ON PLANNING REPEATED MEASURES EXPERIMENTS

by

Walter T. Federer, Cornell University

BU-909-M July 1986

Abstract

Literature on repeated measures design deals almost entirely with construction of plans and statistical procedures for analyzing results. Little has appeared on methods for determining the length of treatment period, methods for eliminating carry-over effects, nature of responses during and after a treatment period, nature of pre- and post-treatment, desirability of using the same treatment repeatedly or alternating treatments, and choice of a model for a particular variable in a particular repeated measures experiment. Criteria for selecting a specified repeated measures design may be based on other than statistical considerations. The philosophical part of planning experiments and selecting a repeated measures design requires considerable strengthening in order to utilize present theory in an intelligent manner. Examples which depict several of the above problems and situations are discussed. The examples were encountered during the course of statistical consulting over the years and arose from diverse subject matter areas and situations.

ON PLANNING REPEATED MEASURES EXPERIMENTS

Walter T. Federer, Cornell University

1. Introduction

One of the most important aspects of an experiment using a repeated measures design is the planning of such an experiment. There is a paucity of literature on the topic; the literature starts with some specific plan and a specific response model equation and proceeds from there. In contemplating the use of a repeated measures design for a forthcoming experiment, it is expedient to specify

a) the nature of a treatment as to whether it is a single, intermittent, or continuous application for a given period,
b) the types of response to a treatment through time,
c) whether or not there will be pre-treatment, post-treatment, or/and a wash-out period measurement(s),
d) a response model equation for the experiment,
e) the types of treatment effects anticipated,
f) the length of a period,
g) the admissibility of treatment sequences,
h) the basis for optimality,
i) the number of stages or sets of treatments to be applied to the experimental units, and
j) whether or not a design needs to be constructed to meet the requirements of the investigator.

Failure to adequately consider the above points may doom the experiment to failure or to providing only partial information. A discussion with a number of examples is presented to exemplify the above points.

2. Nature of Response to a Treatment

The treatment design for each experiment often requires unique considerations. One of the first items to consider is the type of treatment application. Is the treatment applied once, intermittently, or continuously? In treating a headache, a single application may be given, e.g. two aspirin tablets. In studying the effect of a diet, the diet is given every day. In experiments involving irrigation, the water may be applied every three weeks throughout the growing season. The last intermittent treatment application is common to many medical treatments.

For a repeated measures experiment, a measurement of treatment response may be monitored continuously, or it may be obtained at the end of a period. In Figure 1 various types of positive treatment responses are depicted. Note that the negative effects are also possible. In Figure 1(a), the treatment may be applied at the time marked 1, the beginning of period 2. If the treatment is effective only when applied and is applied continuously, then the solid response is indicated. Or, the treatment may be applied at the time marked 1 but is effective for the length of the period. The response would again follow the solid line. If the treatment was applied only once at time 1 and if its effect is immediate but then diminishes, the response could follow the dotted line, indicating no effect at time 2.

In Figure 1(b), the response builds up to a maximum at the end of time 2 but then diminishes slowly thereafter. This type of effect (dotted line) is called a carry-over or residual effect. If a measurement was recorded at time 3, there would still be a carry-over effect. If the measurement was taken at time 4, there would be no carry-over effect. In Figure 1(c), the maximum response is obtained at time 3. If the measurement was at time 2, the maximum response would have been missed. In Figure 1(d), the solid line response indicates an immediate effect of the treatment but the response continues or is permanent. For example, in treating patients, this would indicate a cure of the ailment. This effect was noted in a diet study in that the diet destroyed the bacteria in the intestine and the digestibility was greatly reduced. The effect lasted for six months to three years after the diet was discontinued, whereas the length of the treatment period was only two weeks. In educational studies, the effect of a treatment may be long-lasting if not permanent. The dotted line would indicate that maximum response was not obtained until time 2 and that the effect was permanent after time 2.

