

## Cross Over Design

Cross over design is commonly used in various type of research for its unique feature of accounting for within subject variability. For studies with short length of treatment time, illness that will not be altered its baseline characteristics after treatment, or endpoints that are individual-dependent and subjective, etc., a cross-over study design will be a good choice since it usually requires less patients and more cost efficient. However, a cross-over study design will not be appropriate or feasible if the study length is long or the testing procedures will change the baseline of the illness (such as cancer research).

A cross over design allows each subject serves as his/her own control such that the within subject variability can be accounted for and therefore reduces the random error. In some cases, a cross over design provides a more sensitive testing because it renders a more “precise” estimate of variability than a parallel study design will. For “precise” or “precision”, we define “precision” as the number of experimental units needed for the estimate of variance from a parallel study design will have the same variance as from a cross-over study design. To demonstrate it, let the variance from a cross-over study be

$var(CR) = \frac{\sigma_w^2}{c}$  and the variance from a parallel study be  $var(par) = \frac{2(\sigma_w^2 + \sigma_b^2)}{p}$ , where  $\sigma_w^2$  = within subject variability and  $\sigma_b^2$  = between subject variability. The “precision” of a parallel study to a cross-

over study =  $\left(\frac{parallel}{cross\ over}\right) = \frac{p}{c} = \frac{\frac{2(\sigma_w^2 + \sigma_b^2)}{p}}{\frac{\sigma_w^2}{c}} = \frac{2(\sigma_w^2 + \sigma_b^2)}{\sigma_w^2} > 1$ . That is, a parallel study design will require

more study subjects in order to have the same level of variance from a cross-over design (note that the more study subjects a study has, the lower the variance likely to be since  $variance = \frac{sum\ of\ error}{n}$ ).

However, a cross over study requires careful planning and execution. A cross-over study with statistically significant carry-over effect will be difficulty to draw definite conclusions regarding the testing effect. Careful planning and execution of the study may reduce or avoid the carry-over effect by placing adequate length of wash-out period or utilizing statistically valid testing sequences. Also, a less complicated cross-over study may reduce the chance of having carry-over effect, study subjects drop out, or withdraw (therefore less missing data).

In this article we will first discuss the statistical model and analyses for a 2\*2 cross-over design, followed by a bioequivalence study design as an example. Lastly we will briefly discuss the statistical analyses for a higher-order 2\*2 cross-over design.

### Statistical Model for a 2\*2 Cross Over Study – Continuous Endpoint

A 2 (treatment)\*2 (period) cross-over study design indicates a study with 2 treatments and 2 periods for each study subject. For a 2\*2 cross-over design where the endpoints of interest are continuous, the statistical model can be expressed as :

$$\mu_{ijk} = \mu + Sub_k + Period_j + Treatment_i + \varepsilon_{ijk}$$

where  $k=1,2, \dots, ni$  = number of study subjects in each testing group;

$j=1,2$  period;

*i=1,2 treatment;*

One big difference in analyzing a cross-over study is that we should test for period effect using the Type I sum of square error before we proceed to testing for treatment effect. If the period effect is statistically significant, the treatment effect at the 2<sup>nd</sup> period will not be feasible for interpretation since the results from the 2<sup>nd</sup> period consist of the residual effect from the 1<sup>st</sup> period and the treatment effect from the current period. Therefore, the type I sum of square error should be assessed first to ensure no period effect prior to proceeding to assessing the treatment effect. If there was no evidence of period effect, the treatment effect can be tested using the random error from the within-subject random error. Collecting data prior to the beginning of the 2<sup>nd</sup> period can be informative since it can be used to assess whether the baseline at the beginning of the 2<sup>nd</sup> period is comparable to that of the 1<sup>st</sup> period.

A 2\*2 cross-over design can be analyzed using SAS® PROC GLM procedure with SUBJECT, SEQUENCE, TREATMENT, and PERIOD in the model. If the period effect is not statistically significant, we can proceed to estimate the least squared mean treatment effect. However, if the period effect is statistically significant, one should consider assess the treatment effect using the 1st period data only.

