



Crossover Designs in Nutrition and Dietetics Research



Jeffrey E. Harris, DrPH, MPH, RD, LDN*; Hollie A. Raynor, PhD, RD, LDN*

ARTICLE INFORMATION

Article history:

Submitted 2 December 2016

Accepted 21 March 2017

Available online 4 May 2017

Keywords:

Crossover

Washout

Carryover effect

Phase effect

Sequence effect

Supplementary materials:

PowerPoint presentation available at www.jandonline.org

2212-2672/Copyright © 2017 by the Academy of Nutrition and Dietetics.

<http://dx.doi.org/10.1016/j.jand.2017.03.017>

*Both authors share equal authorship of this article.

ABSTRACT

This article is the 12th installment in a statistical series exploring the importance of research design, epidemiologic methods, and statistical analysis as applied to nutrition and dietetics research. The purpose of this series is to assist registered dietitian nutritionists in interpreting nutrition research and aid nutrition researchers in applying scientific principles to produce high-quality nutrition research. This article focuses on the use of crossover designs in nutrition and dietetics research. The purpose is to distinguish the crossover design from the randomized clinical trial, define important terms, illustrate a 2×2 crossover design, discuss potential confounding variables in the crossover design, describe the analysis and interpretation of crossover data, present sample size considerations, provide examples of the use of the crossover design in nutrition and dietetics, and discuss additional considerations when the independent variable has more than two levels.

J Acad Nutr Diet. 2017;117:1023-1030.

THE CROSSOVER AND THE RANDOMIZED CLINICAL trial are two commonly used experimental designs for examining the effects of a dietary manipulation or intervention. In both of these designs, the independent variable has at least two levels, with both levels being a unique dietary manipulation (intervention) or with one level being a dietary manipulation and the other level being a control comparison. The dependent variable is a health-related outcome. In a randomized clinical trial, participants are randomized to a unique manipulation or control group for the duration of the trial.¹⁻⁴ Identical dependent outcome variables are measured in all groups in the trial, with outcomes compared between groups, or between participants, to determine treatment effectiveness. This type of randomized clinical trial is commonly referred to as a parallel-group trial or design.

In contrast, a crossover design has all participants receive all levels of the independent variable at some point in the study, but participants do not receive all levels at the same

time.⁵⁻¹³ In the most basic illustration of a crossover design, participants are randomized to one of two groups. One group receives the dietary manipulation while the other group receives the control comparison. Once this phase is complete, the groups switch—or cross over—to the next phase, with the group having received the dietary manipulation receiving the control comparison and the other group receives the dietary manipulation. To determine treatment effectiveness, dependent outcome variables are measured for the two levels of the independent variable and are compared within the same participant.

DEFINITIONS OF KEY TERMS

Several key terms are commonly used when describing aspects of a crossover design.^{5,11,12} Sequence refers to the order of the manipulations (levels of the independent variable) presented to a group, such as AB or BA. The phase is each period during the study when a manipulation is implemented, so for sequence AB, manipulation A is administered in Phase 1 and manipulation B in Phase 2. A washout period is a time period between the end of one phase and the start of the next phase when no manipulation/independent variable is present.

SPECIFIC ILLUSTRATION OF A CROSSOVER DESIGN

The simplest crossover design, the 2×2, will be used for illustration throughout this article. This is the most common crossover design used in nutrition research. Figure 1 illustrates this type of design. In Figure 1, participants are randomized to receive the dietary manipulations in sequence AB

The Continuing Professional Education (CPE) quiz for this article is available for free to Academy members through the MyCDRGo app (available for iOS and Android devices) and via www.eatrightPRO.org. Simply log in with your Academy of Nutrition and Dietetics or Commission on Dietetic Registration username and password, go to the My Account section of My Academy Toolbar, click the "Access Quiz" link, click "Journal Article Quiz" on the next page, then click the "Additional Journal CPE quizzes" button to view a list of available quizzes. Non-members may take CPE quizzes by sending a request to journal@eatright.org. There is a fee of \$45 per quiz (includes quiz and copy of article) for non-member Journal CPE. CPE quizzes are valid for 1 year after the issue date in which the articles are published.

