Cross-over trials

Martin Bland

Professor of Health Statistics

University of York

http://martinbland.co.uk

Cross-over trials

Use the participant as their own control.

Each participant gets more than one treatment.

Also known as change-over trials.

Cross-over trials

Example: a two treatment cross-over trial:
pronethalol vs placebo for the treatment of angina.

Patients received placebo for two periods of two weeks and pronethalol for two periods of two weeks, in random order.

Completed diaries of attacks of angina.

Pritchard BNC, Dickinson CJ, Alleyne GAO, Hurst P, Hill ID, Rosenheim ML, Laurence DR. Report of a clinical trial from Medical Unit and MRC Statistical Unit, University College Hospital Medical School, London} *BMJ* 1963; **2**: 1226-7.

Cross-over trials

Example: a two treatment cross-over trial: pronethalol vs placebo for the treatment of angina.

Attacks of angina recorded over four weeks:

Placebo:	2	3	7	8	14	17
	23	34	60	79	71	323
Pronethalol:	0	0	1	2	7	15
	16	25	29	41	65	348

Mann Whitney U test: P = 0.4.

But this ignores the data structure.

These observations should be paired.

Results of a trial of pronethalol for the treatment of angina pectoris (Pritchard *et al*., 1963)

Patient	Placebo	Pronethalol	Placebo-Pronethalol
1	71	29	42
2	323	348	−25
3	8	1	7
4	14	7	7
5	23	16	7
6	34	25	9
7	79	65	14
8	60	41	19
9	2	0	2
10	3	0	3
11	17	15	2
12	7	2	5

Paired analysis (sign test):

P = 0.006.

Two sample analysis, ignoring pairing, (Mann Whitney U test):

P = 0.4.

Advantages of cross-over designs

- ❖ each participant acts as their own control
- ❖ removes variability between participants,
- ❖ fewer subjects needed.

Disadvantages of cross-over designs

- ❖ short term treatment,
- ❖ no follow-up.

Cross-over trials are suitable for:

- chronic diseases (angina, asthma, arthritis),
- symptomatic treatment,
- quick, quantitative outcome (attack frequency, lung function, pain scores),
- early stage in treatment development.

Cross-over trials are *not* suitable for:

- acute conditions (myocardial infarction, pneumonia),
- treatment to cure (clot-busters, antibiotics),
- slow or qualitative outcomes (time to recurrence, death),
- later stages in development (side effects of long term treatment).

Estimation and significance tests

Trialists are encouraged to present results as estimates with confidence intervals rather than use significance tests, i.e. give P values.

Cross-over trials are typically small, so t methods are required for this.

In the pronethalol example, only P values were given.

Does this matter?

Not so much as in a larger trial, as cross-over trials are usually at an early stage in treatment development.

P values are often more important than estimates.

Analysis for an AB/BA cross-over trial

A trial where there are two treatments, each given once, in random order, is called a simple two period two treatment cross-over trial or AB/BA design.

Analysis will be illustrated using a cross-over trial of a homeopathic preparation intended to reduce mental fatigue.

This was a trial in healthy volunteers. On different occasions, paid student and staff volunteers received either the homeopathic preparation or a placebo.

They underwent a psychological test to measure their resistance to mental fatigue.

Analysis for an AB/BA cross-over trial

There were two treatments labelled A and B, one is a homeopathic dose of potassium phosphate and the other a control.

This is a triple blind trial, in that I do not know which is which.

Subjects took A or B, in random order, on different occasions, and carried out a test where accuracy was the outcome measurement.

There were 86 subjects, 43 for each order.

Analysis for an AB/BA cross-over trial

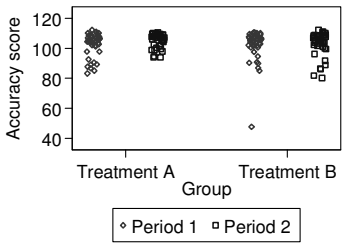
A first				B first			
acc1	acc2	acc1	acc2	acc1	acc2	acc1	acc2
84	108	106	104	50	101	105	107
85	108	106	107	86	99	105	108
88	82	106	107	89	106	106	96
88	89	106	107	91	102	106	108
88	107	106	108	92	100	106	108
91	104	106	108	93	106	106	108
92	107	106	108	93	.	106	.
93	89	107	100	97	106	107	105
98	89	107	104	99	106	107	106
98	107	107	105	101	103	107	106
101	80	107	107	102	95	107	106
101	90	107	107	102	99	107	107
101	99	107	108	102	101	107	107
103	98	107	108	102	101	107	108
103	106	108	94	102	106	108	107
103	107	108	104	102	108	108	107
104	107	108	106	102	108	108	108
104	108	108	108	103	105	108	108
105	106	108	108	103	108	108	108
105	107	108	108	104	90	108	108
105	108	108	108	105	104	108	108
106	100			105	107		

Sorted by first observation.