The nature of a treatment response may determine whether or not a completely randomized repeated measures design, CRRM, or a change-over repeated measures design, CORM, is to be used. For the former, the variance, among subjects or units within treatments, σ_B^2, will be a measure of error variance for treatment differences. For the CORM where two or more treatments are applied to the same subject or unit in sequence, the variance within subjects forms a measure of an error variance, σ_W^2, for treatment differences. The latter variance is usually much smaller, i.e. one-fourth to one-half, than σ_B^2. Regardless of this, CRRM design may be required.

3. Pre-, Post-, and/or Washout-Periods

The experimental material should be treated uniformly before the experiment begins. This is the pre-treatment period. Responses may or may not be obtained for the pre-treatment period or periods. As an example, patients were to be used in an experiment in which an asthma attack was to be induced and then treated. In order to ascertain that the patients were suitable subjects for inducing an asthma attack, responses on the patients were obtained for two times (periods) using a placebo, i.e. a light asthma attack was induced but no treatment was given. Thus, two pre-treatments periods were used. In this experiment, the patient did not know which treatment was given but the doctor did in periods 1 and 2, i.e. the study was singly blind. When the treatments were applied in test-periods 3, 4, 5, and 6, neither the doctor nor the patient knew which treatment was given, i.e. the experiment was doubly blind. The treatments were believed to relieve an asthma attack.

In the above experiment, the treatment was applied and the response measured immediately.