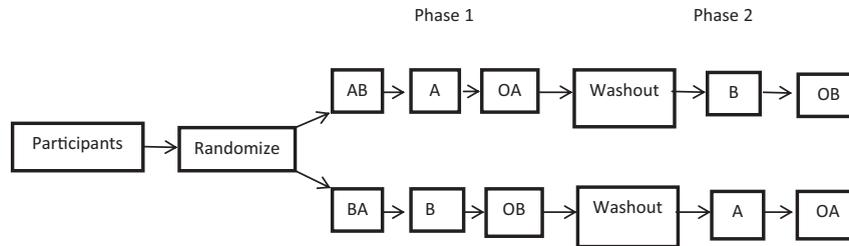


Figure 1. A 2×2 crossover design. A or B denotes one of two levels of the independent variable, AB or BA denotes sequence of manipulation of the levels of the independent variable in the crossover design, OA denotes dependent outcome variable measured after manipulation A, OB denotes dependent outcome variable measured after manipulation B, and Washout denotes period of removal of manipulation for specified time interval prior to initiating second manipulation of crossover design. (NOTE: Information from this figure is available online at www.jandonline.org as part of a PowerPoint presentation.)

or BA. Those in sequence AB receive dietary manipulation A and those in BA receive dietary manipulation B during Phase 1. At the completion of Phase 1, the dependent outcome variable is measured. These measurements are followed by a washout period in which both groups receive no dietary manipulation. After the washout period, those in sequence AB receive dietary manipulation B and those in BA receive dietary manipulation A during Phase 2. At the completion of Phase 2, the dependent outcome variable is measured again. Outcomes measured after dietary manipulation A are compared with outcomes measured after dietary manipulation B in the same participant.

To illustrate a 2×2 crossover design, suppose 50 Hispanic women aged 30 to 50 years with high blood levels of C-reactive protein (CRP), measured with the high-sensitivity CRP (hs-CRP) test, are recruited for a study. Elevated hs-CRP levels are associated with high levels of body inflammation. The investigators desire to answer the question: Does tart cherry juice reduce elevated levels of hs-CRP in Hispanic women aged 30 to 50 years with elevated hs-CRP? Participants are randomized to either treatment sequence AB or BA. In Phase 1, participants in the AB sequence receive 16 oz tart cherry juice each day for 6 weeks with participants in BA receiving 16 oz control fruit juice that has a similar taste to the tart cherry juice. The control fruit juice contains no anthocyanins, the supposed active ingredient in tart cherry juice. The comparable tastes help in blinding participants to the cherry juice vs the control. At the end of the 6-week period, serum hs-CRP levels are determined for all participants. During the following 4 weeks both groups receive no fruit juice and consume their standard diet (washout). After this period, in Phase 2 those in sequence AB receive the fruit juice and those in BA receive the tart cherry juice for 6 weeks. Following Phase 2, serum hs-CRP levels are measured again. To determine whether tart cherry juice reduced elevated levels of hs-CRP effectiveness, serum hs-CRP measured after 6 weeks of consuming tart cherry juice is compared with serum hs-CRP measured after consuming 6 weeks of fruit juice within the same participant.

Potential Confounding Variables in the Crossover Design

Carryover, phase, and sequence effects can all be confounding variables in a study using a crossover design. These confounding variables may influence the dependent outcome

variable in a different way than the independent variable, making it challenging to draw conclusions about the independent variable from the investigation. To address this issue, crossover designs should be designed to minimize these potential confounders.

Carryover Effect Considerations. A carryover effect is the potential persistence of an effect of a manipulation after it has been removed (ie, when the phase in which it was applied ends). This type of effect occurs when it is anticipated that the manipulation makes a more lasting change, either physiologically, behaviorally, or psychosocially. For example, suppose a crossover trial is used to test the influence of fruit and vegetable exposure, the independent variable, on fruit and vegetable intake, the dependent outcome variable. The independent variable has two levels: Intervention A, which is exposure via taste-tests to fruits and vegetable, and Intervention B, which is no exposure (control comparison). Group 1 is randomized to the sequence AB, whereas Group 2 is randomized to the sequence BA. Because Intervention A may have effects on fruit and vegetable intake even after the taste-tests on fruits and vegetables are complete, the effects of Intervention A may carry over to Intervention B and influence measures of fruit and vegetable intake taken after Intervention B. This carryover effect would influence the measure of fruit and vegetable intake following Intervention B only in those participants randomized to the AB sequence, confounding the interpretation of comparing Intervention A with Intervention B on fruit and vegetable intake.

A washout period between phases of a crossover design is used in an attempt to counteract carryover effects. During this period, all participants are taken off all manipulations in an effort to let the effects of the differing levels of the independent variable wear off and theoretically allow the dependent outcome variable to return to baseline levels. When designing a crossover study, it is important to determine the duration of effects of the manipulations being tested so that an appropriate amount of time is built into the washout period. However, some manipulations may actually have permanent changes on physiological, behavioral, or psychosocial outcomes. Manipulations that are designed to enhance nutrition knowledge or produce a lasting behavior change in dietary intake are examples of more permanent changes. No length of washout period will address these lasting changes; thus, these types of manipulations cause carryover effects. Manipulations that may cause permanent

changes that cannot be diminished or removed during a washout period should not be tested with a crossover design and are more appropriately tested using a randomized clinical trial design.

