Cross-Over Experiments

Cross-over experiments are a special class of repeated measures experiments. Up to now the repeated measures experiments that we have been investigating, are where the experimental unit (patient, animal or field plot) is measured repeatedly after treatment and these measurements within patient are assumed to have some type of correlation. Only one treatment is applied to each of the experimental units. With a cross-over experiment, each experimental unit has more than one treatment applied throughout the trial and the treatment usually differs from one measurement period to another and these treatments are given in a randomly assigned sequence. Using a cross-over experiment allows for an increase in precision when less variability is expected within subjects than between subjects. One feature of cross-over designs with more than two periods, is the ability to measure any possible carryover effects. Carryover effects are when the results in subsequent treatments are influenced by treatments given in previous periods or may also be caused by a "learning effect" or "fatigue effect". Researchers may try to get around a possible carryover effect by including a "washout" period between treatment periods. In other trials the carryover effect itself may be of interest.

Latin-Square Designs

Latin square designs are used in cross-over studies in order to separate out the treatment and carryover effects. "Digram-balanced" squares are better than other Latin squares for this. "Digram-balanced" is where each treatment is preceded or followed equally often by each other treatment. For example, for the following pair of Latin squares we have 3 treatments.

Square 1

	Period 1	Period 2	Period 3
Animal 1	A	B	C
Animal 2	B	C	A
Animal 3	C	A	B

Square 2

	Period 1	Period 2	Period 3
Animal 1	A	C	B
Animal 2	B	A	C
Animal 3	C	B	A

We count the number of times each treatment follows every other treatment.

B follow A	2 (rows 1 and 3, S1)
B follow C	2 (rows 1 and 3, S2)
A follow B	2 (rows 2 and 3, S2)
A follow C	2 (rows 2 and 3, S1)
C follow A	2 (rows 1 and 2, S2)
C follow B	2 (rows 1 and 2, S1)

Notice, that we would not be able to obtain a "digram-balanced" Latin square design if we only used one Latin square. The general rule is that if there is an even number of treatments, then it is possible to obtain a "digram-balanced" design with only one square. If there is an odd number of treatments, then two squares are needed to obtain balance.

The basic model for a cross-over design would be

$$y_{ijk} = \mu + \gamma_i + \xi_{i(k)} + \pi_j + \tau_{t(ij)} + \delta_j \lambda_{r(ij)} + e_{ijk}$$

μ = mean
γ_i = sequence effect for sequence i
$\xi_{i(k)}$ = random subject effect nested within sequence
π_j = period effect for period j
$\tau_{t(ij)}$= direct treatment effect for treatment t
$\lambda_{r(ij)}$ = the first-order "carryover" effect due to treatment r, where r is the treatment applied the previous period
δ_j = an indicator variable: 0 if it is the first period
1 if the period is greater than 1

Note that for the treatment direct effect, the subscript t(ij) indicates that treatment t is a function of the sequence i and period j and for the carryover effect, the subscript r(ij) indicates that carryover r is a function of the sequence and the period. The indicator variable allows us to distinguish between the first period, where there is no carryover effect and the subsequent periods. First-order "carryover" means that we are only looking at the carryover effect from the immediately previous treatment. Second- and third-order carryover effects can also be evaluated. Second-order carryover is when the effect of a treatment carries over two periods, while third-order carryover is when the treatment effect carries over three periods. The occurrence of second- and third-order carryover is very rare in reality, so almost all designs and analyses concentrate only on first-order carryover.

Test of the Separability of Treatment and Carryover

When there is carryover in a balanced Latin square design, there is no longer orthogonality, so that the order that the effects appear in the model impact the magnitude of the effect. Different Latin Square designs vary in their ability for separating out the treatment and carryover effects. A design that is poor at separating out the effects may cause the treatment effect to be declared significant, when it may be due to a positive carryover effect, especially if the Latin Square is not "Digram-balanced". A measure of separability can be calculated for each type of design, based on the observed frequencies of carryover and the expected frequencies from an independent model.

Following is the "digram-balanced" Latin Square for a 4 period crossover:

	Period 1	Period 2	Period 3	Period 4
Subject 1	A	B	C	D
Subject 2	B	D	A	C
Subject 3	C	A	D	B
Subject 4	D	C	B	A

The first thing we need to do is build a contingency table relating treatment and carryover incidence:

	Carryover 0	Carryover A	Carryover B	Carryover C	Carryover D	
Treat A	1	0	1	1	1	4
Treat B	1	1	0	1	1	4
Treat C	1	1	1	0	1	4
Treat D	1	1	1	1	0	4
	4	3	3	3	3	16

The expected frequencies under the null hypothesis under the independence model, which means that the probability of an observation falling into a particular column is not a function of the row where the observation occurs, are presented below

	Carryover 0	Carryover A	Carryover B	Carryover C	Carryover D	
Treat A	1	0.75	0.75	0.75	0.75	4
Treat B	1	0.75	0.75	0.75	0.75	4
Treat C	1	0.75	0.75	0.75	0.75	4
Treat D	1	0.75	0.75	0.75	0.75	4
	4	3	3	3	3	16

We can then calculate the Pearson chi-square

$$\chi^2 = \sum\sum[(O_{ij} - E_{ij})^2 / E_{ij}] = 4$$

The chi-square value is then standardized with the following formula called Cramer's V,

$$V = \left[\frac{\chi^2 / N}{\min(r-1, c-1)}\right]^{1/2}$$

where N is the number of incidences (16), r is number of rows (4) and c is number of columns (5). So

$$V = \left[\frac{4/16}{3}\right]^{1/2} = 0.289$$

The "efficiency" of separation of treatment and carryover (S) is calculated from V as follows:

S=100(1-V)=100(1-.289)=71.1.