For example, a pharmaceutical company is planning a study testing the treatment for acute asthma on dilating the bronchial muscle when study subjects have an acute asthma episode. The efficacy endpoint is measured by the Forced Expiratory Volume (FEV). The example data and SAS codes are listed as following

```

data example;
input subject seq $ run_in P1 washout P2;
cards;
1 AB 1.09 1.28 1.24 2.33
2 AB 1.38 1.6 1.9 2.21
...
9 BA 1.74 3.06 1.54 1.38
10 BA 2.41 2.68 2.13 2.10
..
;
proc glm;
class sea subject period trt;
model fev1=seq subject(seq) period trt;
TEST H = SEQ E = SUBject(SEQ) / HTYPE=1 ETYPE=1;
lsmeans trt/pdiff;
run;

```

Note that the SEQ (sequence) should be tested using the SUBJECT error term as each subject is nested in each sequence. The ANOVA table for the analyses is displayed below :

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	18	15.23153546	0.84619641	6.61	0.0003
Error	15	1.91985278	0.12799019		
Corrected Total	33	17.15138824			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
seq	1	1.60796879	1.60796879	12.56	0.0029
subject(seq)	15	10.93591944	0.72906130	5.70	0.0009
period	1	0.07718824	0.07718824	0.60	0.4495
trt	1	2.61045899	2.61045899	20.40	0.0004

Source	DF	Type III SS	Mean Square	F Value	Pr > F
seq	1	1.60796879	1.60796879	12.56	0.0029
subject(seq)	15	10.93591944	0.72906130	5.70	0.0009
period	1	0.03323546	0.03323546	0.26	0.6178
trt	1	2.61045899	2.61045899	20.40	0.0004

Tests of Hypotheses Using the Type I MS for subject(seq) as an Error Term					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
seq	1	1.60796879	1.60796879	2.21	0.1582

The same analyses can be accomplished by using PROC MIXED model with PERIOD and TREATMENT as the fixed effects and the SUBJECT as the random effect.

**proc mixed;**

class seq subject period trt;

model fev1= period trt;

RANDOM SUBJECT(SEQ);

LSMEANS TRT / PDIFF CL E;

run;

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
period	1	15	0.26	0.6178
trt	1	15	20.40	0.0004

Statistical Model for a 2\*2 Cross Over Study – Binary Endpoint

In a 2\*2 cross-over design, if the endpoint of interest is a binary endpoint, several analysis methods can be considered, such as McNemar test or Mainland-Gart test. Although each method has its pros and cons, both methods ignore some data that have the same responses from each period or treatment. In order to maximize the use of the collected data, the repeated measure logistics regression is recommended as the foremost statistical method.

For example, a study is planned to test the safety profile of 2 compounds for the adverse event occurrence. Each subject received compound A/B at each period. The data input and the SAS codes used for analyses for the example are listed below:

```

data binarv;
input pat seq $ trtA trtB;
cards;
1 AB 1 0
2 AB 1 0
...
10 AB 0 1
11 BA 1 0
...
;
run;

proc genmod desc;
class seq pat trt(ref="A") ;
model outcome=trt / dist=bin link=logit ;
repeated subject=pat(seq) / corr=cs corrw;
run;

```

The safety profile difference between A and B can be compared by the logit (which can be back transform  $\exp(0.819)$ ) of the 2 treatments:

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z
Intercept		-0.4055	0.4564	-1.3001	0.4891	-0.89	0.3744
trt	A	0.8109	0.8640	-0.8825	2.5044	0.94	0.3480
trt	B	0.0000	0.0000	0.0000	0.0000	.	.

### Bioequivalence Study

A bioequivalence study is commonly performed in pharmaceutical research when it is necessary to demonstrate the equivalence of 2 conditions of interest. For example, the liquid form of a compound vs. a tablet form, total dose of difference combinations of dosage and the number of tablets (2\*400mg vs. 1\*800mg), a new generic form of compound vs. an existing compound with similar therapeutic purpose, or the metabolism profile of a pharmaceutical agent with or without meals.

When a bioequivalence study is being planned, a 2 (tested group vs. reference group) \*2 (periods) cross-over design will be considered as the study design of choice because of its unique feature of controlling the within subject variability. In the US, per the FDA guideline, the 20% rule is used to determine the bioequivalence of two compounds of interest. For a test compound, it can be claimed to be bioequivalent to a reference compound if the difference in AUC or  $C_{max}$  between 2 compounds is within 20% of the reference group.