Phase Effect Considerations. A phase effect occurs when variables other than the independent variable that can potentially influence the dependent outcome variables vary in the differing phases. For example, a study may be examining the influence of whole grains on serum lipid levels in children. Intervention A provides families with whole grain foods for 4 weeks to assist children in consuming whole grains, whereas Intervention B provides families with foods made with refined grains for 4 weeks so that no whole grains are consumed. Families are instructed to use the products in the amounts provided at meals and snacks for the children and that the provided foods should replace the usual grain products that the children consume. This design allows a comparison between when children are consuming whole grains vs when they are not. Children are randomized to sequence AB or BA. Phase 1 occurs during the summer so children are predominantly eating at home and can consume the foods provided by the investigative team. Phase 2 occurs during the school year, and many of the children consume school breakfast and lunch. Each of the meals provided by the schools contain 1 oz whole grain. Thus during Phase 2, children receiving intervention B would still be consuming whole grains at school breakfast and lunch, which did not occur in children receiving intervention B in Phase 1. This design would influence the measures of serum lipid levels following intervention B only in those participants who received intervention B in Phase 2, confounding the interpretation of comparing intervention A to intervention B on serum lipid levels.

Sequence Effect Considerations. A sequence effect occurs when dependent outcome variables are influenced by the order in which the levels of the independent variable are implemented in a participant. Suppose tart cherry juice was administered in a crossover design to determine whether it can reduce pain that results from hard distance biking among world-class cyclists. The control comparison was given fruit punch, flavored to replicate the tart cherry juice, to assist with condition blinding. Going into the study, the cyclists know that tart cherry juice can promote gastrointestinal discomfort and diarrhea. A sequence effect could occur when those getting the tart cherry juice in the first phase experience the gastrointestinal discomfort, compromising the blinding. These participants now realize they are getting the treatment in Phase 1 and the control comparison in Phase 2. This knowledge could affect their biking effort in Phase 2. These participants might ride with less intensity and thereby have less pain during Phase 2. This may cause less of a difference in pain scores between the two phases when the sequence order is tart cherry juice followed by fruit punch compared with the sequence order of fruit punch followed by tart cherry juice.

Advantages and Disadvantages of Crossover Designs

Crossover designs have two advantages over randomized clinical trials.^{3,14-16} Because participants are exposed to all

levels of the independent variable, each participant can serve as his or her own control. Because of this comparison, the groups' characteristics of those exposed to the levels of the independent variable are exactly the same. Therefore, no confounding can occur due to subtle differences between the comparison groups as you might see in a randomized clinical trial.

The second advantage is that the sample sizes needed for each level of the independent variable can be substantially smaller than those needed for a randomized clinical trial. In addition, even in cases where there was no difference in the sample size needed for each of the levels of the independent variable, the overall sample size needed for the study is reduced, because you are not randomizing different participants into unique groups representing the levels of the independent variable. Using a crossover vs a randomized clinical trial reduces the effort one must make for participant recruitment and associated study expenses.

One disadvantage of the crossover trial is the additional burden that may be required of participants, in terms of time and effort, because they will be involved in all manipulations in the study, compared with just one manipulation in the study. This might lead to increased levels of dropout from participants, which could lead to missing data (outcomes from each manipulation may not be collected). The main disadvantage to conducting a crossover trial is the risk of a carryover, phase, or sequence effect confounding the study. In cases where a carryover, phase, or sequence effect has confounded the data, only the Phase 1 data can be used, thereby reducing the sample size by half and limiting power to detect a significant effect of the manipulation.

Finally, in cases where the population of interest in the study has serious health concerns that are being investigated in the trial, there may be ethical concerns in using a crossover design. These concerns center around moving participants to a level that provides less treatment than an earlier level, as well as removing all treatment during the washout period.¹⁷

Crossover Design Data Analysis

Table 1 contains data to be used to illustrate the analysis of crossover data. The example mentioned previously examining the effect of tart cherry juice on body inflammation as measured by serum hs-CRP is used. In the example, 10 participants are randomized to sequence AB or sequence BA. During Phase 1, those in sequence AB receive tart cherry juice and those in sequence BA receive a similar fruit juice that does not contain anthocyanins. At the end of Phase 1, serum hs-CRP is measured in both groups. Phase 1 is followed by a washout period in which participants consume their usual diet. Next, Phase 2 is initiated with those in sequence AB receiving the fruit juice and those in sequence BA receiving the tart cherry juice. At the end of Phase 2, serum hs-CRP is measured. Table 2 contains the descriptive statistics for the data in Table 1.

