Paper 221

CROSS CROSSOVER STUDIES OFF YOUR LIST

Pippa M. Simpson, University of Arkansas for Medical Sciences, Little Rock, Arkansas
Robert M. Hamer, UMDNJ Robert Wood Johnson Medical School, Piscataway, NJ
Shelly Lensing, University of Arkansas for Medical Sciences, Little Rock, Arkansas

INTRODUCTION

When studying the effect of different drugs or different dose levels (or more generally, simply different treatments) on people, the variability between responses of different people can overwhelm any treatment differences. That is why crossover designs, where each person gets one of each drug or each dose level, are so appealing. Since the order in which people receive a drug may have an effect, the crossover is designed to balance out any period effects. When a carryover effect occurs, that is, when the drug's effect carries over to the next period, there may be difficulties. Nonetheless, the crossover is very appealing theoretically, with power to detect very small differences with relatively much smaller sample sizes. That is the *good* news.

Unfortunately, the *best bad news* is that the analysis for a crossover can be difficult to program in the SAS® System. The *worst bad news* is that the data from a crossover may be nearly useless. Often, even with only a two-armed study (two dose levels or two drugs), dropouts will occur. The period effect is no longer balanced, and only the first period can be analyzed validly. The power of the study is thus gone, and one is left with an independent groups design, with small power resulting from small sample sizes.

Many crossover designs confound carryover and period effects. When there is a physiologic or psychologic carryover (and the latter is more common than realized), it may be impossible to detect the treatment effects for which the study was designed.

We discuss the analysis of crossover designs, procedures in SAS/STAT® for these analyses, the difficulties of doing a proper crossover study, and suggest that perhaps we should cross crossover designs off our list of possible clinical designs.

BACKGROUND

A crossover design is a design where a patient receives two or more treatments in a random order in different periods. It is not necessary for all permutations of all the treatments to be used. Typically Latin square designs (John, 1989), which balance the ordering of treatments, are used as they are both efficient and, with balanced numbers of patients, facilitate analysis.

A crossover design allows the study of differences between treatments and subsequences of treatments. By using a crossover design, blinding can be preserved and possible *period effects* can be considered. Period effects may arise where patients may do better in a subsequent period because their state has changed, for example, their mental or health status has changed, independent of treatment. For example, if during the first period the disease remits, regardless of treatment, so that the individual is disease free by the time the second period occurs, that is a period effect. A carryover effect is one in which the first treatment changes the status of the subject and that change persists into the second period. For example, if the drug given during the first period causes the disease to remit, so that the person no longer has the disease by the time the second period arrives, that is a carryover effect, and is problematic to handle. Crossovers are typically only considered for chronic conditions which are unlikely to change over the period of study.

Since all patients receive more than one treatment, within-patient variation affects treatment variation. Since within-patient variation is typically less than between-patient variation, smaller sample sizes are needed than for a parallel arm trial, and a trial can be more efficiently completed, *if a crossover design is applicable*. Unfortunately, many people are attracted to the small sample sizes needed for the crossover design, but that efficiency is often a mirage.

PROBLEMS WITH CROSSOVER DESIGNS

Carryover of the medication itself or of its effects, physiologic or psychological, is one of the biggest problems with a crossover design. If this problem exists and a crossover design is chosen, then a design must be used so that this effect will not be confounded with the period and treatment effects. These designs typically are less efficient, more time consuming, and more complex than others (Laska et al., 1983; Lasserre, 1991). Models that take into account the possibility of carryover assume that at most there is carryover effect from the previous period. If there are more than 2 periods, theoretically the carryover could change treatment effects in every subsequent period complicating any analysis

considerably! Moreover, it is typically assumed that this carryover from one period to the next is determined by the action of the effect of the drug in the first period on the effect of the drug in the second. If, indeed there is a period effect, this carryover effect may change in subsequent periods (an interaction). Thus, allowing for a simple carryover may not be allowing for enough. If this is a drug study, safety issues may need to be addressed. Depending on the design, carryover effects of a treatment into a subsequent period may be considered in the analysis, but, in general, it renders the use of a crossover dubious.