Clear ceiling effect.

Two students did not come back for the second measurement.

Analysis for an AB/BA cross-over trial



Ceiling effect and negative skewness.

Period 2 may have greater accuracy than period 1.

Analysis for an AB/BA cross-over trial

We can do a simple test of the treatment effect, by estimating the mean difference, A minus B.

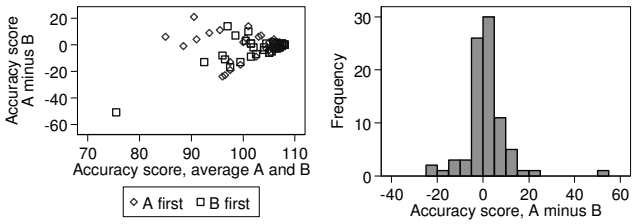
```
. ttest diffamb=0
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 diffamb |      84    1.035714      1.0045    9.206397   -.9621963    3.033625
------------------------------------------------------------------------------
Degrees of freedom: 83

                     Ho: mean(diffamb) = 0

     Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
       t =   1.0311                 t =   1.0311                 t =   1.0311
   P < t =   0.8472             P > |t| =   0.3055             P > t =   0.1528
```

Estimated treatment effect = 1.0 (95% CI –1.0 to 3.0, P=0.3).

Are the assumptions of this analysis met?

Analysis for an AB/BA cross-over trial

Plot of difference against average of the two scores:

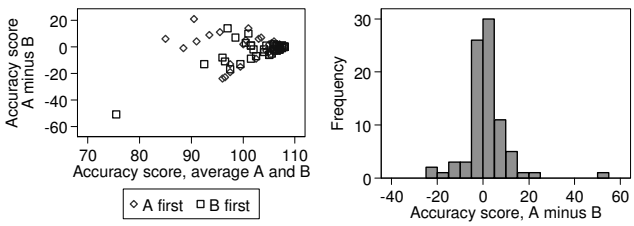


Differences are related to the magnitude of the measurement, cannot assume SD is well estimated.

Assumptions for t method not met.

Analysis for an AB/BA cross-over trial

Plot of difference against average of the two scores:



Differences have symmetrical distribution.

Could use Wilcoxon matched-pairs (signed rank) test.

Analysis for an AB/BA cross-over trial

Non-parametric analysis, Wilcoxon paired test :

```
. signrank diffamb=0
Wilcoxon signed-rank test
        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |       36      1991.5      1739.5
    negative |       35      1487.5      1739.5
        zero |       13          91          91
-------------+---------------------------------
         all |       84        3570        3570
unadjusted variance    50277.50
adjustment for ties      -180.00
adjustment for zeros     -204.75
                     ----------
adjusted variance      49892.75
Ho: diffamb = 0
           z =   1.128
    Prob > |z| =   0.2592
```

P = 0.3, as before.

Conclusion: no evidence for a treatment effect.

Analysis for an AB/BA cross-over trial

Non-parametric analysis, Wilcoxon paired test:

```
. signrank diffamb=0
Wilcoxon signed-rank test
        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |       36      1991.5      1739.5
    negative |       35      1487.5      1739.5
        zero |       13          91          91
-------------+---------------------------------
         all |       84        3570        3570
unadjusted variance    50277.50
adjustment for ties      -180.00
adjustment for zeros     -204.75
                     ----------
adjusted variance      49892.75
Ho: diffamb = 0
           z =   1.128
    Prob > |z| =   0.2592
```

Conclusion: no evidence for a treatment effect.

But we can do better.

The difference between periods has gone into the error.

Analysis for an AB/BA cross-over trial

Using the period effect, step by step method using t tests (Armitage and Hills, 1982).

