The Two-Period Crossover Design in Medical Research

John R. Woods, PhD; James G. Williams, MS; and Morton Tavel, MD

The crossover design has enjoyed popularity with many clinical researchers, but has been criticized by biostatisticians. The central problem is the inability to derive an unbiased estimate of the treatment effect when differences occur because of the different sequences in which treatments are applied. This problem can be traced to a deficiency of the logic of the crossover arrangement itself. Factors that can invalidate the findings of a crossover trial include nonuniform pharmacologic and psychologic carry-over effects, failure to return patients to their baseline state before the crossover, nonuniform changes in the patients over time, and the use of time-dependent response measures. When these problems can be anticipated, a parallel-groups design should be used instead of a crossover trial.

Annals of Internal Medicine. 1989;**110**:560-566.

From Methodist Hospital of Indiana, Inc., Indianapolis, Indiana. For current author addresses, see end of text.

When a researcher wishes to test the relative benefits of two medical treatments there are two methods of comparison. The parallel-groups (between-subjects) design randomly assigns different persons to either of two treatments. The treatment effect is calculated by comparing the two group means. The within-subjects design observes the same persons under both treatments. The treatment effect is calculated by comparing subjects with themselves.

The two-period crossover trial is one of the most popular versions of the within-subjects design. In recent years, however, the crossover design has come under critcism. The Food and Drug Administration's Biometric and Epidemiology Methodological Advisory Committee (BEMAC) has been especially critical of this design. Their report (1) on the crossover design states, "... the two-period crossover design is not the design of choice in clinical trials where unequivocal evidence of treatment effects is required." This conclusion has been supported in more recent technical articles (2, 3).

Because the controversies surrounding the crossover design are essentially statistical, discussion of these controversies is usually presented in mathematical terms. We will show the problems of crossover designs in a manner that is more meaningful to the clinical investigator by appealing to the logic of the design itself. Complete statistical treatment of crossover designs can be found elsewhere (3-8).

The Basic Design

In a two-period crossover design, patients are treated for two periods, using a different treatment in each period. To avoid confusing the effects of treatments with systematic trends that may occur through time, the treatments are counterbalanced. Counterbalancing is the process in which patients are randomly allocated to the two possible treatment orders, or sequences. If the patient's condition changes with the passage of time independently of the treatments received, counterbalancing will help neutralize this effect when all patients are analyzed collectively. The basic crossover arrangement is shown in Figure 1. Patients assigned to group 1 receive treatment A in period 1 and treatment B in period 2. Patients in group 2 receive the treatments in reverse order.

Figure 2 shows the data layout for the crossover design. For simplicity, the elements in the cells can be thought of as the response measured in two patients. By showing the design with only two patients, it is possible to develop an intuitive appreciation of the effects that are calculated in a crossover design without concern for estimates of statistical variance.

The Treatment Effect

As shown in Figure 2, treatment A and treatment B are observed at both time periods (A1 and A2; B1 and B2, respectively). The treatment effect is simply the difference between the overall responses to treatment A and treatment B, calculated by comparing (A1 + A2) with (B1 + B2). Suppose the response measures are self-reported pain relief, then (A1 + A2) might be the total relief seen after treatment with an investigational drug regardless of whether it was applied first or second. Similarly (B1 + B2) might be the total relief seen with morphine, again ignoring the order in which the morphine was applied.

The Period (Time) Effect

Because each patient is observed twice, it is important to determine whether a change has occurred between the first and second period of observation. The period (time) effect is the difference between the responses seen in period 1 and the responses seen in period 2, and is calculated by comparing (A1 + B2) with (B1

+ A2). Referring again to the pain relief example, the period effect would be the total relief in the first period, contrasted with the total relief in the second period, regardless of the specific drug received in either period. Anything that might cause a patient's condition to deteriorate or improve with the passage of time can produce a period effect.

The Sequence Effect

A sequence effect occurs whenever the order in which treatments are given produces a difference that cannot be explained by the specific action of the individual treatments. This effect, because of the order of treatments, is the difference between the column totals, Sequence 1 compared with Sequence 2, and is calculated by comparing (A1 + B1) with (B2 + A2). In the pain relief example, the sequence effect would be the total relief experienced when the investigational drug is applied first followed by morphine, contrasted with the total relief experienced when morphine is applied first followed by the investigational drug. When the ordering of treatments causes a difference in the overall responses of patients, it is difficult or even impossible to separate the specific contributions of each treatment. For this reason it is wise to avoid using a crossover design if it is likely that a sequence effect may occur.

