

Crossover design

Shengping Yang PhD, Gilbert Berdine, MD

ABSTRACT

I am planning a clinical trial to compare two diets on reducing the risk of type II diabetes. Because there is a restriction on the total budget, I would prefer to enroll a small number of participants. Meanwhile, it is important that there is sufficient statistical power to detect a clinically meaningful difference. Is there any study design that can be utilized?

Keywords: crossover design, carryover effect, sample size

Although smaller sample size is often associated with lower statistical power, certain study designs can be utilized to achieve improved power with the same sample size. One of such designs is called the crossover design. Specifically, a crossover design is a longitudinal study design that has multiple intervention periods, with a “washout” period(s) in-between. Specifically, each subject receives a sequence of interventions in the order of intervention — “washout” — intervention . . . In nearly all crossover designs, both the number of interventions a subject receives, and the number of periods a subject participates are the same for all subjects.

There are both advantages and limitations of utilizing a crossover design. We will use the most commonly used crossover design, which is the 2×2 crossover design, to illustrate how it works.

1. 2×2 CROSSOVER DESIGN

1.1 THE LAYOUT OF A 2×2 CROSSOVER DESIGN

There are 2 sequences, 2 periods and 2 interventions in a 2×2 crossover design. To implement such

Corresponding author: Shengping Yang
Contact Information: Shengping.yang@pbrc.edu
DOI: 10.12746/swrccc.v7i30.564

Table 1. The layout of a two-period crossover design

	Period 1	Washout period	Period 2
Sequence 1	A	—	B
Sequence 2	B	—	A

a design, all subjects are first randomized into one of the two sequences (Table 1). Subjects randomized to sequence 1 receive intervention A in period 1 and “crossover” to intervention B in period 2; and those randomized to sequence 2 receive intervention B in period 1 and “crossover” to intervention A in period 2. Note that there is a hallmark “washout” period between periods 1 and 2, i.e., after the completion of period 1 intervention, instead of starting period 2 intervention immediately, there will be a period of no intervention. The “washout” period is designed to affirm that the effect of the period 1 intervention is worn off, so that the effect of period 2 intervention attributes to that intervention, not period 1 intervention.

1.2 THE KEY ASSUMPTION

The key assumption of a crossover design is that there is no carryover effect (the effect of intervention from one period, even after the “washout” period, continues to affect a subject in the subsequent period). Ideally, in a 2×2 crossover study, the effect of the first period disappears completely before the start of the second period. However, in reality, there are virtually

always subtle differences between the two periods; for example, subjects might be more familiar with the intervention/protocol in the second period compared to the first period, and thus respond differently to the second intervention should it be received in the first period. Due to this assumption, carryover design is most useful for studying chronic and stable diseases, and is usually inappropriate for studying curable diseases. For example, if a subject is cured in the first period, then no matter what intervention this subject receives in the second period, the outcome is always cured. Therefore, applying crossover design to curable diseases invalidates the non-carryover effect assumption.

1.3 INTERPRETATION OF CROSSOVER DESIGN AS A TWO-GROUP COMPARISON

We present the cell mean model of a crossover design in Table 2. In period 1, subjects in sequence 1(2) receive intervention A(B), and thus the expected value is the summation of overall mean, effect of intervention A(B) and effect of period 1. In period 2, subjects in sequence 1(2) receive intervention B(A), and thus the expected value is the summation of overall mean, effect of intervention B(A), effect of period 2, and the carryover effect $\lambda_A(\lambda_B)$ due to the first period intervention.

Like parallel-group trials, the primary goal of a crossover study is to compare the effect of interventions A and B. However, because each subject receives both interventions, a direct comparison

Table 2. The cell mean model

	Period 1	Period 2
Sequence 1	$E(Y_{i11}) = \mu + \tau_A + \rho_1$	$E(Y_{i12}) = \mu + \tau_B + \rho_2 + \lambda_A$
Sequence 2	$E(Y_{i21}) = \mu + \tau_B + \rho_1$	$E(Y_{i22}) = \mu + \tau_A + \rho_2 + \lambda_B$

μ : Overall mean
 $i = 1, \dots, n_i$, where n_i is the number of subjects in sequence i
 τ_A : Effect of intervention A
 τ_B : Effect of intervention B
 ρ_1 : Effect of period 1
 ρ_2 : Effect of period 2
 λ_A : Carryover effect of intervention A
 λ_B : Carryover effect of intervention B

of subjects received A with those received B is not possible, and thus the following analysis procedure should be used: (1) obtain the differences between periods 1 and 2, for subjects in both sequences. According to the cell mean model, $E(Y_{i11} - Y_{i12}) = \tau_A - \tau_B + \rho_1 - \rho_2 - \lambda_A$, and $E(Y_{i21} - Y_{i22}) = \tau_B - \tau_A + \rho_1 - \rho_2 - \lambda_B$. Assuming that there is no carryover effect, i.e., $\lambda_A = \lambda_B$, then $E(Y_{i11} - Y_{i12}) - E(Y_{i21} - Y_{i22}) = 2\tau_A - 2\tau_B$, which equals to the difference between interventions A and B. Furthermore, assuming that the distribution of the outcome measurement follows a normal distribution, the comparison of the two interventions can be made by comparing the two-period difference between the two sequences using a two-sample t test. In situations where a normal distribution assumption is questionable, such a comparison can be made by replacing the t test with a Mann-Whitney U test. Note that such a comparison is valid only when there is no carryover effect. Therefore, before making the above test, a test of carryover effect should be performed. Specifically, let $E(Y_{i11} + Y_{i12}) = 2\mu + \tau_A + \tau_B + \rho_1 + \rho_2 + \lambda_A$, and $E(Y_{i21} + Y_{i22}) = 2\mu + \tau_A + \tau_B + \rho_1 + \rho_2 + \lambda_B$, then $E(Y_{i11} + Y_{i12}) - E(Y_{i21} + Y_{i22}) = \lambda_A - \lambda_B$. Then, testing equal carryover effect is equivalent to compare the sum of periods 1 and 2 between the 2 sequences.