The item of greatest interest in the analysis of the data from this 2×2 crossover design is the treatment effect of the tart cherry juice on serum hs-CRP. A major mistake can be made in data analysis by simply combining all the treatment data values (serum hs-CRP at the end of the tart cherry juice phase) and comparing them with the combined control values (serum hs-CRP at the end of the fruit juice phase) to see whether there

Table 1. Example data for use in 2×2 crossover design examining the influence of tart cherry juice and control fruit juice on body inflammation as measured by serum high-sensitivity C-reactive protein (hs-CRP)

Participant	Sequence ^a	End of Phase 1	End of Phase 2	Difference
		hs-CRP level (mg/L) ^b	hs-CRP level (mg/L)	
1	AB	1.3	3.0	1.7
2	AB	1.5	2.8	1.3
3	AB	2.0	3.1	1.1
4	AB	1.0	2.7	1.7
5	AB	1.1	2.5	1.4
6	BA	2.5	1.1	1.4
7	BA	3.0	1.5	1.5
8	BA	3.5	2.1	1.4
9	BA	2.8	1.0	1.8
10	BA	3.0	1.3	1.7

^aThe treatment sequence is tart cherry juice administration (A) or the control comparison of fruit juice administration (B).

^bThe dependent outcome variable is body inflammation as measured by serum hs-CRP (mg/L) taken at the end of each phase.

^cAbsolute value of the difference between Phase 1 and 2 within each participant.

NOTE: Information from this table is available online at www.jandonline.org as part of a PowerPoint presentation.

is a statistically significant difference through the use of the paired *t* test. This type of analysis does not allow carryover, phase, and sequence effects to be considered that have potential confounding effects on serum hs-CRP.

Table 3 illustrates how the data for the example 2×2 crossover examining the influence of tart cherry juice on inflammation would be entered into a spreadsheet for statistical analyses. There are different approaches to analyzing

the data. The data can be analyzed for treatment and sequence effects by using a mixed-design (split plot) analysis of variance. This type of analysis has at least one within-subjects variable and one between-subjects variable, and in this example treatment is the within-subjects variable (all participants get all levels of the treatment; that is, tart cherry juice and fruit juice) and sequence is the between-subjects variable (participants are in only one level of this because they are randomized to either sequence AB or BA). Another approach is using a mixed-linear model with repeated measures, controlling for phase. Consultation with a statistician during the designing phase of a study (ie, before seeking institutional review board approval or funding) can assist with identifying the appropriate analyses to determine treatment effects while controlling for carryover, phase, or sequence effects.

In the detailed example, the dependent outcome variable, serum hs-CRP level, was only measured once after the manipulation of each level of the independent variable. In a crossover design, the dependent outcome variable may be measured more than once in each phase. For example, the dependent outcome variable may be measured before the manipulation and after the manipulation (see Table 4). These additional measures can assist with statistically determining the occurrence of carryover, phase, or sequence effects. Additional resources for the appropriate analysis of crossover designs can be found online or in a variety of text books.^{18,19}

Sample Size Considerations

As previously mentioned, one of the advantages of using a crossover vs a randomized clinical trial design is the reduction in the number of participants needed for the study. The main difference in determining sample size in these two designs is the formulas for determining variances.^{12,20,21} For the randomized clinical trial design formula the sum of the variance for each group's outcome data is used. In contrast, the crossover design formula uses the variance of the treatment differences between outcome data. For the tart cherry juice and inflammation data in Table 1, the variance for the

Table 2. Example descriptive statistics for 2×2 crossover design examining the influence of tart cherry juice and control fruit juice on body inflammation as measured by serum high-sensitivity C-reactive protein (hs-CRP) calculations

Measure	Within-Sequence Group Comparison								Between-Group Comparison	
	AB ^a				BA ^b				TCJ _{AB} - TCJ _{BA}	FJ _{AB} - FJ _{BA}
	TCJ _{AB} ^c	FJ _{AB} ^d	Diff _{AB} ^e	Avg _{AB} ^f	TCJ _{BA}	FJ _{BA}	Diff _{BA} ^g	Avg _{BA} ^h		
Mean	1.4	2.8	1.4	2.1	1.4	3.0	1.6	2.2	-0.0	0.1
Standard deviation	0.4	0.2	0.3	0.3	0.4	0.4	0.2	0.4	0.2	0.4
n	5	5	5	5	5	5	5	5	5	5

^aAB=sequence AB (tart cherry juice, control fruit juice).

^bBA=sequence BA (control fruit juice, tart cherry juice).

^cTCJ=tart cherry juice.

^dFJ=control fruit juice.

^eDiff_{AB}=mean difference in serum hs-CRP levels between tart cherry juice and control fruit juice in sequence AB.

^fAvg_{AB}=mean average of the serum hs-CRP levels for tart cherry juice and control fruit juice values in sequence AB.