Advantages of the Crossover Design

The principal advantage of the two-period crossover design is efficiency. From a practical point of view, an investigator can obtain an equal number of response measures using only half the patients needed in a parallel-groups design. Because patient recruitment is often a problem in clinical research, the advantage of using consenting patients for more than one measure is considerable. The crossover trial might permit adequate enrollment of patients at a single center rather than requiring a multicenter trial.

The crossover design can also provide statistical efficiency. Different patients may respond with wide variation to treatments, whereas variation within the same patient may be considerably less. When this is the case, the within-subjects comparison provided by the crossover design produces a more precise estimate of the treatment difference. In general, the statistical precision of the crossover design increases in proportion to the magnitude of the statistical correlation between period 1 and period 2 observations (2).

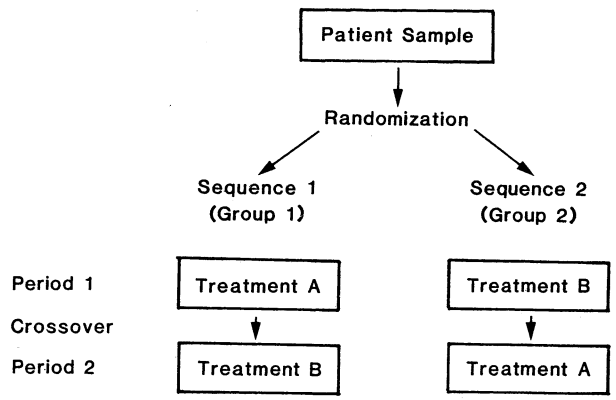

Figure 1. The basic two-period crossover design. Patients assigned to group 1 receive treatment A in period 1 and treatment B in period 2. Patients in group 2 receive the treatments in reverse order.

	Sequence 1 (Group 1)	Sequence 2 (Group 2)	Totals
Period 1	A1	B2	A1+B2
Period 2	B1	A2	B1+A2
Totals	A1+B1	B2+A2	

Treatment Effect = (A1 + A2) vs (B1 + B2)
Period (Time) Effect = (A1 + B2) vs (B1 + A2)
Sequence (Order) Effect = (A1 + B1) vs (B2 + A2)

Figure 2. Crossover effects estimates. Calculation of treatment, period, and sequence effects in a crossover design.

Deficiencies of the Crossover Design

The advantages of the crossover design are well known. Its deficiencies are more subtle. Paradoxically, the greater efficiency of this design may sometimes be a liability. Because relatively few patients are required, patient attrition (due to protocol violations, mortality, and so on) and extreme observations (outliers) can cause considerable distortion of results.

Because each patient in a crossover design must be followed through two periods using two different treatments, the total trial time will be longer. In some circumstances the longer time requirement may offset the advantage of smaller patient samples. Crossover designs will also be impractical for situations where a lengthy washout interval between treatments is necessary, and where there is a risk that lingering effects from the first treatment may still be present when the second treatment is applied. Crossover designs also make it difficult to determine which of two treatments is responsible if an adverse experience is seen.

A study of the structure of the crossover design provides a better understanding of its deficiencies. Because of the design's structure, there is a lack of independence among the various estimated statistical effects. Of particular concern is the crossover trial that yields a statistically significant sequence effect. When an investigator conducts a crossover trial a tacit assumption is made that the sequence of treatments will not produce a significant effect.

