 2. The sums of squares for Sequence, Subjects in Sequence, Time, Drugs, and (Subjects by Time) in Sequence which may be called (Subjects by Drugs) in Sequence are the same. To show the calculations of the sums of squares when drugs replace time one may set up Table 5:

TABLE 5

Sequence

	1		2		3		4		5		6		
	Subjects		Subjects		Subjects		Subjects		Subjects		Subjects		
Drugs	1	2	3	4	5	6	7	8	9	10	11	12	Total
A	5	4	5	7	6	7	2	5	6	4	4	1	56
B	3	2	4	2	3	5	4	3	3	1	2	5	37
C	6	8	5	8	9	6	5	6	7	4	5	7	76
Total	14	14	14	17	18	18	11	14	16	9	11	13	169

*QxT is really a mixture of many sources of variation because of the Latin Square peculiarities of this design (Refer to Table 8.1.3, p.215 of Anderson and McLean (1974)) but the one source of interest is drugs. (Refer to Table 1 for the original design of the experiment and to Table 3 for the data arrangement associated with drugs A, B, and C).

**SS Drugs: $\frac{(\Sigma A)^2 + (\Sigma B)^2 + (\Sigma C)^2}{12} - \frac{(\Sigma \text{ all})^2}{36} = \frac{56^2 + 37^2 + 76^2}{12} - \frac{(169)^2}{36} = 63.39$

Sequence

Drug	1	2	3	4	5	6	Total
A	9	12	13	7	10	5	56
B	5	6	8	7	4	7	37
C	14	13	15	11	11	12	76

SS Sequence x Drugs:

SS Subtotal - SS Sequence - SS Drugs

= 98.14 - 17.80 - 63.39 = 16.95

The analysis of variance for this new arrangement of the responses is given in Table 6:

TABLE 6

ANOVA using equation (10)

Source	df	SS	MS
Sequence (Q)	5	17.80	3.56
Subjects in Sequence (S)	6	11.83	1.97
δ	0	0	
Drugs (D)	2	63.39	31.70
*QxD	10	16.95	1.70
Time	2	3.39	1.70
Remainder (QxD)	8	13.56	2.64
**SxD	12	31.67	
Total	35	141.64	

*QxD is really a mixture of many sources of variation because of the Latin Square peculiarities of this design. In fact after removing time from this source the remainder has the same SS as for the remainder (QxT) in Table 4. This remainder has DxT, QxT, QxD, QxTxD pieces of interactions.
**SxD is also a mixture of SxD, SxT and SxDxT because of the Latin Square arrangement.

It is seen that Table 4 and Table 6 give the same analysis. The preferred way to write the ANOVA is given in Table 7:

TABLE 7

ANOVA (Preferred)

Source	df	SS	MS
Sequence (Q)	5	17.80	3.56
Subjects in Sequence (S)	6	11.83	1.97
δ	0	-----	
Times (T)	2	3.39	1.70
Drugs (D)	2	63.39	31.70**
Remainder Interaction	8	13.56	1.70
Error	12	31.67	2.64
Total	35	141.64	

In general for this type problem one should use equation (11), which will lead to the correct analysis as demonstrated in Table 7. The results of Table 7 could be further analyzed using an accepted multiple comparison test on the drug means.

If the number of sequences is too large for the experimenter to run he may make use of a balanced incomplete block design (BIBD) Westlake (1974). Two other approaches to evaluate bioequivalence are (1) to use confidence bands on differences of means, Westlake, (1979), and (2) to use a Bayesian methods, Selwyn, Dempster and Hall (1981). Mandallaz and Mau (1981) compare different methods for decision making in bioequivalence evaluation.

Kirkwood (1981) presents a case against Westlake's bioequivalence method and Westlake (1981) responds. A note by the editor follows those two comments.

A supplement to the Biometrics March 1982 journal covers topics in Biostatistics and Epidemiology in honor of Jerome Cornfield.

C. Other Designs

Bioavailability research personnel should also be aware of designs of experiments that may be used when many factors must be evaluated simultaneously. Only factors with two levels are described here to give the reader a synopsis of the techniques. A detailed coverage of designs with different number of levels of the factors and with more than two equal number of levels of factors is given in Chapters Box, Hunter and Hunter (1978).