^gDiff_{BA}=mean difference in serum hs-CRP levels between tart cherry juice and control fruit juice in sequence BA.

^hAvg_{BA}=mean average of the serum hs-CRP levels for tart cherry juice and control fruit juice values in sequence BA.

Table 3. Example of data entry spreadsheet for 2×2 crossover design examining the influence of tart cherry juice and control fruit juice on body inflammation as measured by serum high-sensitivity C-reactive protein (hs-CRP) with one measure of inflammation collected in each phase

Participant	Group ^a	TCJ _{post} mg/L ^b	FJ _{post} mg/L ^c
1	1	1.3	3.0
2	1	1.5	2.8
3	1	2.0	3.1
4	1	1.0	2.7
5	1	1.1	2.5
6	2	1.1	2.5
7	2	1.5	3.0
8	2	2.1	3.5
9	2	1.0	2.8
10	2	1.3	3.0

^aRandomized groups with 1=sequence AB (tart cherry juice, fruit juice) and 2=sequence BA (fruit juice, tart cherry juice).

^bTCJ_{post}=serum hs-CRP level after tart cherry juice manipulation.

^cFJ_{post}=serum hs-CRP level after control fruit juice manipulation.

NOTE: Information from this table is available online at www.jandonline.org as part of a PowerPoint presentation.

Table 4. Example of data entry spreadsheet for 2×2 crossover design examining the influence of tart cherry juice and control fruit juice on body inflammation as measured by serum high-sensitivity C-reactive protein (hs-CRP) with two measures of inflammation collected in each phase

Participant	Group ^a	TCJ _{pre} mg/L ^b	TCJ _{post} mg/L ^c	FJ _{pre} mg/L ^d	FJ _{post} mg/L ^e
1	1	2.0	1.3	2.1	3.0
2	1	2.1	1.5	2.0	2.8
3	1	2.5	2.0	2.5	3.1
4	1	1.8	1.0	1.8	2.7
5	1	1.9	1.1	1.9	2.5
6	2	2.0	1.1	2.0	2.5
7	2	2.1	1.5	2.1	3.0
8	2	2.8	2.1	2.8	3.5
9	2	1.9	1.0	1.9	2.8
10	2	2.4	1.3	2.5	3.0

^aRandomization groups where 1=sequence AB (tart cherry juice, fruit juice) and 2=sequence BA (fruit juice, tart cherry juice).

^bTCJ_{pre}=serum hs-CRP level before tart cherry juice manipulation.

^cTCJ_{post}=serum hs-CRP level after tart cherry juice manipulation.

^dFJ_{pre}=serum hs-CRP level before control fruit juice manipulation.

^eFJ_{post}=serum hs-CRP level after control fruit juice manipulation.

NOTE: Information from this table is available online at www.jandonline.org as part of a PowerPoint presentation.

tart cherry inflammation data is 0.154 and that for the placebo is 0.090. The variance for the differences between the tart cherry and placebo inflammation data is 0.049. When we add the separate variances for each treatment and divide the sum by the variance for the differences the answer is 5. That means for a randomized clinical trial design with the same variables, five times the number of participants would be needed in each group to yield a similar study power and level of significance. The sample size calculations demonstrate the advantage of a lower sample size for the crossover design when there are no carryover or period effects.

The ability to calculate an appropriate sample size for both designs needs preliminary data that capture means and variances, as well as minimal meaningful clinical outcomes. These data may come from pilot work of investigators, or from previously published literature using similar samples and manipulations.

Examples of Crossover Design in Nutrition Research

Nutrition research questions tested within crossover designs can examine the effects of manipulating levels of the independent variable within a laboratory setting or within a field or applied setting. Crossover designs used in laboratory settings commonly examine the effects of a dietary manipulation over a brief period (minutes to hours) on amounts and types of foods consumed in a meal and/or a later eating occasion; sensations of hunger, satiation, and satiety; and/or hormonal and cardiometabolic responses occurring after an eating occasion. In these types of studies, the food or nutrient component being examined in the dietary manipulation is provided to participants. In addition, because the dietary manipulation is short term, the washout period is brief (1 day to 1 week), allowing the study to be completed within a few weeks for each participant. Crossover designs used in field or applied settings commonly examine the effect of a dietary manipulation over a lengthier time period (potentially a few days to several months) on anthropometric and/or biochemical data, nutrition-focused physical findings, and other health parameters (ie, blood pressure). For the dietary manipulation to occur, participants may be provided with the food or nutrient component to be eaten in their usual daily life, or they may be instructed on the types of foods and nutrients to be consumed, with the instruction being such that a permanent change in intake is not expected (ie, being provided with a very structured menu to follow, but strategies to assist with behavior change, such as motivational interviewing, would not be provided). In these studies, because the dietary manipulation is over a longer period of time and the dependent variables measured may need a longer period of time to recover to baseline levels due to the relationship between diet and the health parameter, the washout period is longer than that found in studies conducted in laboratory settings. Thus, the overall length of the study for a participant is usually several weeks to several months.