The Problem of Interactions

The problem that sequence of treatments introduces into a crossover design can be viewed as a problem of statistical interaction. Two variables are said to interact whenever the effect of one variable changes depending on the particular state of a second variable. Of greatest concern in the crossover design is the interaction of treatment and period. This interaction occurs when the effectiveness of one treatment changes, relative to the other, in the move from period 1 to period 2. This type of interaction cannot be distinguished

a) Treatment effect only:

	Sequence 1 (Group 1)	Sequence 2 (Group2)	Totals
Period 1	A1=10	B2 = 6	16
Period 2	B1 = 6	A2 = 10	16
Totals	16	16	

Treatment Difference = A − B = (10+10) − (6+6) = 8

b) Period (time) effect only

	Sequence 1 (Group 1)	Sequence 2 (Group 2)	Totals
Period 1	A1 = 8	B2 = 8	16
Period 2	B1 =10	A2 =10	20
Totals	18	18	

Treatment Difference A − B = (8+10) − (10+8) = 0

c) Treatment and Period effect

	Sequence 1 (Group 1)	Sequence 2 (Group 2)	Totals
Period 1	A1 = 8	B2 = 4	12
Period 2	B1 = 8	A2 = 12	20
Totals	16	16	

Treatment Difference = A − B = (8+12) − (8+4) = 8

d) Period by Treatment Interaction:

	Sequence 1 (Group 1)	Sequence 2 (Group 2)	Totals
Period 1	A1 = 8	B2 = 8	16
Period 2	B1 = 4	A2 = 12	16
Totals	12	20	

Treatment Difference A − B = (8+12) − (4 + 8) = 8

Figure 3. Four possible outcomes in a crossover design.

from a sequence effect in the crossover design. To illustrate this point, Figure 3 shows four possible outcomes of a crossover design.

Figure 3a shows the pattern occurring when there is a treatment effect and nothing more. The total effect of treatment A is better by 8 response units than the effect of Treatment B. A comparison of the two row marginals (period 1 compared with period 2) shows there is no period effect, and a comparison of the column marginals (sequence 1 compared with sequence 2) shows there is no sequence effect. A treatment effect without a period or sequence effect is the type of outcome an investigator would hope to obtain from a crossover design.

Figure 3b shows the pattern occurring if there is a general trend over time (period effect), but no treatment effect. Comparison of the row marginals shows that the total effect in period 2 for both treatments combined is 4 response units better than the total in period 1. The total effect of treatment A, however, is no different than the total for treatment B. This pattern of results might arise if patients were generally improving over time independently of the treatments, and the treatments themselves were both equally effective.

Figure 3c shows the pattern occurring when there is an effect of both treatment and period. The treatment effect is the same as in 3a. Treatment A is better by 8 units than treatment B. In addition, subjects fare better in period 2 regardless of the specific treatment they receive. Note that this period effect merely adds a constant increment to the effectiveness of both treatments in period 2, but the relative difference between the two treatments is left undisturbed. This outcome is more complex, but the treatment effect continues to be interpretable. This pattern might arise if patients were generally improving over time independently of the treatments, but treatment A was better than treatment B by the same amount in both periods.

Figure 3d shows a sequence effect. The treatment difference shown appears identical to those in Figures 3a and 3c, the total effect of treatment A being 8 response units better than the total for treatment B. In this case, however, the treatment difference has arisen in a peculiar manner and its meaning is ambiguous. If only period 1 observations are considered, there is no difference between the treatments. It is only in period 2 that treatment A appears more effective than B. With this type of outcome, it is difficult to evaluate the meaning of the treatment effect, because the difference between treatments is influenced by the period in which the treatments were seen. Failure to recognize this period by treatment interaction (sequence effect) might lead one to conclude, as in the first two examples, that treatment A is the therapy of choice, when in fact it may be no more effective than treatment B.

Incompleteness of Counterbalancing

The problem of sequence effects can also be understood in terms of the incomplete counterbalancing of treatments in the crossover design. When subjects are observed at more than one time, they may change in ways unrelated to the treatments tested. Such change is common in medical practice and may occur either as a result of previous exposure to treatment or because of some factor that varies with the passage of time. There is no satisfactory way to prevent such change from taking place, but its effect can be neutralized by the technique of counterbalancing, that is, by reversing the order in which treatments are given for half the patients in the study. When counterbalancing is complete, the effects of any change occurring through time will fall equally on both treatments. As a result, one treatment will not be influenced more than the other when the orders of the treatments for all subjects are considered collectively.