For demonstration purposes only consider an experiment in which there are two factors (a,b) each at two levels (0,1), where 0 represents the lower level and 1 represents the upper level of each factor. One possible layout for this completely randomized design is

		b = 0	b = 1
a	0	00 or (1)	01 or b
	1	10 or a	11 or ab

which allows the reader to use two notations to describe the factorial treatment combinations. In this case 00, 01, 10, 11 represents the treatment combinations (ab) as

00 = low level of a, low level of b

01 = low level of a, high level of b

10 = high level of a, low level of b

11 = high level of a, high level of b.

The second notation (1), b, a, ab defines the treatment combinations in the same order identically as the other notation. Notice that this notation, however, uses (1) when all factors are at the low level, the letter is missing when the low level of one factor is used in the treatment combination and the letter is present when the high level of the factor is present.

The ANOVA of the data from such an experiment is shown in Table 8.

TABLE 8

ANOVA of 2x2

Source	df
A	1
B	1
AB	1
Total	3

The usual method for calculating sums of squares is

$$\text{SSA: } \frac{(a + ab)^2 + ((1) + b)^2}{2} - \frac{(a + ab + (1) + b)^2}{4} ,$$

which is identically equal to

$$\frac{(a + ab - (1) - b)^2}{4} ,$$

a much more efficient procedure.

It follows that SSB and SSAB may be calculated directly as:

$$\text{SSB: } \frac{(b + ab - (1) - a)^2}{4},$$

$$\text{SSAB: } \frac{((1) + ab - a - b)^2}{4} .$$

This idea of using plus and minus signs on the yields comes from expecting the high levels of the factors to have the higher yields. It does not matter, however, if the yields actually reverse because the SS will still account for the effects of the treatments in all cases and all estimates of the treatment effect will have the correct sign. The nicety of this method of obtaining SS is that interactions may be calculated directly from the data, not subtracting the main effects (m.e.) SS from

the subtotal SS to obtain the two factor interaction (2 f.i.) SS as is the usual practice. It must be understood, however, that this differencing process is only applicable for two-leveled factorials. Whenever the number of levels is three or larger one must resort to the usual procedure to find the SS.

Table 9 summarizes the plus, minus notation.

TABLE 9

Treatment Combinations and Effects in 2^2

	Treatment Combination			
Effect	(1)	a	b	ab
A	-	+	-	+
B	-	-	+	+
AB	+	-	-	+
Mean	+	+	+	+

Notice that the mean uses all pluses and, of course, would require a division of the total yield by 4 here. Also notice that AB signs are obtained by multiplying the signs of A by those of B for each column or treatment combination. This methodology generalizes for any number of factors if all have two levels.

The most important value of this notation is to allow easy investigation of experiments with several factors in which the number of treatment combinations is too large to be run under one condition. This situation demands blocks of treatment combinations that do not contain all the combinations for the experiment or "incomplete" blocks. Also if the number

of factors is large enough and certain higher factor interactions can be assumed zero, fewer than all the treatment combinations may be run and still allow a very good experiment. This type design is called a "fractional factorial" design.

For demonstration purposes let us show the incomplete block and fractional factorial concepts using 2^3 factorial. If only 4 of the 8 treatment combinations can be run under homogeneous environmental conditions, there must be 2 blocks of four treatment combinations in each block. The experimenter must recognize that for this condition, the df for blocks must be confounded with one of the seven main effects or interactions from the factorial because each of A, B, AB, C, AC, BC, and ABC has one df and adds to the total of 7 df. In other words there must be a loss of information if fewer than all treatment combinations are run at one time or place.

The experimenter must now decide which of the 7df to lose. Since ABC should be of least interest let us confound ABC with blocks. Table 10 shows the plus, minus notation one may use to select the correct treatment combination for each block when ABC is confounded with blocks.