In order to avoid carryover a suitable 'washout' period, where neither of the treatments are given between the two treatment periods, is recommended, but may not be feasible. The washout period length will be chosen based on the knowledge about the treatments' effects, if this is available. If it is not feasible to have an inactive washout then the study has to be designed so that measurement is only taken when there is no longer a carryover effect.

A strong case can be put that a decision *a priori* has to be made about the existence of a carryover effect. Senn (Senn, 1993b) and others state that if it is believed *a priori* that carryover does not exist then the model should disregard the possibility of its existence. The Grizzle two stage (Grizzle, 1965; Grieve, 1982), and the Kenward and Jones three-stage procedures (Kenward et al., 1987), for studies with baseline, used for assessing carryover in the two period, two treatment crossover described below (AB/BA study) seem to have been discredited (Senn, 1991; Senn, 1994).

In any study, when data are not missing completely at random, analysis is problematic. However, missing data in a crossover is particularly a problem since the within patient comparisons are a main aspect of the study. In addition, since the design is (theoretically) more efficient, the sample size will be smaller, and the missing data will have more impact (Woods et al., 1989). Should patients not complete all treatments, their data may not give a valid comparison of the treatments . The data are difficult to incorporate but omitting them will not be an intention to treat analysis and may bias the results. Data may not be missing at random since the particular sequence that the patient experienced may be related to the probability of dropping out. Inclusion of the patient in a model, (Little et al., 1987; Little, 1993) which allows unequal numbers of observations for each patient may thus be inappropriate. All other things being equal, the more periods the design has, the higher the probability will be for dropout. Designs which purport to allow for carryover, typically have more periods (Richardson et al., 1996). An example is the three period two-treatment design which is used to allow for carryover. It would replace the two treatment two period design discussed below.

Some effort has been made to reduce the number of periods in crossovers. Therefore, there have been crossover designs which may have fewer periods than treatments. An alternative crossover design for a seven-treatment study might be a cyclic design with four periods (John, 1989; Matthews, 1994). This could be relatively efficient and will have the advantage of potentially fewer dropouts. However, these incomplete designs can not be as efficient as some designs with the same number of periods as treatments.

With a crossover there may be a treatment effect, a period effect, a subject effect and even a carryover effect. All too often the data are analyzed without reference to the design and consideration of all possible effects and a biased estimate of treatment effect is obtained.

We will first state a general model for crossover designs with continuous data. We will illustrate this model by discussing the two-period two-treatment crossover design denoted by the AB/BA design. We show how analyze it using the SAS system. We will discuss briefly methods for nonnormal data, including nonparametric methods, and methods for binary and ordinal data. Then we use some general examples to illustrate how crossovers might not be as useful as they might first appear.

GENERAL MODEL FOR CROSSOVER WITH CONTINUOUS DATA

General Description: Suppose that we have t treatments and p periods with n_i patients allocated randomly to sequence group i,
i =1, 2, …g. If the number of patients is the same for each sequence, ie $n_i = n$, for all i, then the design is more efficient. However, this is not required.

Let $y_{i(j)k}$ be the observed value of the random variable $Y_{i(j)k}$, such that

$$Y_{i(j)k} = \alpha_{ik} + s_{i(j)} + \varepsilon_{i(j)k} ,$$

where $s_{i(j)}$ is a an effect due to subject j of sequence i, j = 1, 2,… n_i and α_{ik} is an effect indexed by sequence i and period k, k=1,2,..p. The $\varepsilon_{i(j)k}$ are a random "error" term with expectation 0 and variance γ^2. Usually it is assumed that the $\varepsilon_{i(j)k}$ are independent of each other and the $s_{i(j)}$ are randomly and identically distributed for every i,j,k with mean 0 and variance σ^2. Normality for the $\varepsilon_{i(j)k}$ is a common distributional assumption.