Contrary to common belief, the counterbalancing in a crossover design is not complete. Consequently, the control it provides is limited. The incompleteness may be seen by referring again to Figure 1. Each treatment appears equally in both periods and in both sequences. However, each treatment does not appear in every possible period-sequence pairing. For example, treatment A appears in period 2 only in sequence 2. Therefore, we can only observe how treatment A behaves in

period 2 when it is preceded by treatment B; we cannot observe how treatment A behaves in period 2 when it is not preceded by treatment B. If treatment A acts differently in period 2, it will be impossible to decide whether this difference is due to the order of testing (sequence), or to the time of testing (period). More importantly, it will be impossible to decide how much, if any, of the overall difference between the two treatments is because of the treatments themselves and how much is because of a confounding of the treatments with period or sequence effects.

This confusion can be removed by using extended designs that manipulate each variable independently of every other variable (5, 9, 10, 11, 12). Extended crossover designs continue to make use of the statistical efficiency of within-subject comparisons, but all require either additional patients, that the patients be observed for additional periods, or both. From a practical point of view the enforcement of such experimental protocols can present significant problems in clinical research. Thus, extended designs are usually no more economical or efficient than a parallel-groups design and cannot be recommended under most circumstances.

Clinical Sources of Confounding in Crossover Designs

Attention has been given to the problems that arise when a sequence effect occurs in the data of crossover designs, and the source of these problems has been traced to the logic of the design itself. There are circumstances in the research setting likely to produce such sequence effects. If an investigator is aware of these circumstances, steps can be taken to control their occurrence or to abandon the crossover design and substitute an alternative design.

Residual (Carry-Over) Effects

A residual or carry-over effect occurs when the effect of the first treatment extends beyond its period of application to influence the action of a subsequent treatment. If the carry-over is uniform, affecting both treatments equally, residual effects will appear as a period effect and will not bias the estimate of treatment differences. If the carry-over is not uniform, affecting the two treatments differently, then there will be a sequence effect, obscuring the true treatment difference. The most familiar type of residual is the drug carry-over effect that occurs when traces of the period 1 drug are present when responses to the period 2 drug are measured. Drug carry-over can be prevented by inserting a suitable washout period between applications of the two drugs.

A more subtle residual effect may arise because of certain psychologic influences established during first-period treatment or testing. In the behavioral sciences (13, 14), learning or expectancy can influence observed outcomes, but the problem is not well understood among medical researchers. Psychologic carry-over effects may occur in various ways.

For example, an investigator may use a crossover design to test the effects of a new analgesic compared with a placebo. Subjective estimates of pain are influenced by the standard against which judgments are made (15, 16). In a crossover design, period 1 estimates are made in the absence of any explicit standard of reference, but period 2 estimates are made in light of effects experienced in period 1. A placebo is likely to be judged less effective in period 2, where it can be compared with the previous effects of the active drug. Conversely, an active drug may be judged more effective in period 2, where it can be compared with the previous effects of placebo. This mixing of absolute and relative judgments can result in a sequence effect, violating the assumptions of the crossover design. The insertion of a washout period may help reduce this problem, but the required length of a psychologic washout period is usually unknown. The problem may also be remedied, in part, by blinding the subjects not only to the identities of the drugs, but also to the occurrence of the crossover point. Ultimately, however, the crossover design remains vulnerable to psychologic carry-over and alternate designs should be considered when the measurement of treatment responses relies on subjective reports. Learning effects may continue to threaten the validity of a crossover design, even when response measures are based on objective behavioral measures, rather than subjective reports (17).

In another example, an investigator wishes to test two anti-emetic drugs for control of nausea and vomiting in cancer patients undergoing chemotherapy. Some evidence suggests that nausea and vomiting in patients receiving chemotherapy are influenced by a process of classic (Pavlovian) conditioning to the offending chemical agent (18, 19). If drug A is a more effective anti-emetic agent, nausea will be controlled when it is applied in period 1 and the conditioned response will be less likely to develop. Drug B will profit from drug A's effectiveness when it is applied in period 2. Conversely, when drug B is given first, conditioning will take place in period 1, and drug A will be at a disadvantage in period 2. Again, the result will be a sequence effect and conclusions about treatment differences will be made ambiguous. This learning effect will threaten the validity of a crossover design even though an objective behavioral measure (observed frequency of vomiting) is used.