If we added a fifth period, where the subjects receive the same treatment as in the fourth period as follows, with the corresponding contingency table, the efficiency of the design becomes 100%.

	Period 1	Period 2	Period 3	Period 4	Period 5
Subject 1	A	B	C	D	D
Subject 2	B	D	A	C	C
Subject 3	C	A	D	B	B
Subject 4	D	C	B	A	A

Contingency table:

	Carryover 0	Carryover A	Carryover B	Carryover C	Carryover D	
Treat A	1	1	1	1	1	5
Treat B	1	1	1	1	1	5
Treat C	1	1	1	1	1	5
Treat D	1	1	1	1	1	5
	4	4	4	4	4	20

We will look at this measure further in the next section.

Two-Treatment, Two-Period, Two-Sequence Design

This design is one of the most widely used designs in the pharmaceutical industry, especially in the area of bioequivalence. It is also known as the "Grizzle" design or the "Hills-Armitage" design for the authors who were the first to describe the statistical analysis of this design. The following table presents the design.

Period 1	**Period 2**
Treatment A	Treatment B
Treatment B	Treatment A

Subjects or animals are randomly allocated to one of the two sequences (AB or BA). This design is the most powerful statistically when the numbers of animals in each sequence are equal, so this design could also be considered a replicated two treatment Latin square.

The model for this design would be

$$y_{ijh} = \mu + \gamma_i + \xi_{i(k)} + \pi_j + \tau_t + e_{ijk}$$

μ = mean
γ_i = sequence effect
$\xi_{i(k)}$ = random subject effect nested within sequence
π_j = period effect
τ_j = treatment effect

Note that there are only four cell means for this design, which means that there are only four parameters that can be estimated. One of these is the overall mean. The others would be period effect, treatment effect and sequence effect. Therefore, assumptions of no carryover effect and no treatment-by-period interaction must be made with this design because these two terms are aliased with the sequence effect. If the sequence effect is statistically significant, the usual practice in the pharmaceutical industry is to look at the first period results only.

It is not unusual for subjects not to complete both periods of the crossover study. If we were to treat the subject effect as fixed, then none of the data for subjects with missing data would be used. However, treating subject as random, the data from the first period would still be used in the analysis. The following are the within-subject and between-subject comparison variances for the unbalanced data, where N=number of subjects, p= proportion of subjects who provide data for only one period.

Close to the beginning of the course, we saw that the variance of individual observations is the sum of the variance components

Within-subject variance of the estimate of treatment differences

If there are missing data and subjects are treated as fixed, then this within-subject variance is used.

$var_W(A\text{-}B) = 2\sigma_r^2/N(1-p)$ where there are N(1-p) subjects who have complete data

and σ_r^2 is the residual variance

Between-subject variance of the estimate of treatment differences

$var_B(A\text{-}B) = 2(\sigma_r^2 + \sigma_p^2)/(Np/2) = 4\sigma_r^2(1+\gamma)/Np$ where $\gamma = \sigma_p^2/\sigma_r^2$

and σ_p^2 is the between-subject variance component.

Pooled comparisons or weighted average of the treatment effect

When subject is treated as random and there are missing data then the following equation for calculating the pooled variance where we use a weighted average of the within and between-subject estimates is used.

$var_P\ (A-B)= 4\sigma_r^2(1+\gamma)/(Np+2N(1-p)(1+\gamma)$

The textbook looks at the relative efficiency of using subjects as fixed versus subjects as random with varying proportions of missing data and varying ratios of σ_p^2/σ_r^2. When the ratio of σ_p^2/σ_r^2 was small (0.2), then using random subjects was more beneficial. However when σ_p^2/σ_r^2 was large (3.0), then there was not much benefit using random subjects over fixed subjects unless the proportion of missing data was large.

Example of the Two-Treatment, Two-Period, Two-Sequence Design

This design is also called the AB/BA crossover design. We will go over the example of analyses of this type of design that is presented in the book.

The data presented is from a cross-over trial comparing two diuretics in patients with heart failure. There were 94 patients who entered into the study and only two of these did not complete both periods. The authors removed approximately 20% of the second period data in order to compare four different models. Following is a printout of the data for the first 6 patients.

Obs	patient	treatment	period	edema0	dbp0	edema	dbp
1	1	B	1	45	60	45	55
2	1	A	2	45	60	45	60
3	2	A	1	51	50	48	60
4	2	B	2	51	50	48	65
5	3	A	1	53	70	50	70
6	3	B	2	53	70	52	80
7	4	B	1	49	68	47	60
8	4	A	2	49	68	47	60
9	5	A	1	46	65	45	60
10	6	A	1	61	95	60	95
11	6	B	2	61	95	59	97

The amount of edema and diastolic blood pressure were measured both before the start of the study (edema0, dbp0) and at the end of each treatment period (edema, dbp). Note that patient 5 completed only the first period of the study.