It can be seen that

$$\alpha_{ik} = E[Y_{i(j)k}] - E[s_{i(j)}].$$

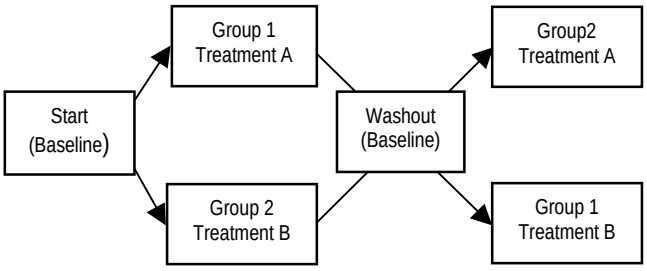
Interest centers around α_{ik} which can be expressed in terms of treatment, period and possibly carryover effects of one treatment from a previous period on a treatment in a subsequent period.

$$\alpha_{ik} = \mu + \tau_{d(i,k)} + \pi_k + \lambda_{d(i,k-1)}$$

Here $\tau_{d(i,k)}$ is the effect of the treatment in period k from sequence i. π_k is the effect of period k, $+\lambda_{d(i,k-1)}$ is a carryover effect arising from treatment in period (k-1) from sequence i, which will change the effect of treatment in period k from sequence i.

THE AB/BA CROSSOVER, THE MOST COMMON DESIGN

The simplest crossover is the AB/BA design which is below.



The study may or may not have a baseline and in some cases there will be no washout period. When the study has no baseline and no washout period, the crossover is in its simplest form. One simply has, for each subject, two measurements, one taken under each treatment, and an additional variable to indicate whether the treatments were administered in the AB or BA order.

NORMALLY DISTRIBUTED ANALYSIS

Treatment effects: With normality, for each sequence the average of the difference of the second period from the first is calculated and the difference of these two averages is calculated. This is a good unbiased estimate of the treatment effect. The period effect drops out using these differences. An (unpaired) t-test (PROC TTEST) can be used to assess the difference.

Period effects: With normality, for each sequence the average of the difference of the two periods is calculated and the sum of these two averages calculated. This is a good unbiased estimate of the period effect difference. An (unpaired) t-test (PROC TTEST) can be used to assess the effect by multiplying the differences for one sequence by –1, so that the two average differences are essentially summed.

In the SAS system, an approach which will generalize for any crossover with normally distributed data is to use the GLM procedure. Assume that the data are stored in a SAS data set named DAT1, and consist of an observation for each patient at each period (PERIOD) with the outcome (OUTCOME), the treatment (TREAT), patient number (PATIENT) and the sequence group (GROUP).

```
proc glm data=dat1;
   class group patient period treat;
   model outcome = group patient(group) period
                   treat;
   random patient(group)/test;
   * tests carryover effect;

   estimate 'Treatment Difference' treat –1 1;
   * treatment difference for 2-treatment design;
run;
```

Where there are baseline measures (BASE) for each period, the MIXED procedure should be used, fitting the above model with BASE added to the model statement. Other ways are discussed in Senn (Senn, 1994)

NONNORMAL DATA

Consider further, that the dependent variable might be ordinal or categorical. If there is a period effect then the crossover analysis becomes more difficult.

NONPARAMETRIC APPROACHES

Treatment effects: Without normality, a Wilcoxon rank sum test can be used to compare the differences of the two periods for both sequences pooled over periods; in the SAS system, the NPAR1WAY procedure can be used. With carryover, the difference above contains the carryover effects and cannot be separated from them. Therefore this design should not be used if there is a carryover effect.

Period effects: Without normality, a Wilcoxon rank sum test can be used to compare the difference of the first period and -1 times the difference of the second period. In the SAS System, PROC NPAR1WAY can be used. With this design the period effect would be confounded with any carryover effect.

A simple nonparametric analysis that assumes no carryover is given below. Assume that the data are

stored in a SAS data set named DAT2, and consist of an observation for each patient at both periods (PERIOD1 and PERIOD2) with the outcome (OUTCOME), the treatment (TREAT), patient number (PATIENT) and the sequence group (GROUP).

```
data dat2;
  set dat2;
  trt_diff=period1-period2;
  if group=1 then per_diff=trt_diff;
  else if group=2 then per_diff=-1*trt_diff;
run;

proc npar1way wilcoxon data=dat2;
  class group;
  var trt_diff per_diff;
run;
```

For data which are metric, but not normally distributed, bootstrapping or randomizatoin tests represent an alternative to normal-theory analyses. PROC MULTTEST can be used to produce bootstrap or permutation tests.