Failure to Return Subjects to Baseline State

Failure to return subjects to their baseline state can also produce a nonuniform carry-over effect, giving rise to the previously mentioned problems. In some instances, withdrawal of an apparently effective treatment may be deemed unethical. In other cases, if the first-period treatment effects a cure, it will be impossible to return subjects to their original state. In either case, if it is likely that exposure to treatments in period 1 will leave the two groups in relatively permanent but unequal states, a crossover design must be avoided.

Consider, for example, the comparison of two methods of nutritional support for increasing body weight

in patients with cystic fibrosis. If method A is more effective, then body weight may rise in the group that received method A first, and method B will profit from this when it is used in period 2. On the other hand, body weight may decline in the group that receives method B first. Consequently, method A will be at a disadvantage because persons treated in period 2 will have deteriorated physically. In this case, a sequence effect is virtually assured, because it will be impossible, on both ethical and practical grounds, to return subjects to their baseline body weight before Period 2.

Nonuniform Changes in the Patient Over Time

In many cases a patient's medical condition may change *independently* of the specific treatment offered. If the treatments are influenced differentially by this change, the results of a crossover design will be uninterpretable.

For example, an investigator wishes to test a new non-narcotic agent against morphine for relief of chronic cancer pain. If pain is relatively constant over the course of the study, then given a reasonable washout period, a crossover design may produce a fair comparison. Suppose both medications, however, are equally effective for mild pain, but morphine is superior for intense pain. Under these circumstances, if the level of pain increases as the cancer progresses, a sequence effect will result. The two drugs will appear equally effective in period 1, but morphine will be superior in period 2, and the crossover design will be unable to provide an unbiased treatment comparison.

Even if a subject's biologic state remains constant over the course of the study, systematic changes in the treatment environment may pose problems similar to those described above. For example, the movement of subjects from an intensive care area to a general medical unit, or from hospital to home care, between periods 1 and 2 can produce sequence effects if these changes have a differential effect on either the treatment responses themselves or the conditions under which those responses are observed and recorded.

Time-Dependent Response Measures

Crossover designs are inappropriate where treatment effectiveness depends on measure of elapsed time. Under these circumstances, the point at which period 2 testing begins will be governed by the outcome of the period 1 treatment, giving rise to nonindependence of the two response measures. This problem has been reported by Meier and colleagues (20).

For example, an investigator uses a crossover design to compare the duration of postoperative pain relief from two analgesics. Drug effectiveness is measured in terms of elapsed hours after injection in which the subject continues to report at least 50% pain relief. Subjects remain in period 1 until they fall below 50% pain relief, at which time the crossover to the second drug and period 2 begins.

Suppose the average duration of relief is 6 hours for drug A and only 4 hours for drug B. When drug A, the long-acting drug, is applied first, drug B will have the advantage of being applied approximately 6 hours postoperatively when pain may have begun to subside. When drug B, the shorter-acting drug, is applied first, drug A will have the disadvantage of being applied only 4 hours postoperatively when pain may still be relatively intense. Under these circumstances, the relative effectiveness of the two drugs will depend on the period in which they are observed, creating the conditions for a sequence effect and obscuring the measurement of true treatment differences.

Statistical Implications

The logical deficiencies of the crossover design have important implications for the statistical analysis. Because it is difficult to evaluate the meaning of treatment differences when the orderings of the treatments cause an effect, an investigator obtaining a statistically significant sequence effect will be unable to say with any confidence whether one treatment is actually superior to the other. Because this type of result is always possible with a crossover trial, it is important that the problem be dealt with in the statistical analysis. It should be emphasized that no statistical analysis is completely effective in untangling the potential confounding factors inherent in the crossover design. A partial solution, however, has been suggested. A method proposed by Grizzle (7) involves a separation of the analysis into two phases. The sequence effect is evaluated in phase 1 by comparing the column totals (12 and 20 in Figure 3d). If this comparison is nonsignificant (using a liberal criterion of $P = 0.10$), then the investigator may proceed to estimate the treatment effect, as already described. If the sequence effect is significant ($P < 0.10$), then the assumption of no sequence effect has been violated, and an analysis based on the crossover model will result in a biased estimate of the treatment effect. Under these circumstances, the investigator should discard the period 2 data and analyze only the period 1 treatment difference using a parallel-groups comparison (7).