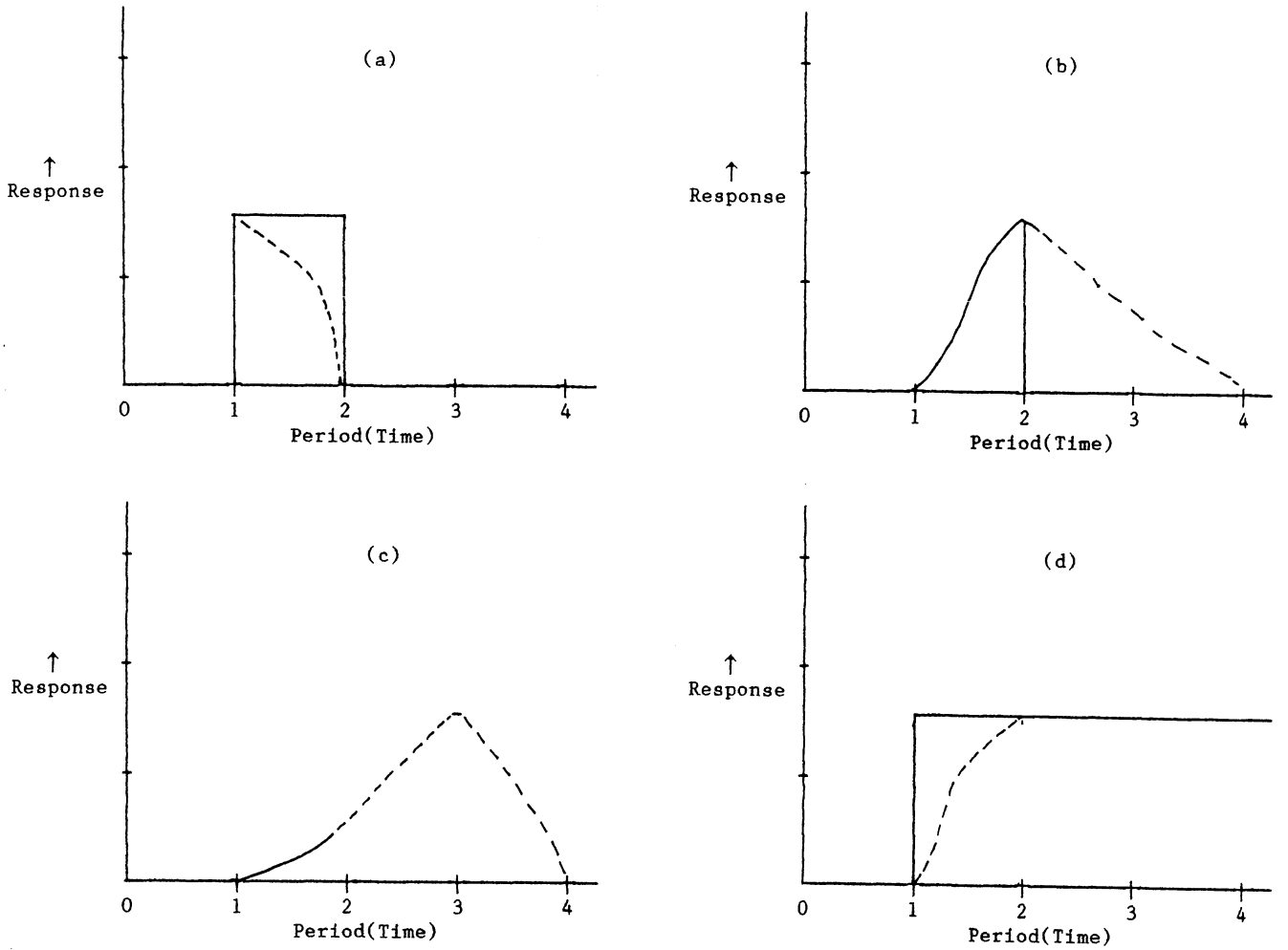

Figure 1. Types of treatment response.

The patient then left and did not return for a month. There was a month's time between treatments. This could be considered as a time or period in which the patient would return to normal life, i.e. a washout-period. There should be no carry-over or lasting effects from the previous treatment in this experiment. The purpose of washout-periods is to eliminate the effect of the previous treatment. During the washout-period all experimental material should be uniformly treated. The nature of the treatment response must be known in order to ascertain that the effect of the previous treatment is eliminated. A long-lasting or permanent effect is not eliminated in a washout-period. If these effects are present, a CRRM design would be indicated.

Post-treatment period responses are often obtained to study drifts or changes in response from beginning to end of the experiment, given that a pre-treatment-period response was also obtained. Both pre-treatment and post-treatment responses may be used as covariates for CRRM designs but not for CORM designs. In order for a CRRM design with p periods to be more efficient than a CORM design with p periods the squared correlation between pre- or/and post-treatment responses and test period responses, say ρ^2, must be greater than one minus the ratio of the error variances, i.e.

$$\rho^2 > 1 - \sigma_W^2/\sigma_B^2 .$$

This means, e.g., that if $\sigma_W^2/\sigma_B^2 = 1/4$, then the ρ must be greater than $\sqrt{.75} = .87$.

The use of washout-period and post-treatment period responses can have a serious disadvantage in lengthening the time in which an experiment is conducted. In experiments involving human subjects, the longer the experiment the greater the likelihood of losing subjects because of illness, subjects dropping out of the study, or drift in responses. If conditions during pre-treatment and during washout-periods differ, this may induce an interaction with treatments. Unless this can be taken care of in the analysis of data from the experiment, the results could be affected and lead to misinterpretation of data.

4. Type of Treatment Responses

In a simple comparative experiment for which a randomized complete or Latin square design would be appropriate and for which treatment responses are not measured through time (repeated measures), there is one type of treatment

main effect. However, in RM experiments there may be many types of effects. This increases the conceptual ideas of treatment effects by an order of magnitude. Explanations and interpretations become much more difficult and complex. Subject matter specialists often must reorient their entire way of thinking to accommodate this complexity. These difficulties should not be overlooked by a consulting statistician. They are real and must be appropriately handled in order to make effective use of RM designs.

The types of responses depicted in Figure 1 depict some types of effects. For completeness sake, we discuss a number of treatment effects below. A direct effect of a treatment has been defined as the response of a treatment in the period in which it is applied. Note from Figure 1, that the magnitude of an effect is time-dependent. If the dotted line in Figure 1(a) were the response to a single application of a treatment at time 1, its magnitude would depend entirely upon when the response was measured.

A one-period carry-over treatment effect has been defined as the effect of the treatment in the first period following the period in which a treatment was applied. Here again the magnitude of this effect is time dependent. A two-period carry-over treatment effect has been defined as the effect of a treatment in the second period following the period in which the treatment was applied. Likewise a k-period carry-over treatment effect has been defined as the treatment effect in the k*th* period following the period of application. A continuing or permanent treatment effect is the unchanging effect of a treatment effect in all the periods following the period in which the treatment was applied. The sum of the direct plus all the residual effects is called the cumulative treatment effect. It has also been called a permanent effect but here permanent and continuing effect are considered to be the same. Note again that the effects defined above are all time dependent and that this must be taken into account in the interpretation of the results from RM experiments.