A mixed effect regression can also be used to model a crossover design, and is more flexible. However, this is not the focus of this article.

2. THE ADVANTAGES OF USING A CROSSOVER DESIGN

2.1 EACH SUBJECT SERVES AS HIS/HER OWN CONTROL

In general, each subject in a crossover study receive all intervention at least once. The advantage of this design is that comparison between interventions can be made without taking into account of confounding factors, such as gender and age, because each subject serves as his/her own control. This is highly desirable in data analysis especially when the sample size is small. Due to small number of subjects, the distributions of the confounders might be very different (after randomization) between groups in a parallel-group design, which makes it difficult to separate the true intervention effect from the confounder

effect. Note that in a crossover study, because each subject serves as his/her own control, confounding effects are removed by design.

2.2 IMPROVED EFFICIENCY

Because each subject is measured multiple times, e.g., 2 times (one in period 1 and another in period 2) in a 2×2 crossover design, data variance can be substantially reduced due to correlation between repeated measurements on the same subject, which can be directly translated into improved statistical power. We will demonstrate this advantage by comparing sample size required by a crossover design and a two-sample t test.

2.2.1 SAMPLE SIZE CALCULATION FOR A TWO-SAMPLE t TEST

For a given type I error of α and type II error of β , assuming that both groups have the same group size, sample size for a two-sample t test can be approximated (using z approximation) by the following,

$$N = 2 \frac{(z_{1-\alpha/2} + z_{\beta})^2 (\sigma_1^2 + \sigma_2^2)}{\Delta^2},$$

where z_{γ} is the γ^{th} percentile of the standard normal distribution, and $\Delta = \mu_1 - \mu_2$ is difference between the two group means, σ_1^2 and σ_2^2 are the variance of the two groups, respectively.

2.2.2 SAMPLE SIZE CALCULATION FOR CROSSOVER DESIGN

Sample size for a 2×2 crossover design can be calculated by using,

$$N = \frac{(z_{1-\alpha/2} + z_{\beta})^2 \sigma_m^2}{\Delta^2},$$

where σ_m^2 is the variance in crossover design. Because of correlation between repeated measures, σ_m^2 is usually much smaller than σ_1^2 and σ_2^2 (details not

shown), and thus sample size required for a crossover design can be much smaller than that required for a two-sample t test, given the same type I and type II errors.

3. LIMITATIONS

While crossover design has its advantages over parallel-group design, it can sometimes have serious limitations.

3.1 CARRYOVER EFFECT IS NOT NEGLIGIBLE

The success of a crossover design is largely dependent on how large the carryover effect is. If the carryover effect is negligible, then a crossover design has better efficiency than a parallel-group design. On the other hand, if carryover effect is strong, then it confounds the intervention effect, which prevents the intervention effect from being estimated correctly. Not surprising, testing for carryover effect is a critical step in analyzing crossover design data. However, statistical power for testing carryover effect is always much lower than that for testing intervention effect; thus a non-significant carryover effect test result does not necessarily mean that carryover effect should be ignored. Additionally, should a significant carryover effect be detected, the arguably best approach for data analysis is to perform comparison by using data collected from the first period only, which is not efficient.

4. OTHER CONSIDERATIONS

In a crossover design, although the number of subjects required is much less compared to a parallel-group design, the costs associated with a crossover study might not be reduced proportionally, especially for those with a large number of repeated measurements.

Depending on the nature of a specific study, a crossover study could have long duration (considering there is also a “washout” period(s)), which might result in high dropout rate. Appropriate adjustments

need to be made to make sure that enough subjects complete all interventions.

In summary, crossover design has improved efficiency compared to a parallel-group design, and is highly useful when recruitment is a challenge and/or cost associated with a trial is very high. Meanwhile, crossover design is most suitable for trials on chronic diseases, ideally with minimal carryover effect, given a sufficient “washout” period.

Article citation: Yang S, Berdine G. Crossover design. *The Southwest Respiratory and Critical Care Chronicles* 2019;7(30):63–66

From: Department of Biostatistics (SY), Pennington Biomedical Research Center, Baton Rouge, LA; Department of Internal Medicine (GB), Texas Tech University Health Sciences Center, Lubbock, Texas

Submitted: 7/2/2019

Conflicts of interest: none

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

REFERENCES

1. Chow SC, Shao J, Wang H. *Sample Size Calculations in Clinical Research*. Chapman and Hall/CRC Press, New York, USA. 2003.
2. Grizzle JE. The two-period change-over design and its use in clinical trials. *Biometrics* 1965;21(2):467–480.
3. Siyasinghe NM, Sooriyarachchi MR. Guidelines for calculating sample size in 2×2 crossover trials: a simulation study. *J Natn Sci Foundation Sri Lanka* 2011;39(1):77–89.