Grizzle's method provides a way of determining, after the data have been collected, whether the design can yield a valid treatment comparison. It does not, however, offer a complete solution because the test of the sequence effect is based on a between-subjects comparison that has low statistical power. Brown (2) has calculated that to detect a sequence effect with sufficient statistical power (0.95), the number of subjects required in the crossover design would greatly exceed the number required in a comparable parallel-groups design. It is possible to increase the sample size, thereby increasing the statistical power of Grizzle's test, but this negates the orginal advantage of using a crossover design. The same effort put into a simple parallel-groups design would yield a treatment estimate of equal power, while avoiding the complexities of the crossover design.

There are several basic rules that should be followed in reporting the results of a crossover trial. These rules are also useful in evaluating other investigators' cross-

over results. First, the issue of subject attrition must be addressed. Because crossover designs often have few subjects, each dropout must be carefully evaluated. Unequal dropouts from different causes can invalidate the statistical analysis. Investigators should also always report period, sequence, and treatment means and do a preliminary test of the sequence effect. The outcome of this test determines whether both periods of data can be analyzed, or whether the analysis should be restricted to period 1 data alone (7).

If the full crossover design is to be analyzed, then the effects of period, treatment, and sequences must be reported. It is not appropriate to compare only the average effect of treatment and ignore the effects of period and sequence. Figure 3d shows that an average difference between treatments does not necessarily indicate the existence of a genuine treatment effect.

If an analysis of variance model is used to analyze the data of a crossover design, certain classic assumptions must be met. Specifically, it must be assumed that the observations have been sampled randomly from normal population distributions, that the variances are the same for each treatment condition, and that errors in measurement are uncorrelated across observations.

Efficient Alternatives to the Crossover Design

When the assumptions of a crossover design are in doubt, the simpler parallel-groups design is preferred. A parallel-groups design requires more subjects than a crossover design, but each subject is observed under one treatment for only one period. Therefore, the problems of protocol violation and subject withdrawal are minimized, and the statistical analysis is straightforward, involving none of the assumptions of crossover designs.

Various authors (1-3, 21, 22) have discussed the incorporation of baseline measures to increase the statistical efficiency of the standard parallel-groups design. This type of design, shown in Figure 4, involves two periods of testing but no treatment crossover. Baseline measures taken during the first period may involve a placebo treatment, no treatment, or one of the two study treatments. Like the crossover design, this design involves a within-subject comparison in the sense that each subject's response is recorded as a difference between baseline and active treatment. In most clinical situations, such difference measures result in an increase in efficiency, the gain being proportional to the statistical correlation between the baseline and treatment observations. Baseline measures can usually be incorporated with little additional time and effort. They can also provide assurance of pre-experimental group comparability independent of randomization and can be used as a basis for stratification if within-group differences are found.

When all factors are taken into account (time, complexity, subject attrition, and so on), the parallel-groups design, with the incorporation of a baseline measure, can approach the crossover design in efficiency even when the restrictive crossover assumptions

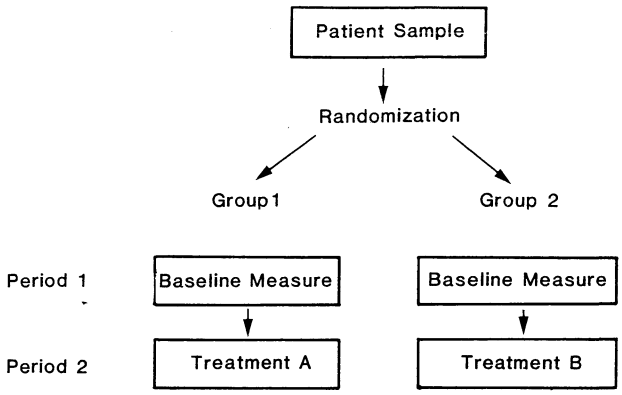

Figure 4. Parallel-groups design with baseline measures. A design involving two periods of testing but no treatment crossover. Baseline measures are taken during the first period.

concerning sequence effects are met. When the crossover assumptions are not met, however, the parallel-groups design is clearly superior to the crossover.

