L11: Crossover designs for repeated measures data

BIOS6643 Longitudinal

EJC

Department of Biostatistics & Informatics, CU Anschutz

1. Introduction and definitions
2. Carry-over effects
3. Examples
4. Summary

Associated reading

- ‘Crossover designs for repeated measures data’ section (5.4) in **Course notes**

Crossover designs for repeated measures data

- In some cases a researcher may want to have each subject try multiple treatments in an experiment, rather than just one.
  - In the simplest case, there are 2 treatments, which can be assigned to each subject in a 2 period, 2 treatment (2×2) crossover design.
- For the 2×2 design, subjects are usually randomly assigned an order of treatments, AB or BA, in equal amount.
  - This helps to eliminate confounders associated with time.
- If there are 3 treatments, then one may set up a 3 period, 3 treatment crossover design.
- Crossover designs are often used in clinical trials when the cost of tracking subjects longitudinally for extended time does not impose major difficulties.

Carry-over effects

Carry-over effects: One limitation of crossover designs is that receiving one treatment first may have an influence on subjects' responses in the subsequent period in which they receive the other treatment.

- If this carry-over effect differs between treatment sequences, then estimates of effect of interest may be biased.
- The difficulty with the 2×2 design is that carry-over effect estimates are aliased with other effects (i.e., they are completely confounded with each other).
  - Specifically, the sequence, carry-over and period*treatment effects are aliased.
- If sequence and period*treatment effects are assumed to not exist, then we can test for carry-over effects by including the sequence term in the model. But the validity of the test relies on that assumption…
- In more complex models, it may be easier to estimate carryover effects by examining interactions. Including a term in the model for treatment used in the previous period may help in estimating (differential) carryover effects.

- For any crossover design, including a washout period of suitable length between treatment periods may help to eliminate carryover effects that a treatment might have.
  - Most researchers do include some washout period in their crossover experiment, however one of the issues that arises is planning in advance how long this should be since it is often uncertain how long it will take to 'wash out' the treatment.
- If some carryover effects are expected for a given study or experiment, then the researcher may also consider using alternative designs. Here, we focus on crossover experiments with repeated measures within periods.
- For more examples and details about modeling data from crossover designs, see **Littell et al, SAS System for Mixed Models**, and **Jones and Kenward, Design and Analysis of Cross-Over Trials** (in particular, see Chapter 5).

Example: Cherry juice to improve muscle damage

Consider a crossover experiment that was performed and reported in **Connolly et al. (2006)**, entitled *Efficacy of a tart cherry juice blend in preventing the symptoms of muscle damage (British Journal of Sports Medicine, 40: 679-683)*.

- In the experiment, subjects were randomized to receive cherry juice twice a day or placebo drink for 8 consecutive days.
- At day 4, subjects performed 'eccentric elbow flexion contractions'.
- Measures of strength and pain after the challenge (relative to baseline) were then taken on subjects on the last 4 days of the period, after the challenge.
- Subjects then repeated the experiment with the treatment they did not have in Period 1 (crossover), using the opposite arm.
- Mean strength was greater and pain was less when subjects had the cherry drink, relative to placebo.
  - Strength loss relative to BL was 22% for placebo but only 4% for cherry juice.
- This is considered a 2-period, 2-treatment crossover design, with repeated measures.

Hypothetical example

- In the spirit of this experiment, consider a hypothetical data set involving muscle soreness measurements on 4 successive days after an exercise challenge.
- This soreness score ranges from 0 to 10 but is typically in a range of 1 to 4.
- These scores are adjusted for baseline soreness before the experiment (e.g., if a subject has a soreness score of 2 coming into the study and a score of 6 one day after the challenge, then their soreness score on that day would be 4).
- This was designed like the reported experiment (2×2 crossover, 4 repeated measures within each period).