5. Response Model Equation

The standard response model equation for CRRM designs states that there is an effect for a treatment. Occasionally, a direct effect and, a one-period residual effect will be postulated. For CORM designs, a subject, a period, a direct effect of treatment, and a one-period carry-over treatment effect are almost universally postulated. Computer software packages use this response model. The acceptance of a single response model for all CRRM or CORM designs can be inappropriate and incorrect. In fact, in any given experiment, each variable measured may have a different response model. There is no such thing as the linear model. The best that can be said is that there is a linear model which will be assumed appropriate for this experiment and this variable.

Note that in a particular experiment in which measurements are taken on a number of variables, some variables may show no one-period, no two-period, or no continuing or permanent effect of a treatment. Other variables may show both a one-period and a two-period carry-over effect. For other variables, it may be necessary to transform the data in order to have a linear model. All these points should be carefully considered in planning a RM experiment. The experiment design used must be such that the effects in the model can be estimated. If the experimenter is uncertain which response model is appropriate, he should postulate a set of plausible response models and select a design which allows estimation of the effects in the set of models (see Kershner and Federer, 1981). The nature of the treatment response greatly affects the choice of a response model equation. For responses as depicted in Figure 1(a), solid line, only a direct effect of a treatment is required. The length of the period is not important. This is not true for the remaining types of responses depicted in Figure 1. A response model should take into account all types of treatment effects present in the experiment.

A usual response model for a CORM designed experiment is

$$Y_{ghij} = \mu + \beta_g + \gamma_h + \delta_i + \pi_j + \varepsilon_{ghij},$$

where Y_{ghij} is the response of the i*th* treatment in the g*th* period and from the h*th* subject, μ is an effect common to all observations, β_g is a g*th* period effect, γ_h is a h*th* subject effect, δ_i is a direct effect of the i*th* treatment, π_j is the one-period carry-over effect of treatment j, and ε_{ghij} is a random effect. The ε_{ghij} are identically and independently distributed random variables with mean zero and variance σ^2_ε. The treatment and period effects are fixed and the subject effects γ_h are random.

The above response model is appropriate for a single measurement of response for a treatment applied for one period. Usually responses are measured at the end of the period, or the sum of all the observations in a period is the response. For example, in a diet study a measure of digestibility may be taken at the end of a two-week period for a given diet. In a marketing experiment, the total sales of a product, say apples, for a one week period may be the response measured.

An unresolved problem is how to handle continuous responses for a treatment during the periods. For example, digestibility could be measured daily in a two-week period in which the subject is on a specified diet. In Figure 1, suppose that the time intervals between the numbers represent two weeks and that daily measurements are taken. What type of analysis would be appropriate for data of this nature? Above it was stated that the sum was used for the apple experiment. It is suggested that the following procedure be used. First determine an appropriate response function for the treatments in the RM experiment. Second for each experimental unit ghij, fit the response function to the data and obtain estimates of the parameters of the function, e.g. intercepts, slope, change in slope, etc. and use the above response model for each of these in a univariate analysis of variance. Alternatively, the set of p parameter estimates could be used in a multivariate analysis of variance to determine what linear combinations of the estimates best discriminate among the treatments. Whether the former or/and

the latter are used will depend upon the nature of the experiment and the goals of the experimenter.

6. Length of Treatment Period

The choice of the length of treatment period is an extremely important aspect of any RM experiment. Literature is almost devoid of discussion of this topic. The choice of period length can greatly affect the magnitude and presence/absence of the various treatment effects (See Figure 1). It is imperative that the nature of treatment response be known in order to adequately select an appropriate length of period. The response model will be very much dependent upon the period length. In certain cases the period length can be selected to eliminate the carry-over effects of previous treatments. Note that a period length which eliminates carry-over effect can be undesirable if the experimenter is interested in the nature and magnitude of these effects. Elimination of carry-over effects is not always a good idea. Also, as depicted in Figure 1, the period length must be long enough to obtain an expression of maximum treatment response. The length could be, and often is affected by the particular treatment. Since the time to maximum response may vary from treatment to treatment in a particular experiment, the length of period for the experiment should be long enough to obtain a maximum response for the treatment requiring the longest period. Measurements for each treatment would be made at the time when maximum response occurs. Then, analyses would involve maximum response and time to maximum response. The latter variable could be studied via a linear trend measurement. Differences in trends would indicate differences in time to maximum response.