Example of a Crossover Study: Laboratory Setting

Wijnhoven and colleagues¹⁹ examined whether providing a greater variety of foods within a meal would increase energy intake in the meal in older women who self-reported having poor appetite. In this study, the independent variable was dietary variety within a meal, with the independent variable

having two levels, variety (three different proteins, vegetables, and starch components provided in a dinner meal) and no variety (only one protein, vegetable, and starch component provided in a dinner meal). The overall gram amount of food, energy density, and macronutrient composition of the provided meals in the two conditions were similar. The primary dependent variable was amount of energy consumed in the meal, with measures of hunger, fullness, and liking of the foods also obtained.

Participants were randomized to a sequence: test meal (variety), control meal (no variety) or control meal, test meal. The two meal sessions were completed within a 2-week period. Because the dependent variables were measured within the actual meal, and the influence of the independent variable was believed to have an effect only within the meal, the washout period was brief. Each meal session was approximately 4 hours in length, with the manipulated meal occurring at the end of the session. The sessions were conducted in a laboratory setting, with the meals provided to participants by the research team. The dependent outcome variables were collected during each meal session.

In this investigation, the authors anticipated having an effect size of $d=0.68$ on energy intake between the two test meals (d represents the differences between two means divided by the pooled standard deviation of the data). With that effect size, the investigators anticipated needing 17 participants total for the study, not accounting for participant attrition, with 80% power and significance level <0.05 . In a case where this same effect size was used to determine the sample size needed while achieving 80% power to find significance with $\alpha<.05$ within a randomized clinical trial, the number of participants needed would be 70. The difference in the number of participants required in the two types of experimental designs, 17 in the crossover design and 70 in the randomized clinical trial, highlight one of the strengths of the crossover design.

In the statistical analyses, the treatment (diet) effect was examined using a mixed-linear model with repeated measures, with phase controlled. Sequence and carryover effects were not examined. This may be due to the belief that the independent variable would have no influence on the dependent variables outside the actual meal in which it was implemented. However, the events occurring during the two phases may be different (ie, a participant's level of hunger going into the session), and by controlling for phase in the analysis, the independent influence of the diets on the dependent variables could be tested.

Reported results were that the mean energy intake in the test meal with variety was 427 ± 119 kcal and 341 ± 115 kcal for the control meal without variety. A statistically significant ($P<0.05$) mean difference of 79 kcal (95% CI 25 to 134 kcal, meaning that there is 95% confidence that the true value of the mean difference is between 25 and 134 kcal, which is different from 0 kcal difference) in consumption occurred between the two meals, with the meal with variety producing greater energy intake.

Example of a Crossover Study: Field/Applied Setting
Santiago-Torres and colleagues²⁰ examined metabolic responses to two diets, a traditional Mexican diet and US diet, in first- and second-generation healthy women of Mexican

descent. The independent variable in this study, diet, had two levels. One level was a traditional Mexican diet, which was based on peer-reviewed publications and an historical review of food composition of traditional Mexican diets. The second level was a traditional US diet, which was based on the contribution of foods and beverages to the total intake of food items reported in the 2003-2004 National Health and Nutrition Examination Survey. The diets were similar in energy and macronutrient content. Each diet was implemented for 24 days. The dependent outcome variables included fasting serum glucose, insulin, insulin-like growth factor 1, insulin-like growth factor binding protein 3, adiponectin, CRP, and interleukin-6, and the homeostasis model assessment of insulin resistance.

Participants were randomized to a sequence: traditional Mexican diet, traditional US diet, or traditional US diet, traditional Mexican diet. A 28-day washout period occurred between the diet phases to eliminate the potential effects of the previously implemented diet on metabolic markers. The diets were implemented using a 7-day menu rotation, with all foods and beverages carefully prepared, packaged, and provided to participants three times per week. The amount of energy provided in the food and beverages to participants was designed to maintain participants' weight within 3%, and weight was monitored three times per week. Participants were instructed to eat or drink only the foods and beverages provided to them, except for drinking water, and reported when foods or beverages were consumed that were not provided by the study. Dependent variables were measured at the beginning and end of each diet phase.