TABLE 10

Treatment Combinations and Effects in 2^3

	Treatment Combinations							
Effects	(1)	a	b	ab	c	ac	bc	abc
A	-	+	-	+	-	+	-	+
B	-	-	+	+	-	-	+	+
C	-	-	-	-	+	+	+	+
ABC(multiply)	-	+	+	-	+	-	-	+

To obtain the SS for ABC one subtracts the 4 "minus" treatment combinations yields from the 4 "plus" treatment combination yields, squares that difference and divides by 8. If blocks are to be confounded with ABC that is, the exact information on ABC is to be mixed up with that of blocks, the SS for blocks will be calculated identically to that of ABC. It follows that the make up of one block is the minuses of ABC and the pluses of ABC must be in the other block. Hence from Table 10 one can construct the two blocks to be run in the experiment as:

Blocks	
1	2
minus	plus
(1)	a
ab	b
ac	c
bc	abc

The ANOVA for the analysis of the data from this experiment is given in Table 11:

TABLE 11

ANOVA (2^3 in 2 blocks of 4, ABC confounded)

Source	df
Blocks and/or ABC	1
δ	0
A	1
B	1
AB	1
C	1
AC	1
BC	1
Total	7

Notice that there is no error for testing the m.e.'s and 2 f.i.'s in this case. That is the reason this experiment should be considered a demonstration of the technique.

Of course, if a few treatment combinations (at least one) were repeated in each block a few df could be avilable for an error estimate.

Using this same 2^3 experiment for demonstration purposes one can show how a fractional factorial experiment can be constructed. If Block 2, | a, b, c, abc | , is used, only four treatment combinations instead of 8 would make up a 1/2 replicate or fractional factorial. To find out what effects may be estimated from only those four treatmetn combinations Table 12 is set up.

TABLE 12

Treatment Combinations and Effects

for 1/2 replicate of 2^3

	Treatment Combinations			
Effects	a	b	c	abc
A	+	-	-	+
B	-	+	-	+
AB	-	-	+	+
C	-	-	+	+
AC	-	+	-	+
BC	+	-	-	+
ABC	+	+	+	+
Mean	+	+	+	+

Notice that the mean = ABC, A = BC, B + AC and C = AB.

These equal effects may be indicated as aliases or complete confounded with each other. Further the mean may be called I or identity and one has I = ABC. If we multiply each side of the equality by A, we obtain

$$A = A^2BC,$$

but the levels of the factors are only 0 or 1 and any even number becomes 0 or any odd number becomes 1 (modulo 2). Hence

$$A^2 = A^0 = 1$$

and

$$A^2BC = BC,$$

or

$$A = BC.$$

Similarly for $B = AB^2C = AC$ and $C = ABC^2 = AB$.

The ANOVA for this 1/2 fractional factorial of 2^3 is given in Table 13.

TABLE 13

ANOVA (1/2 replicate of 2^3)

Source	df
A and/or BC	1
B and/or AC	1
C and/or AB	1
Total	3

It must be understood that this is NOT a good design and is given here to demonstrate a methodology for handling fractional factorials.

A good example of a fractional factorial is a $1/4^{th}$ replicate of a

2^9 where the factors are a,b,c,d,e,f,g,h, and j. If three factor and higher interactions are zero, the effects can be identified as:

$$I = ABCDEF = DEFGHJ = ABCGHJ$$

$$A = BCDEF = ADEFGHJ = BCGHJ$$

$$AB = CDEF = ABDEFGHJ = CGHJ$$

$$ABC = DEF = ABCDEFGHJ = GHJ$$

Hence, in general, m.e.'s are confounded only with 5 f.i. and higher: 2 f.i.'s are confounded with 4 f.i.'s and higher and 3 f.i.'s (assumed zero) are confounded with other 3 f.i.'s and higher.

This allows the following ANOVA shown in Table 14.

TABLE 14

ANOVA ($1/4^{th}$ replicate of 2^9)

Source	df
m.e.	9
2 f.i.	$\binom{9}{2} = 36$
(Residual) error	82
Total	127

A booklet listing almost all useable 2^n fractionals with numerators one is "Fractional Factorial Experiment Designs for Factors at two Levels" by the Statistical Engineering Laboratory of the National Bureau of Standards Applied Mathematics Series 48 (1957).