Tudor and Koch (Tudor et al., 1994) gives a description of nonparametric methods and their limitations for models with baseline measurements and carryover. Essentially they are an extension of the Mann-Whitney, Wilcoxon or Quade statistic. Without a design for carryover, there are more analyses available. More recently Tsai and Patel have proposed robust regression techniques using M-estimators for considering the baseline. The SAS system does not have this option available, but they can be obtained by programming an iteratively reweighted regression routine. For more than two treatments, Senn (Senn, 1993b) suggests several pairwise comparisons as discussed above for the AB/BA design. As in most designs, nonparametric analysis is more limited than a parametric analysis, and this is very apparent once a design with more than two treatments is considered.

ORDINAL OR DICHOTOMOUS DATA

Over the years there has been extensive work on ways to deal with longitudinal data of this type. These methods can be applied in the analysis of crossover designs. Taking a marginal approach which involves only the counts in the table indexed by each pair of time points (treatment in which period) Koch and others (Landis et al., 1977) using a weighted least squares approach and later (Liang et al., 1986; Landis et al., 1977) developed a GEE approach. Koch's approach can be implemented using the CATMOD procedure in the SAS system (Koch et al., 1988). However, large sample sizes are needed, and it is precisely large sample sizes which are not usually found in crossover designs. For Zeger et al' s (Diggle et al., 1995),(Zeger et al., 1986),(Zeger et al., 1988) approach, the GENMOD procedure can be used. It has been shown that ignoring the variance and covariance structure does not necessarily have to be considered and a fixed structure can be chosen, to simplify the process. For subject specific approaches, logistic regression conditional on a subject's preferences using the PHREG procedure is also applicable. However, if the subject effect is modeled as random, PROC GENMOD is again a possibility. All marginal models can deal with missing data. With the different assumptions of the models different estimates of the treatment effect may be obtained. There are different problems with each approach. The conditional approach is restricted to a generalized logit link function and loses information about between patient behavior. The marginal approaches may not be efficient and the estimates are calculated with a lack of accuracy. A more comprehensive discussion of approaches can be found in Kenward and Jones (Kenward et al., 1994). Senn (Senn, 1993b; Senn, 1993a) tends to advocate taking, at least at first, a fairly simple approach of comparing pairs of treatments and tailored to the type of results. An example of an analysis of a two-period three-treatment crossover with dichotomous outcome using LOGISTIC in the SAS system can be found in Stokes et al. (Stokes et al., 1995)

For binary outcomes, PROC MULTTEST has bootstrap and permutation tests which may be useful.

OTHER SAS APPLICATIONS

A SAS macro to generate crossover designs and to assess the impact of imbalance on the power of tests can be obtained at: http://www.sas.com/techsup/download/observations/2q96/broberg/testpow.sas.

EXAMPLES

To illustrate the limitations of the crossover design an example will be discussed. Various examples from the literature will then be used to illustrate some points.

Suppose we wish to study three different levels of a medication and placebo in the treatment of Attention Deficit Hyperactivity Disorder (ADHD) in children. We will measure various outcomes, some of which are parent assessment of behavior and reaction to several tests. Because of the wide variability of patients with this disorder, it is attractive to use a design where the patients are their own controls. The medication used is known to have a half-life of six hours, and we assume that there will be no

carryover of the medication. So, according to one researcher a crossover is applicable.

A crossover for this study would have four treatments. The classic way to design this would be to have a four period crossover, using a Latin square repeatedly for patient assignment. An optimal balanced uniform design for treatments A, B, C, and D like this is presented below:

		Treatment Period			
		1	**2**	**3**	**4**
Sequence	**1**	A	D	B	C
	2	B	A	C	D
	3	C	B	D	A
	4	D	C	A	B

Note: This Williams design (John, 1989) is such that every treatment occurs equally often in each period and each patient receives each treatment equally often. Moreover, each treatment follows directly each other treatment equally often. This is a very efficient design for looking at treatment effects since the period effect can be separated from the treatment effect due to the orthogonal design. Balanced designs such as this will require as many periods as treatments.