It was anticipated that the general effect size that would be found between the two diets would be $d=0.40$. Given that effect size, the investigators anticipated needing 50 participants total for the study, not accounting for participant attrition, with 80% power and significance level <0.05 . When this same effect size was used to determine the sample size needed while achieving 80% power to find significance with $\alpha<.05$ within a randomized clinical trial, the number of participants needed would be 200. This difference in sample size needed to acquire the same amount of power in the investigation given a specific effect size again highlights this strength in the crossover design.

Data were analyzed using linear mixed models so that treatment (diet) effect could be examined. For this analysis, mean changes from pre- to postdiet phase were compared, with sequence, phase, and baseline and washout values controlled. This type of analysis allows testing of the

Table 5. Latin square showing sequence order of three different levels, represented by A, B, and C, of one independent variable in a crossover design

Order	Phase 1	Phase 2	Phase 3
1	A	B	C
2	B	C	A
3	C	A	B

NOTE: Information from this table is available online at www.jandonline.org as part of a PowerPoint presentation.

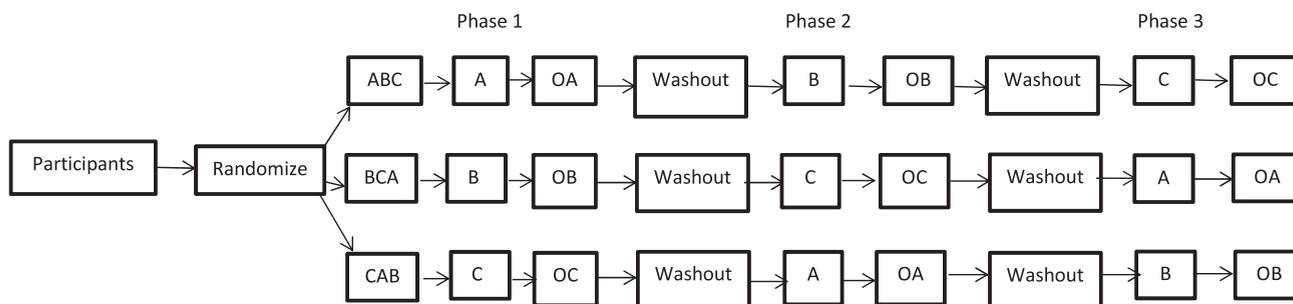


Figure 2. A 3×3 crossover design. A, B, or C denotes one of three levels of the independent variable; ABC, BCA, CAB denotes sequence of manipulation of the levels of the independent variable in the crossover design; OA denotes dependent outcome variable measured after manipulation A; OB denotes dependent outcome variable measured after manipulation B; OC denotes dependent variable outcome measured after manipulation C; and Washout denotes period of removal of manipulation for specified time interval before initiating second or third manipulation of crossover design. (NOTE: Information from this figure is available online at www.jandonline.org as part of a PowerPoint presentation.)

occurrence of sequence and phase effects while controlling for them in examining treatment effect (allowing for examination of the independent effect of the diets on the change in the dependent variables). The carryover effect was not examined, but by controlling for baseline and washout values and using change scores in the analyses, carryover effects are controlled for in the analysis, allowing for the examination of the independent effect of the diets on change in the dependent variables.

Compared with a traditional US diet, a traditional Mexican diet significantly ($P<0.05$) reduced insulin by 14%, homeostasis model assessment of insulin resistance by 15%, and insulin-like growth factor binding protein 3 by 6%. There were no significant diet effects on serum concentrations of glucose, adiponectin, CRP, interleukin-6, or insulin-like growth factor 1.

Additional Considerations: Independent Variable with More than Two Levels

In nutrition research using crossover designs, the independent variable may have more than two levels. It is not uncommon for there to be three levels of the independent variable, two different dietary manipulations, and a control comparison. When there are more than two levels of the independent variable, the different number of sequences in which the levels of the independent variable can be implemented in participants increases. For example, when there are three levels of the independent variable, A, B, and C, there are six different sequences in which the levels can be implemented (ABC, ACB, BCA, BAC, CAB, and CBA). Ideally, to examine the potential for sequence effects in a crossover design, participants need to be randomized to all potential sequences. This design may increase the number of participants needed in the study, potentially more than the number needed for adequate power to find the treatment effect. When one does not want to use all possible sequences due to the number of participants required, one common strategy used in a crossover design with more than two sequences is to use a Latin Square design (Table 5) in determining the sequences to use in the trial. The Latin Square uses sequence orders that allow each manipulation, or level of the independent variable, to occur

in each phase. The sequence orders of ABC, BCA, and CAB allow each level to occur in each phase, reducing the occurrence of phase effects while minimizing the number of sequences used in the study, decreasing the number of potential participants required. However, because not all sequences are used, it is not possible to completely ascertain whether a sequence effect does or does not occur in the investigation. See Figure 2 for an example of a 3×3 crossover design in which there are three levels in the independent variable and a Latin Square (see Table 5) has been used to determine the sequences to which participants should be randomized.