To illustrate some of the above, consider some examples. The first one relates to a marketing experiment on sales of apples in stores in a grocery chain. At the time the experiments were conducted, customers were purchasing apples about once a week. Hence, if the length of period was one day, there could be no carry-over effect of sales on Monday, say, on sales on Tuesday. This is true because customers buying apples on Monday would not buy apples on Tuesday, Wednesday, Thursday, Friday, or Saturday. Hence, their purchase of apples on a given day would not affect the sales of apples for five to six days following date of purchase. There could be no carry-over effect of previous treatment. Now if the treatment period were one week or even two, a purchase of a large quantity of apples would affect purchases in succeeding weeks and a carry-over effect of treatment would be present. Both period lengths were used in marketing studies on apples and on other commodities.

A second example illustrates a different effect. In comparing prices on roses sold in supermarkets, it was found that the first purchase of roses in a supermarket set the price level of roses in a supermarket for a customer who did or did not purchase the roses. Thus, there was a continuing effect of the first treatment in a store. The period length here was one week. Now if the period length had been one day this effect probably would not be observed since there is a different set, mostly, of customers in a store on different days of the week.

In a third example on the effect of different fiber content diets on many body responses of healthy young males, it was necessary to determine the length of period during the course of the experiment. The experiment was carried according to the plan in Figure 2. It was originally designed as periods 2,4,6, and 8 for subjects 1 to 12 and as periods 1,2,3, and 4 for subjects 13 to 24, because it was believed that the length of period would have to be 21 to 25 days in order for a diet to stabilize for a subject. The experiment had to be conducted within a limited time frame of about 11 weeks.

		Subject											
Period	Week	1	2	3	4	5	6	7	8	9	10	11	12
1	1	A	B	C	D	A	B	C	D	A	B	C	D
	2	A	B	C	D	A	B	C	D	A	B	C	D
2	3	B	A	D	C	D	C	B	A	C	D	A	B
	4	B	A	D	C	D	C	B	A	C	D	A	B
3	5	C	D	A	B	B	A	D	C	D	C	B	A
	6	C	D	A	B	B	A	D	C	D	C	B	A
4	7	D	C	B	A	C	D	A	B	B	A	D	C
	8	D	C	B	A	C	D	A	B	B	A	D	C
5	9	E	E	E	E	E	E	E	E	E	E	E	E
	10	E	E	E	E	E	E	E	E	E	E	E	E
	11	E	E	E	E	E	E	E	E	E	E	E	E

		Subject											
Period	Week	13	14	15	16	17	18	19	20	21	22	23	24
1	1	B	C	D	A	C	D	A	B	D	A	B	C
	2	B	C	D	A	C	D	A	B	D	A	B	C
2	3	A	A	A	B	B	B	C	C	C	D	D	D
	4	A	A	A	B	B	B	C	C	C	D	D	D
3	5	A	A	A	B	B	B	C	C	C	D	D	D
	6	A	A	A	B	B	B	C	C	C	D	D	D
4	7	A	A	A	B	B	B	C	C	C	D	D	D
	8	A	A	A	B	B	B	C	C	C	D	D	D
5	9	A	A	A	B	B	B	C	C	C	D	D	D
	10	A	A	A	B	B	B	C	C	C	D	D	D
	11	A	A	A	B	B	B	C	C	C	D	D	D

Figure 2. Experiment design to study the effects of five diets (A,B,C,D,E) on many measurements of body response on 24 healthy young males.

It was believed that at most three periods would be completed. However, after starting the experiment, daily measurements were taken, and it was found that diet effects were stabilizing in 10 to 14 days, indicating a period length of two weeks. For many of the characteristics, weekly measurements were obtained, hence, the plan in

Figure 2. For the character digestibility, it was found that there were no carry-over effects using responses at the end of a two-week period but there was a continuing effect for diet C. In some types of RM experiments, it may be necessary to determine period length after the experiment is initiated.