Two approaches can be used to demonstrate the equivalence:

1. Confidence Interval Approach:
  - *difference*  $(\mu_t, \mu_R) = \mu_t - \mu_R \leq \pm 20\%$  of  $\mu_R$ ; or
  - $80\% \leq \text{ratio}(\mu_t, \mu_R) \leq 120\%$ ;
2. Interval Hypotheses Testing
  - $T$  is not worse than  $R$  by  $\Delta$  and  $T$  is not superior to  $R$  by  $\Delta$ ;  
 $\rightarrow H_0: \mu_t - \mu_r \leq \theta_L$  or  $\mu_t - \mu_r \geq \theta_U$ ;  
 $H_a: \theta_L < \mu_t - \mu_r < \theta_U$

One can obtain the estimate of the variance and the estimated mean difference using the method described above.

The following example illustrates the calculation. A manufacture is planning a study to demonstrate its generic compound being bioequivalent to the branded compound on the market. Each of the 7 volunteer subjects received 2 compounds, test compound and the branded compound in each of the 2 periods. The area under the curve (AUC) was used as the endpoint to demonstrate the equivalence. The example data and analyses are shown below.

```
data BE;
input seq $ subj Period1 Period2 ;
cards;
TR 1 75 73
TR 2 96 93
..
TR 8 74 37
TR 9 86 51
..
;
run;
proc glm;
class seq subj period trt;
model AUC1=seq subj(seq) period trt;
lsmeans trt;
run;
```

The partial results that are relevant to the calculation from the SAS output are listed below.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	10718.25000	714.55000	4.83	0.0045
Error	12	1773.85714	147.82143		
Corrected Total	27	12492.10714			

trt	AUC1 LSMEAN
R	84.0714286
T	89.5714286

As the ANOVA table shows, the MSE (estimate of variance) is 147.82143. Based on the 20% rule, using the 90% confidence interval approach,

$$\begin{aligned}
 \text{90\% CI for } (\bar{Y}_T - \bar{Y}_R) &= (L_1, U_1) \\
 &= (89.57 - 84.07) \pm t_{0.05,12} * \text{MSE} \\
 &= 5.5 \pm 1.77 * \sqrt{147.82} \\
 &= (-16.02, 27.02);
 \end{aligned}$$

The 20% limit based on the reference group (R) = 20% of  $\bar{Y}_R = 0.2 * 84.07 = 16.81$ ; so the equivalence limit = (-16.81, 16.81). The upper bound of the 90% CI exceeds the upper limit of 16.81. Therefore the T and R are not bioequivalent.

We can also use the ratio of the 2 means to assess the equivalence. The bioequivalence can be established if the 90% CI for the ratio is within (80%, 120%).

$$\begin{aligned}
 \text{The 90\% CI for } \frac{\mu_T}{\mu_R} &= (L_2, U_2) \\
 &= [ (\frac{L_1}{\bar{Y}_R} + 1) * 100\%, (\frac{U_1}{\bar{Y}_R} + 1) * 100\% ] \\
 &= (80.9\%, 132\%)
 \end{aligned}$$

Since the upper bound of the confidence exceeds the upper limit of 120%, T and R can not be claimed to be bioequivalent based on the 20%.

#### Higher-order designs for 2\*2 cross-over design

A high-order design is a 2 treatment cross-over with more than 2 periods or sequences. The main reason that a higher-order design is needed is that the carry-over effect from a single 2\*2 cross-over replicate is aliased with the treatment by period effect. In order to separate the carry-over and the interaction effects, a 2-treatment cross over design with more than 1 replicate is required. The goal of an optimal higher-order design is to select a study design that will render a minimum random error. Following are some examples of higher-order designs for a 2-treatment cross over study.

<i>Design</i>	<i>Sequence</i>	<i>Period</i>			
		1	2	3	4
1	1	A	A		
	2	B	B		
	3	A	B		
	4	B	A		
2	1	A	B	B	
	2	B	A	A	

3	1	A	A	B	B
	2	B	B	A	A
	3	A	B	B	A
	4	B	A	A	B
4	1	A	B	B	A
	2	B	A	A	B

Any of the designs above will provide an optimal outcome that the random error will be minimized and the carry-over effect can be separated from the treatment and period interaction with adequate degree of freedoms in the model.