It will be difficult to obtain required efficiency if some doses cannot follow others. It might be that placebo should not follow the high dose because of potential rebound or withdrawal effects. Hence, it may be impossible to obtain a balanced design and the efficiency will be lower. In addition, if there is any possibility of dropout due to length of study for more than a two-treatment study such an “optimal” design may not be attractive. One would not wish to design a study that it is impossible to realize. Again, often in trials of sick people, retention of patients is difficult (and it is more difficult if some of the doses used in the trial are not effective.

For many crossovers, the timing or spacing of periods is loose. Patients will be included in a study in a staggered way. The gap between periods may vary. Periods may correspond to when the patient can come in. Provided this does not change the status of the patient, the periods can be, and commonly are, taken to mean the sequence numbers of the treatments. In some trials this may or may not matter. For children with ADHD this changing meaning of period may have a bearing on the study, since the time of year, school, vacation, weather and activities may affect the behavior.

It is difficult in the ADHD example to see how carryover could not be a problem. The environment of the patient and the patient’s behavior would be affected by previous behavior. Indeed it might be suspected that there might be a cumulative carryover depending on the order and the period when an effective dose for the patient was given. So, despite the attractiveness of the efficiency of the design, we would argue that this is a case, where due to carryover and other reasons, the crossover must be crossed off the list.

This may seem an exaggerated example. Nevertheless several studies have been done on ADHD patients using a crossover design (Mattes et al., 1985; Mayes et al., 1994). Moreover, with sick people even a chronic illness may change status over time and the response to a drug treatment may last longer than expected, especially due to a psychologic response of feeling better. The use of a crossover design to evaluate two treatments for an acute attack of asthma in children might be suspect since some of the affects of asthma are compounded by the child’s reaction to the symptoms (Amirav et al., 1997). An extreme case where it would seem that carryover form a treatment can occur is in studies of two treatments for infertility (Cohlen et al., 1998; te Velde et al., 1998)! Yet there has been discussion that it should be used. This may be due to a misunderstanding where the advocates are really arguing in favor of N of 1 trials where a patient has many sequences of treatments in order to determine the best treatment for that patient (Mahon et al., 1996; Menard et al., 1990).

CONCLUSION

Crossovers are useful for pharmacokinetic studies in normal volunteers where a single dose of a drug is given and the concentration of the drug in the body is studied over a relatively short time (1-3 days typically). Sometimes physiologic effects of the drug are also measured. This design may be used to show equivalent properties of two drugs (bioequivalence). Crossovers are also useful for repeated dose studies where it is important to find the level of a drug when a patient is taking medication over a period of time - the steady state pharmacology.

In such studies, properly designed, there it is unlikely to be carryover or dropout. However, when studying sick people, even with a chronic stable disease, there may be many problems in using the crossover design and a statistician may crossly find these too heavy a cross to bear**.**

ACKNOWLDEGMENTS

REFERENCES

Amirav, I., & Newhouse, M.T. (1997). Metered-dose Inhaler Accessory Devices in Acute Asthma: Efficacy and Comparison with Nebulizers: A Literature Review. *Archives of Pediatrics & Adolescent Medicine*, **151**(9), 876-882.

Cohlen, B.J., te Velda, E.R., Looman, C.W.N., Eijckemans, R., & Habbema, J.D.F. (1998). Crossover or Parallel Design in Infertility Trials? The discussion continues. *Fertility and Sterility,* **70**(1), 40-45.

Diggle, P.J., Liang, K.Y., & Zeger, S.L. (1995). Analysis of Longitudinal Data. Oxford: Clarendon Press.

Grieve, A.P. (1982). The Two-period Changeover Design in Clinical Trials. *Biometrics*, **38**, 517

Grizzle, J.E. (1965). The Two-period Changeover Design and its Use in Clinical Trials. *Biometrics*, **21**, 467-480.

John, P.W.M. (1989). Statistical Design and Analysis of Experiments. New York: The Macmillan Company.

Kenward, M.G., & Jones, B. (1987). The Analysis of Data From 2 x 2 Cross-over Trials With Baseline Measurements. *Statistics in Medicine*, 6(a), 911-926.