7. Admissibility of Treatment Sequences

In RM experiments, an item of prime importance is whether to use a CRRM or a CORM design. If a CORM is desired, then the experimenter must consider sequences of treatments and whether or not some sequences of treatments are undesirable or inadmissible. This should be determined for each RM experiment being conducted. Three examples are used to illustrate some of the considerations necessary in determining inadmissible treatment sequences and designs to take this into account.

For the first example, consider the asthma experiment discussed above. The plan for the experiment is given in Figure 3. It was required that there be two periods, pre-test periods, in which only a placebo be the treatment. The doctor conducting the experiment knew a placebo was being used, and he would not induce an asthma attack which would progress very far.

Patient Number	Treatment for Visit Number and Mediator*						Sequence of Treatments
	1-H	2-H	3-H	4-M	5-H	6-M	
801	P	P	S	I	P	C	s_1
802	P	P	S	I	C	P	s_2
803	P	P	I	S	P	C	s_3
804	P	P	P	S	I	C	s_4
805	P	P	I	P	S	C	s_5
806	P	P	I	S	C	P	s_6
807	P	P	S	P	I	C	s_7
808	P	P	P	I	S	C	s_8
809	P	P	I	S	P	C	s_3
810	P	P	P	S	I	C	s_4
811	P	P	I	S	C	P	s_6
812	P	P	S	I	C	P	s_2
813	P	P	S	I	P	C	s_1
814	P	P	P	I	S	C	s_8
815	P	P	S	P	I	C	s_7
816	P	P	I	P	S	C	s_5
817	P	P	P	S	I	C	s_4

*Treatments are P=placebo, S=Scholl's 10000-BR, I=Isoproterenol, and C=S+I in combination. The mediators are histamine(H) and methacholine(M). There were seven different treatment sequences s_1 to s_7.

Figure 3. Experimental plan of drug administration to 17 patients on 6 successive visits.

Both mediators, histamine and methacholine, needed to be tested on a patient to determine whether or not adverse reactions would occur. If they did, a patient was considered unsuitable for the study. For the 17 patients in the study, it was considered medically inadmissible to attempt a combination of drugs S + I = C before trying each drug separately on a subject. Thus, C could not occur in the sequence until both S and I had been administered to a subject. If an adverse reaction to either would have occurred, C would not be used. Thus C cannot be used before periods or visits 5 and 6. With the randomization scheme used, C occurred only four times in period 5 and 13 times in period 6. A better scheme would have been to randomly assign eight Cs to period 5 and the remaining nine to period 6 and then randomize from there. This would have resulted in less confounding of the C effect with periods 5 and 6 effects. It would also be desirable to balance the occurrence of P,S, and I over periods 3,4, and 5 as much as possible. However, in this particular experiment there was a negligible effect of visit so that the above suggestions would not need to have been considered.

In an experiment involving diet and two aerobic dancing-exercise treatments, three treatments were involved. A diet alone, D and two exercise programs in combination with diet. The three treatments are denoted as D, DA and DB. The admissible sequences for three periods were considered to be:

Period	1	2	3	4	5	6
1	DA	DA	D	DB	DB	D
2	DB	D	DA	DA	D	DB
3	D	DB	DB	D	DA	DA

An aerobics class was used for each sequence. Since the class were volunteers, it was considered inadmissible to have two periods without any exercise. If such occurred it was believed that the volunteers would drop out of the study.

For the third example, there appears to be some evidence that the production of milk by dairy cows cannot be increased very much during the latter part of a lactation period of a cow. It can be decreased easily but not increased very much, especially if the production has been lowered beyond a certain point. This means that the sequence of treatments may be very important in RM experiments involving diets for dairy cows. Results on milk production in RM experiments should be interpreted with caution.