Conclusions

The two-period crossover design offers the advantage of efficiency, but it does so at the risk of validity. If an investigator is confident that differences between the treatments under study will not change depending on the period in which the treatments are observed, then a crossover design may be economical and efficient for generating valid treatment comparisons. It is important, however, to recognize the limitations of this design. It is often difficult to rule out the possibility of a sequence effect before conducting a clinical trial. Investigators who conduct a crossover trial in the absence of such confirmatory evidence, on the grounds that it is an efficient way to proceed, must be aware that they may produce biased results. Investigators unsure whether data collected in their particular clinical context will support the assumptions of a crossover analysis should design their study as a simple parallel-groups trial.

Requests for Reprints: John R. Woods, PhD, Department of Medical Research, Methodist Hospital of Indiana, Inc., 1701 N. Senate Boulevard, Box 1367, Indianapolis, IN 46206.

Current Author Addresses: Drs. Woods and Tavel: Department of Medical Research, Methodist Hospital of Indiana, Inc., Indianapolis, IN 46202.
Mr. Williams: Academic Computer Services, Indiana University-Purdue University at Indianapolis, Indianapolis, IN 46206.

References

1. **O'Neill R.** A report on the two-period cross-over design and its applicability in trials of clinical effectiveness. *Report of the Food and Drug Administration's Biometric and Epidemiological Methodology Advisory Committee:* Washington, D.C.: Food and Drug Administration; 1977.
2. **Brown BW Jr.** The crossover experiment for clinical trials. *Biometrics.* 1980;**36:**69-79.
3. **Hills M, Armitage P.** The two-period cross-over clinical trial. *Br J Clin Pharmacol.* 1979;**8:**7-20.
4. **Fleiss JL.** *The Design and Analysis of Clinical Experiments.* New York: John Wiley & Sons; 1986:263-90.
5. **Cochran WG, Cox GM.** *Experimental Designs.* 2nd ed. New York: John Wiley & Sons; 1957:127-43.
6. **Taulbee JD.** A note on the use of nonparametric methods in the statistical analysis of the two-period changeover design. *Biometrics.* 1982;**38:**1053-5.

7. **Grizzle JE.** The two-period change-over design and its use in clinical trials. *Biometrics.* 1965;**21:**467-80.
8. **Winer BJ.** *Statistical Principles in Experimental Design.* New York: McGraw-Hill; 1971:685-751.
9. **Lucas HL.** Switch-back trials for more than two treatments. *J Dairy Sci.* 1956;**39:**146-54.
10. **Patterson HD.** Change-over trials. *J R Stat Soc Series B.* 1951;**13:**256-71.
11. **Williams EJ.** Experimental designs balanced for the estimation of the residual effects of treatments. *Aust J Sci Res Series A.* 1949;**2:**149.
12. **Kershner RP, Federer WT.** Two treatment crossover designs for estimating a variety of effects. *J Am Stat Assoc.* 1981;**76:**612-9.
13. **Greenwald AG.** Within-subjects design: to use or not to use? *Psychol Bull.* 1976;**83:**314-20.
14. **Grice GR, Hunter JJ.** Stimulus intensity effects depend upon the type of experimental design. *Psychol Rev.* 1964;**71:**247-56.
15. **Corso JF.** *The Experimental Psychology of Sensory Behavior.* New York: Holt, Rinehart & Winston; 1967:508-49.
16. **Helson H.** Current trends and issues in adaptation-level theory. *Am Psychol.* 1964;**19:**26-8.
17. **Olver IN, Simon RM, Aisner J.** Antiemetic studies: a methodological discussion. *Cancer Treat Rep.* 1986;**70:**555-63.
18. **Bernstein IL.** Learned taste aversions in children receiving chemotherapy. *Science.* 1978;**200:**1302-3.
19. **Neese RM, Carli T, Curtis GC, Kleinman PD.** Pretreatment nausea in cancer chemotherapy: a conditioned response? *Psychosom Med.* 1980;**42:**33-6.
20. **Meier P, Free SM, Jackson GL.** Reconsideration of methodology in studies of pain relief. *Biometrics.* 1958;**14:**330-42.
21. **Chassan JB.** A note on relative efficiency in clinical trials. *J Clin Pharmacol.* 1970;**10:**359-60.
22. **William AR, Pater JL.** Using baseline measurements in the two-period crossover clinical trial. *Controlled Clin Trials.* 1986;**7:**282-9.