Below is a description of the predictors in the model and what they can be used to test:

- **Period:** 1 or 2; test accounts for differences between first and second time periods.
- **Treatment:** placebo vs. cherry drink; test is for main effect of treatment (comparing treatment means).
- **Time:** the 4 days that measures were taken following the exercise challenge; test is for main effect of time (comparing means for 4 days following the exercise challenge); time modeled as a categorical (class) variable.
- **Period** $\times$ **time:** Will test for differences between time patterns between the two periods. Can be thought of as the general time variable.
- **Treatment** $\times$ **time:** Will test whether changes over time (with a period) differ between the placebo and cherry drink.
  - If treatment $\times$ time is significant, then comparisons can be made between treatments for individual days (with multiple comparison adjustments, if desired).

Here is the SAS code for the analysis.

- **Sequence:** Compares AB versus BA treatments. Since there are only 2 treatments, this sequence effect is aliased with carry-over effects.
  - We can use this term to test for carry-over effects assuming that there are no true treatment $\times$ period or (other) sequence effects.

```
data cross; format trt $9.;
input id pd trt $ time seq y @@; datalines;
1 1 Control 1 1 1.7 1 1 Control 2 1 2.9 1 1 Control 3 1 3.4
1 1 Control 4 1 2.8 1 2 Treatment 1 1 1.5 1 2 Treatment 2 1 3.0
1 2 Treatment 3 1 3.1 1 2 Treatment 4 1 1.9 2 1 Control 1 1 1.5
. . .
8 1 Treatment 2 2 3.4 8 1 Treatment 3 2 3.0 8 1 Treatment 4 2 1.7
8 2 Control 1 2 2.0 8 2 Control 2 2 4.2 8 2 Control 3 2 3.3
8 2 Control 4 2 2.8
; run;
proc sort data=cross_mv; by time;
proc gplot data=cross;
  plot y*time=trt / vaxis=axis1 haxis=axis2;
  axis1 label=(h=2 angle=90 'muscle soreness score')
     order=1 to 4 by 1 value=(h=2);
  axis2 label=(h=2 'time (days after treatment)') value=(h=2);
  symbol1  i=std1mtj l=1 c=blue   v=none w=2 mode=include w=2;
  symbol2  i=std1mtj l=1 c=red    v=none w=2 mode=include w=2; run;
proc mixed data=cross order=data;
 class id pd trt time seq;
 model y = pd trt time pd*time trt*time seq / dfm=kr solution;
 random id;  repeated time / type=ar(1) subject=id*pd;
 lsmeans trt*time; run;
```

Abbreviated output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
id		0.2434
AR(1)	id*pd	0.7168
Residual		0.5308

Fit Statistics

-2 Res Log Likelihood	113.4
AIC (smaller is better)	119.4
AICC (smaller is better)	120.0
BIC (smaller is better)	119.7

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
pd	1	5.7	0.74	0.4247
trt	1	5.7	4.81	0.0733
time	3	36.1	19.96	<.0001
pd*time	3	36.1	0.15	0.9261
trt*time	3	36.1	2.72	0.0584
seq	1	6	0.18	0.6836