8. Basis for Optimality

Statisticians are prone to consider only one type of an optimality criterion, i.e. variance optimality. Other criteria, especially from the subject matter fields, are most frequently ignored by the statistician. To illustrate, consider the RM experiment of Figure 2. The reader may have wondered why one design was used for subjects 1 to 12 and another one for subjects 13 to 24. The reason for this was to satisfy two criteria. The proposed design using weeks 2,4, 6, and 8 for subjects 1 to 12 is variance optimal while the design for subjects 13 to 24

and weeks 2,4,6, and 8 is variance minimal for a pairwise balanced connected design. The nutritionist had to satisfy a nutritional criterion and this was done for subjects 13 to 24. Present knowledge at the time of planning the experiment indicated that somewhere between 20 and 60 days would be required for a diet to stabilize responses. Note that in the design for subjects 13 to 24, a treatment was given continuously for nine weeks. Other nutritionists would have to believe the results from subjects 13 to 24. If the results for treatments from subjects 1 to 12 and from subjects 13 to 24 were the same, they would be believeable to other nutritionists. Hence, the statistician was happy because a variance optimal design was used for subjects 1 and 12 and the nutritionist was happy because a nutritionally optimal design was used for subjects 13 to 24. As the above example indicates, several optimality criteria may need to be satisfied. The experiment design should be constructed in such a manner as to take this into account. To use only a variance optimality criterion is too narrow and limited for research investigations.

9. Multistage RM Designs

Little has been written on the design and statistical analysis of experiments conducted simultaneously on the same experimental units (see e.g. Federer, 1984a). Several papers have appeared on the statistical design and analysis using a succession of experiments on the same experimental units as, e.g., in crop rotation experiments. The problem of carrying out simultaneous experiments occurred in designing experiments to be conducted in grocery supermarkets where the experiments were to be run simultaneously and on different commodities, e.g. apples, carrots, and paper products. Each experiment was a repeated measures one. A CORM design was used for each commodity. The sets of treatments in each CORM design should be orthogonal or at least pairwise balanced with respect to each other. They definitely should not be completely confounded. The design suggested for v = 3 treatments on each of two commodities is:

Store (Commodity one)

Period	1	2	3	4	5	6
1	A	B	C	A	B	C
2	B	C	A	C	A	B
3	C	A	B	B	C	A

Store (Commodity two)

Period	1	2	3	4	5	6
1	α	β	γ	α	β	γ
2	γ	α	β	β	γ	α
3	β	γ	α	γ	α	β

The Greek letter symbols are orthogonal to the Latin letters.

If there were interactions between the two sets of treatment, the experiment design must take this into account in order to obtain estimates of interaction effects. Also, it should be noted that a RM plan for even numbers can use only one Latin square. For example the following Latin square of order 4 is balanced for one period carryover effects:

Column

Row	1	2	3	4
1	A	B	C	D
2	D	A	B	C
3	B	C	D	A
4	C	D	A	B

One cannot find another Latin square which is orthogonal to this one. Hence, it would not be a candidate for a multistage experiment. The following square would be a candidate:

Column

Row	1	2	3	4
1	A	B	C	D
2	B	A	D	C
3	C	D	A	B
4	D	C	B	A

10. Catalogued or Specific RM Designs

As should be evident from the preceding the sole use of catalogued or computer-constructed designs would not appropriate in general. A design has to be constructed to meet the needs and requirements of the experiment. The axiom (see Federer, 1984b) that one should "design for the experiment; do not experiment for the design" applies here as well as to all other experiments. For any design that is constructed for an experiment, it should be checked to ascertain that effects can be estimated, i.e. the design is connected. Statistical criteria of optimality should be taken into account but also subject matter optimality needs to be considered. The planning aspects of RM experiments are crucial to their success. Ignoring any one or more of the points discussed above can lead to the experiment becoming only an experience without any tangible results.

11. Literature Cited

Federer, W. T. (1984a) Statistical analyses for multistage experiment designs. *Biometrical J.* 26(5):535-553.

Federer, W. T. (1984b) Principles of statistical design with special reference to experiment and treatment design. In *Statistics: An Appraisal* (editors, H. A. David and H. T. David) The Iowa State University Press, pp. 77-104.

Kershner, R. P. and Federer, W. T. Two treatment crossover designs for estimating a variety of effects. *J. American Statistical Association* 76:612-619.