We will analyze these data with four different models presented in the following table.

Model	Fixed Effects	Random Effects
1	Treatment, period, patient	-
2	Treatment, period	patient
3	Treatment, period, patient, baseline	-
4	Treatment, period, baseline	patient

The following SAS program reads in the data and analyzes the edema data using the four different models. The same models are used for the blood pressure data.

```
options ps=65 ls=65;
filename hrtfl 'i:\kathy mixed model\hrtfail.dat';
data dbp;
infile hrtfl;
input patient treatment $ period edema0 dbp0 edema dbp;

proc mixed noclprint;
class treatment period patient;
title 'Model 1: Fixed effect Analysis without baseline';
model edema=treatment period patient;
lsmeans treatment/diff pdiff;

proc mixed noclprint;
class treatment period patient;
title 'Model 2: Random effect Analysis without baseline';
model edema=treatment period/ddfm=kr;
random patient;
lsmeans treatment/diff pdiff;

proc mixed noclprint;
class treatment period patient;
title 'Model 3: Fixed effect Analysis with baseline';
model edema=treatment period patient edema0;
lsmeans treatment/diff pdiff;

proc mixed noclprint;
class treatment period patient;
title 'Model 4: Random effect Analysis with baseline';
model edema=treatment period edema0/ddfm=kr;
random patient;
lsmeans treatment/diff pdiff;
```

The following SAS output is from Model 1 for edema.

```
                    Model Information
Data Set                      WORK.DBP
Dependent Variable            edema
Covariance Structure          Diagonal
Estimation Method             REML
Residual Variance Method      Profile
Fixed Effects SE Method       Model-Based
Degrees of Freedom Method     Residual

                       Dimensions
          Covariance Parameters            1
          Columns in X                    99
          Columns in Z                     0
          Subjects                         1
          Max Obs Per Subject            170
```

Number of Observations

Number of Observations Read	170
Number of Observations Used	168
Number of Observations Not Used	2

Covariance Parameter Estimates

Cov Parm	Estimate
Residual	0.5297

Fit Statistics

-2 Res Log Likelihood	217.1
AIC (smaller is better)	219.1
AICC (smaller is better)	219.2
BIC (smaller is better)	221.4

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treatment	1	72	6.45	0.0132
period	1	72	4.31	0.0414
patient	93	72	218.37	<.0001

Least Squares Means

Effect	treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
treatment	A	55.2267	0.08487	72	650.73	<.0001
treatment	B	54.9226	0.08438	72	650.88	<.0001

Model 1: Fixed effect Analysis without baseline 25

Differences of Least Squares Means

Effect	treatment	_treatment	Estimate	Standard Error	DF
treatment	A	B	0.3041	0.1197	72

Differences of Least Squares Means

Effect	treatment	_treatment	t Value	Pr > |t|
treatment	A	B	2.54	0.0132

The results from all four analyses are summarized in the following table

	Treatment effect: A-B(SE)		**Variance components (SE)**			
Model			**Edema**		**Diastolic BP**	
	Edema	**Diastolic BP**	**Patient**	**Residual**	**Patient**	**Residual**
1	0.3041(0.1197)	0.7427(0.7789)	-	0.5297	-	22.4312
2	0.3009(0.1197)	0.8615(0.7692)	66.8251	0.5299	76.9105	22.4929
3	0.3041(0.1197)	0.7427(0.7789)		0.5297		22.4312
4	0.3085(0.1184)	0.9581(0.7518)	3.7632	0.5260	25.3284	22.1425

Notice that the results from Models 1 and Models 3 are exactly the same. This is a well known result that with a fixed effect analysis, including a single baseline has no effect. This can be seen further looking at the SAS output from Model 3. Note the 0 degrees of freedom associated with the baseline effect. For each "level" of patient, there is only one level of the baseline effect associated.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treatment	1	72	6.45	0.0132
period	1	72	4.31	0.0414
patient	92	72	14.21	<.0001
edema0	0	.	.	.

For Model 4, where patient is a random effect, we get the following results for edema. Note that the baseline edema was a significant source of variation.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treatment	1	75.3	6.79	0.0110
period	1	74.9	4.02	0.0487
edema0	1	94.1	1433.62	<.0001

We can discuss a little about the relative efficiencies of the fixed versus random patient effect models. For edema, the between-patient variation was so large relative to the residual, that there was no gain treating the patient random versus fixed. This is further verified by the standard errors of the treatment differences between Models 1 and 2 being the same. Even though adding the baseline edema covariate reduced the between-patient variation substantially, it was still seven times larger than the residual error, again indicating no gain treating the patient effect as random. For diastolic blood pressure, the between-patient variation was still larger than the residual variation, although not as extreme as for edema and this can be seen by the somewhat smaller standard errors for the models with random patient compared to the models with fixed patient.