CONCLUSIONS

The crossover design is a useful approach in nutrition research, particularly when treatment effects are temporary and baseline levels are achievable when the dietary manipulation is removed. This research design has a number of advantages, particularly due to the increased utility of within-subject analysis. In addition, sample sizes required are dramatically reduced. Thus, this study design can reduce resources needed for a study, yet allow adequate power to find clinically relevant outcomes, supported by statistical significance. With careful design and statistical analyses, potential effects that can confound the interpretation of results can be controlled. Due to its strengths, registered dietitian nutritionists and nutrition researchers should consider using crossover designs in nutrition research.

References

1. Cornu C, Kassai B, Fisch R, et al. Experimental designs for small randomised clinical trials: an algorithm for choice. *Orphanet J Rare Dis.* 2013;8:48.
2. Foley RN. Randomized controlled trials 2: Analysis. *Methods Mol Biol.* 2015;1281:177-190.
3. Sedgwick P. Randomised controlled trials: "Within subject" versus "between subject" designs. *BMJ.* 2014;349:g6435.
4. Thiese MS. Observational and interventional study design types; an overview. *Biochem Med (Zagreb).* 2014;24(2):199-210.
5. Diaz-Urriarte R. Incorrect analysis of crossover trials in animal behavior research. *Animal Behavior.* 2002;63:815-822.
6. Hallas J, Pottegård A. Use of self-controlled designs in pharmacoepidemiology. *J Intern Med.* 2014;275(6):581-589.

7. Josephy H, Vansteelandt S, Vanderhasselt MA, Loeys T. Within-subject mediation analysis in AB/BA crossover designs. *Int J Biostat.* 2015;11(1):1-22.
8. Kieser M, Rauch G. Two-stage designs for cross-over bioequivalence trials. *Stat Med.* 2015;34(16):2403-2416.
9. Leonard G, Lafrenaye S, Goffaux P. Randomized placebo-controlled cross-over designs in clinical trials: A gold standard to be reassessed. *Curr Med Res Opin.* 2012;28(2):245-248.
10. Matthews JN, Henderson R. Two-period, two-treatment crossover designs subject to non-ignorable missing data. *Biostatistics.* 2013;14(4):626-638.
11. Rosner B. *Fundamental of Biostatistics.* 8th ed. Boston, MA: Brooks Cole; 2016.
12. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: Part 18 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2012;109(15):276-281.
13. Yan Z. Comments on 'Estimation in AB/BA crossover trials with application to bioequivalence studies with incomplete and complete data designs'. *Stat Med.* 2013;32(30):5484-5486.
14. Sedgwick P. Bias in randomised controlled trials: Comparison of crossover group and parallel group designs. *BMJ.* 2015;351:h4283.
15. Tvete IF, Olsen IC, Fagerland MW, et al. An approach to combining parallel and cross-over trials with and without run-in periods using individual patient data. *Clin Trials.* 2012;9(2):164-175.
16. Yan Z, Hosmane B, Locke C. Crossover versus parallel designs: Dose-escalation design comparisons for first-in-human studies. *J Biopharm Stat.* 2013;23(4):804-817.
17. Prasad V, Grady C. The misguided ethics of crossover trials. *Contemp Clin Trials.* 2014;37(2):167-169.
18. Lui KJ. *Crossover Designs: Testing, Estimation, and Sample Size.* Hoboken, NJ: John Wiley and Sons; 2016.
19. Raghavarao D, Padgett L. *Repeated Measurements and Cross-Over Designs.* Hoboken, NJ: John Wiley and Sons; 2014.
20. Chiang C, Hsiao C-F, Liu J-P. Sample size determination for individual bioequivalence inference. *PLoS ONE.* 2014;9(10):1-7.
21. Golkowski D, Friede T, Kieser M. Blinded sample size re-estimation in crossover bioequivalence trials. *Pharm Stat.* 2014;13(3):157-162.

AUTHOR INFORMATION

J. E. Harris is a professor and chair of the Department of Nutrition, West Chester University, West Chester, PA. H. A. Raynor is a professor, Department of Nutrition, University of Tennessee, Knoxville.

Address correspondence to: Hollie A. Raynor, PhD, RD, LDN, Department of Nutrition, University of Tennessee, 1215 Cumberland Ave, Knoxville, TN 37996. E-mail: hraynor@utk.edu

STATEMENT OF POTENTIAL CONFLICT OF INTEREST

No potential conflict of interest was reported by the authors.

FUNDING/SUPPORT

There is no funding to disclose.

ACKNOWLEDGEMENTS

The authors thank the members of the *Journal of the Academy of Nutrition and Dietetics* STATS Committee for their contributions.