Least Squares Means

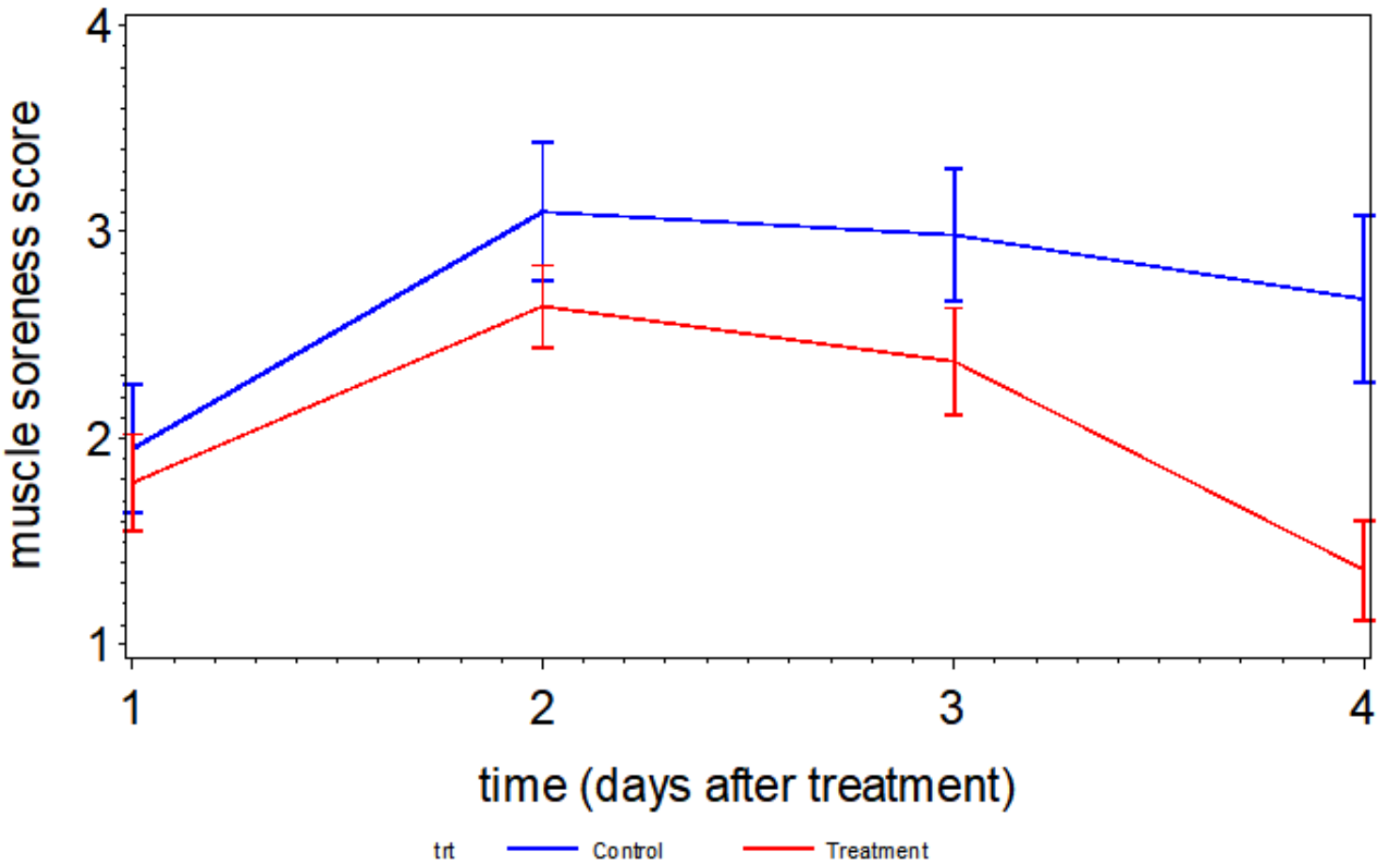
Effect	trt	time	Estimate	Standard Error	DF	t Value	Pr > |t|
trt*time	Control	1	1.9500	0.3111	16.4	6.27	<.0001
trt*time	Control	2	3.1000	0.3111	16.4	9.97	<.0001
trt*time	Control	3	2.9875	0.3111	16.4	9.60	<.0001
trt*time	Control	4	2.6750	0.3111	16.4	8.60	<.0001
trt*time	Treatment	1	1.7875	0.3111	16.4	5.75	<.0001
trt*time	Treatment	2	2.6375	0.3111	16.4	8.48	<.0001
trt*time	Treatment	3	2.3750	0.3111	16.4	7.63	<.0001
trt*time	Treatment	4	1.3625	0.3111	16.4	4.38	0.0004

Differences of Least Squares Means

Effect	trt	time	_trt	_time	Estimate	Standard Error	DF	t Value	Pr > |t|
trt*time	Control	1	Treatment	1	0.1625	0.3643	10.8	0.45	0.6643
trt*time	Control	2	Treatment	2	0.4625	0.3643	10.8	1.27	0.2308
trt*time	Control	3	Treatment	3	0.6125	0.3643	10.8	1.68	0.1212
trt*time	Control	4	Treatment	4	1.3125	0.3643	10.8	3.60	0.0042

How do we interpret the results?

Graph of sample means and SD error bars.



See course notes for other examples: sleep study, similar crossover design with washout.

Summary