Another design that is occasionally of interest to bioavialability research people is the composite design. This design allows estimation of curvature in experiments having many factors whereas the fractional

factorial of 2^n designs does not. The composite design is built around the factorial (or fractional) and center points are used on all the faces of the hypercube from the factorial plus one point in the center of the entire design.

If there are k factors, then there are only

$$2^k + 2k + 1$$

treatment combinations needed for the whole experiment. This is a tremendous savins relative to 3^k factorial (the smallest complete factorial allowing curvature estimation) when $k \geq 4$. For example if $k = 5$

$$2^5 + 2.5 + 1 = 43$$

whereas

$$3^5 = 243.$$

Both of these designs allow estimates of all linear and quadratic m.e. plus the linear x 2 f.i.'s. A thorough coverage of this design is given in Chapter 13 of Anderson and McLean (1974).

One final type of experimentation that is encountered in drug bioavailability studies is that called mixture experiments. The basic indicator of this type of experiment is the fact that all factors of interest must add to a constant. The usual constant is 100 percent and the factor levels are certain percentages of the total. An illustration of a few treatment combinations for an experiment involving factors A, B, and C are as follows (the levels of A, B, and C are percentages):

Treatment Combination	A	B	C
1	10	40	50
2	12	40	48
3	10	35	55
4	12	35	53

Note that in order to lay out a 2 x 2 factorial experiment involving factors A and B that the level C has to be adjusted so that the total composition is 100 percent.

Designs for this type of experiment are quite involved and will not be included here. One type of design that is good when the number of factors is no greater than 5 is the extreme vertices design, the details of which are given in Chapter 13 of Anderson and McLean (1974). A good summarization of the existing literature on mixture experiments is given in Cornell (1979).

III. References

1. Albert, K.S., Brown, S.W., Jr., DeSante, K.A., DiSanto, A.R., Stewart, R.D. and Chen, T.T., "Double Latin Square Study to Determine Variability and Relative Bioavailability of Methylprednisolon", Jour. of Pharm. Sci., Vol. 68, No. 10, October 1979.

2. Anderson, V.L. and McLean, R.A. "Design of Experiments: A Realistic Approach", Marcel Dekker, 1974.

3. Box, G.E.P. Hunter, W.G. and Hunter, J.S. "Statistics for Experimenters", Wiley, 1978.

4. Cole, J.W.L. and Grizzle, J.E. "Applications of Multivariate Analysis of Variance to Repeated Measurements", Biometrics, Vol. 22, No. 4, 810-828, Dec. 1966.

5. Cornell, J.A., "Experiments with Mixtures: An Update and Bibliography" Technometrics, Vol. 21, No. 1, 95-106, February 1979.

6. Fisher, R.A. "The Design of Experiments", Seventh Edition, Hafner, 1960.

7. Greenhouse, S.W. and Geisser, S."On methods in the analysis of profile data", Psychometrika, 24, 95-112, 1959.

8. Grizzle, J.E. "The Two-Period Change-Over Design and its Use in Clinical Trials", Biometrics Vol. 21, p. 467, 1965.

9. Kirkwood, T.B.L. "Bioequivalence Testing - A Need to Rethink". Biometrics, Vol. 37, No. 3, 589-591, September, 1981.

10. Mandallaz, D. and Mau, J. "Comparison of Different Methods for Decision Making in Bioequivalence Assessment", Biometrics, Vol. 37, No. 2, 213-222, June, 1981.

11. Selwyn, M.R., Dempster, A.P. and Hall, N.R. "A Bayesian Approach to Bioequivalence for a 2x2 Changeover Design", Biometrics, Vol. 37, No. 1, 11-22, March, 1981.

12. Westlake, W.J. "The Use of Balanced Incomplete Block Designs in Comparative Bioavailability Trials", Biometrics Vol. 30, p. 319, 1974.

13. Westlake, W.J. "Statistical Aspects of Comparative Bioavailability Trials", Biometrics Vol. 35, p. 273, 1979.

14. Westlake, W.J. "Response (to Kirkwood Sept. 1981 Note)", Biometrics, Vol. 39, No. 3, 591-593, September, 1981.

15. Winer, B.J. "Statistical Principles in Experimental Design", Second Edition, McGraw Hill, 1971.