Kenward, M.G., & Jones, B. (1994). The Analysis of Binary and Categorical Data From Crossover Trials. *Statistical Methods in Medical Research,* **3**, 325-344.

Koch, G.G., & Edwards, S. (1988). Clinical Efficacy Trials with Categorical Data. In K. E. Peace (Ed.), Biopharmaceutical Statistics for Drug Development. (pp. 403-451). New York: Marcel Dekker.

Landis, J.R., & Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, **33**(March), 159-174.

Laska, E., Meisner, M., & Kushner, H.B. (1983). Optimal Crossover Designs in the Presence of Carryover Effects. *Biometrics*, **39**(4), 1087-1091.

Lasserre, V. (1991). Determination of Optimal Designs Using Linear Models in Crossover Trials. *Statistics in Medicine*, **10**, 909-924.

Liang, K.Y., & Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22.

Little, R.J.A. (1993). Pattern-Mixture Models For Multivariate Incomplete Data. *Journal of the American Statistical Association*, **88**(421), 125-134.

Little, R.J.A., & Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York: John Wiley & Sons.

Mahon, J., Laupacis, A., Donner, A., & Wood.T. (1996). Randomised study of n of 1 trials versus standard practice. *BMJ*, **312**, 1069-1074.

Mattes, J.A., Boswell, L., & Oliver, H. (1985). Methylphenidate Effects on Symptoms of Attention Deficit Disorder in Adults. *Archives of General Psychiatry,* **41**(11), 1059-1063.

Matthews, J.N.S. (1994). Multi-period Crossover Trials. *Statistical Methods in Medical Research*, **3**, 383-405.

Mayes, S.D., & Bixler, E.O. (1994). Reliability of Global Impressions for Assessing Methylphenidate Effects in Children With Attention-deficit Hyperactivity Disorder. *Perceptual & Motor Skills*, **77**(3), 1215-1218.

Menard, J., Bellet, M., & Serrurier, D. (1990). From The Parallel Group Design To The Crossover Design, And From The Group Approach To The Individual Approach. *Am J Hypertens*, **3,** 815-819.

Richardson, B.A., & Flack, V.F. (1996). The Analysis of Incomplete Data in the Three-Period Two-Treatment Cross-over Design for Clinical Trials. *Statistics in Medicine*, **15**, 127-143.

Senn, S. (1993a). A Random Effects Model For Ordinal Responses From A Crossover Trial. Letter to the Editor. *Statistics in Medicine*, **12**, 2147-2150.

Senn, S. (1993b). Cross-over trials in clinical research. New York: John Wiley & Sons.

Senn, S. (1994). The AB/BA Crossover: Past, Present and Future? *Statistical Methods in Medical Research*, **3**, 303-324.

Senn, S.J. (1991). Problems With the Two-stage Analysis of Cross-over Trials. *British Journal of Clinical Pharmacology*, **32**, 133

Stokes, M.E., Davis, C.S., & Koch, G.G. (1995). Categorical Data Analysis Using the SAS System. Cary, N.C.: SAS Institute, Inc.

te Velde, E.R., Cohlen, B.J., Looman, C.W.N., & Habbema, J.D.F. (1998). Crossover Designs Versus Parallel Studies in Infertility Research. *Fertility and Sterility*, **69**(2), 357-358.

Tudor, G., & Koch, G.G. (1994). Review of Nonparametric Methods for the Analysis of Crossover Studies. *Statistical Methods in Medical Research,* **3**, 345-381.

Woods, J.R., Williams, J.G., & Tavel, M. (1989). The Two-Period Crossover Design In Medical Research. *Annals of Internal Medicine,* **110**, 560-566.

Zeger, S.L., & Liang, K.Y. (1986). Longitudinal Data Analysis For Discrete And Continuous Outcomes. *Biometrics*, **42**, 121-130.

Zeger, S.L., Liang, K.Y., & Albert, P.S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, **44**, 1049-1060.

CONTACT INFORMATION

Pippa Simpson, Ph.D.
University of Arkansas for Medical Sciences
Arkansas Children's Hospital
800 Marshall St. – PEDS/CARE
Little Rock, AR 72202
Work Phone: (501) 320-6631
Fax: (501) 320-1552
Email: psimpson@care.ach.uams.edu