We will illustrate the statistical model and analyses using the Design 1, which is also called Balaam's design. Balaam's design requires  $t^2$  experiment units. The treatment effect under the Balaam's design can also be adjusted for carry-over effect if the carry-over is statistically significant. Denote the estimated means for each period and sequence as follows:

<u>Sequence</u>	<u>Period</u>		$A \bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.23}$
	1	2			
1	$A \bar{Y}_{.11}$				
2	$B \bar{Y}_{.12}$	B			
3	$A \bar{Y}_{.13}$	B			
4	$B \bar{Y}_{.24}$	$A \bar{Y}_{.24}$			

The difference in the estimated treatment mean adjusted for carry-over becomes:

$$\begin{aligned}
 &= \frac{1}{2} \{ (\bar{Y}_{.23} - \bar{Y}_{.13}) - (\bar{Y}_{.24} - \bar{Y}_{.14}) - (\bar{Y}_{.22} - \bar{Y}_{.12}) + (\bar{Y}_{.21} - \bar{Y}_{.11}) \} \\
 &= \frac{1}{2} ( (B-A)_3 - (A-B)_4 - (B-B)_2 + (A-A)_1 ) \\
 &= \frac{1}{2} \{ (B_3 + B_4) - (A_3 + A_4) \}
 \end{aligned}$$

Note that if the carry-over is not statistically significant (negligible) the terms  $(B-B)_2 + (A-A)_1$  will be close to 0. However, if the carry-over effect is significant the estimated treatment difference will be accounted for with the term  $(B-B)_2 + (A-A)_1$ .

For example, a biotechnology company is testing a compound for treating Parkinson's disease. The efficacy endpoint for the study is the quality of life. Since QOL is a subjective measurement and varies from person to person, a cross-over design will be a better choice of design for the study because the within subject variability can be accounted for and the random error can be reduced. The example data and the SAS® codes for the analyses are displayed below:

```

data balaam;
input seq $ Sub baseline P1 P2;
cards;
AA 1 14 12.5 14

```

```

AA 2 27 24.25 21.5
....
BB 5 21 21 22.51
BB 6 11 12.5 15
....
AB 9 9 8.75 9.75
AB 10 12 10.5 11.75
....
BA 13 23 22 21
BA 14 15 15 17
...
:
run;

```

```

proc glm;
class seq sub period trt;
model QOLI=seq sub(seq) period trt trt*period/*carry-over effect*/;
TEST H = SEQ E = SUB(SEQ) / HTYPE=1 ETYPE=1;
LSMEANS TRT trt*period / PDIF CL E;
run;

```

The partial output for the analyses are displayed below:

Tests of Hypotheses Using the Type I MS for Sub(seq) as an Error Term					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
seq	3	175.3097750	58.4365917	1.20	0.3510

Source	DF	Type III SS	Mean Square	F Value	Pr > F
seq	3	179.0085250	59.6695083	50.30	<.0001
Sub(seq)	12	583.6647250	48.6387271	41.00	<.0001
period	1	3.2896125	3.2896125	2.77	0.1198
trt	1	0.8789062	0.8789062	0.74	0.4050
period*trt	1	14.0812562	14.0812562	11.87	0.0043

The period and treatment interaction term is significant, which indicates the treatment responses differed from period to period. One can estimate the treatment difference based on the least squared mean (LSM) at each period:



period	trt	QOL1 LSMEAN	LSMEAN Number
P1	A	18.6256250	1
P1	B	17.2181250	2
P2	A	17.3906250	3
P2	B	19.7356250	4

Least Squares Means for Effect period*trt				
i	j	Difference Between Means	95% Confidence Limits for LS Mean(i)-LS Mean(j)	
1	2	1.407500	-0.256314	3.071314
1	3	1.235000	-0.205905	2.675905
1	4	-1.110000	-2.550905	0.330905
2	3	-0.172500	-1.613405	1.268405
2	4	-2.517500	-3.958405	-1.076595
3	4	-2.345000	-4.008814	-0.681186

For a 2\*2 higher-order design, there is no clear recommendation as to which design should be used. It depends on the features of the study. For example, if the length of treatment is long then one may choose to have more sequences, instead of more periods, since it will prolong the study time. Once the study design is chosen, the statistical analyses for 2\*2 cross-over higher-order designs are similar to those described above.

### Conclusions

A cross-over design is a more desirable design when the within subject variability is high. With careful planning and execution, a cross-over study design is more efficient. It allows each subject to serve as her/his own control and assesses the difference in each period/treatment that is due to different external causes, instead of internal sources (subjects themselves). The statistical models and analyses are similar among most of the designs, with some designs allowing carry-over effect to be accounted for. Therefore, a cross-over study design is recommended for studies where within subject variability is of concern.