To see how the analysis works, we will use the following notation:

A1 = the mean for A in the first period

A2 = the mean for A in the second period

B1 = the mean for B in the first period

B2 = the mean for B in the second period

Armitage P and Hills M. (1982) The two period cross-over trial. *The Statistician* **31**, 119-131.

Analysis for an AB/BA cross-over trial

First we ask whether there is evidence for a period effect, i.e. are scores in the first period the same as in the second?

For example, there might be a learning effect, with accuracy increasing with repetition of the test.

If there is no period effect, we expect the differences between the treatment to be the same in the two periods.

Analysis for an AB/BA cross-over trial

The period effect, first period mean minus second period mean, will be estimated by

(A1 – A2 + B1 – B2)/2.

We can rearrange this as

(A1 – B2 – A2 + B1)/2 = (A1 – B2)/2 – (A2 – B1)/2

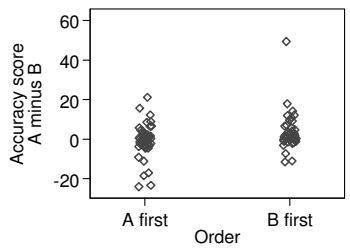
(A1 – B2) is the mean treatment difference for the group with A first, (A2 – B1) is the mean treatment difference for the group with A first.

We can test the null hypothesis that the difference between these two mean differences is zero.

Compare difference A minus B between orders.

Analysis for an AB/BA cross-over trial

Compare difference A minus B between orders.



Analysis for an AB/BA cross-over trial

Compare difference A minus B between orders.

```
. ttest diffamb, by(order)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 A first |      43   -.8604651    1.295952    8.498127   -3.475803    1.754872
 B first |      41     3.02439    1.498978    9.598145   -.0051582    6.053939
---------+--------------------------------------------------------------------
combined |      84    1.035714      1.0045    9.206397   -.9621963    3.033625
---------+--------------------------------------------------------------------
    diff |           -3.884855    1.975746               -7.815243    .0455321
------------------------------------------------------------------------------
Degrees of freedom: 82
                  Ho: mean(A first) - mean(B first) = diff = 0
     Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
       t =  -1.9663                 t =  -1.9663                  t =  -1.9663
   P < t =   0.0263           P > |t| =   0.0527           P > t =   0.9737
```

Weak evidence of a period effect, P=0.05.

Analysis for an AB/BA cross-over trial

Compare difference A minus B between orders.

Non-parametric analysis using Mann-Whitney U test:

```
. ranksum diffamb, by(order)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
       order |      obs    rank sum    expected
-------------+---------------------------------
     A first |       43        1610      1827.5
     B first |       41        1960      1742.5
-------------+---------------------------------
    combined |       84        3570        3570
unadjusted variance    12487.92
adjustment for ties     -151.09
                       ----------
adjusted variance      12336.83
Ho: diffamb(order==A first) = diffamb(order==B first)
             z =  -1.958
    Prob > |z| =   0.0502
```

Weak evidence of a period effect, P=0.05.

Analysis for an AB/BA cross-over trial

We can allow for a possible period effect.

We estimate the average of the differences between A and B for each period: (A1 — B1)/2 + (A2 — B2)/2

(A1 — B1)/2 + (A2 — B2)/2 = (A1 — B2)/2 – (B1 – A2)/2

(A1 — B2) and (B1 – A2) are the differences between periods 1 and 2, for those starting with A and for those starting with B.

Test the difference between difference between periods 1 and 2 for the two orders.

This is called the CROS analysis.

Analysis for an AB/BA cross-over trial

Test the difference between difference between periods 1 and 2 for the two orders.

```
. ttest diff1m2, by(order)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |      Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 A first |       43   -.8604651    1.295952    8.498127   -3.475803    1.754872
 B first |       41    -3.02439    1.498978    9.598145   -6.053939    .0051582
---------+--------------------------------------------------------------------
combined |       84   -1.916667    .9887793    9.062312   -3.883309    .0499756
---------+--------------------------------------------------------------------
    diff |            2.163925    1.975746               -1.766462    6.094313
------------------------------------------------------------------------------
    diff = mean(A first) - mean(B first)                          t =   1.0952
Ho: diff = 0                     degrees of freedom =       82
    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.8617         Pr(|T| > |t|) = 0.2766          Pr(T > t) = 0.1383
```

No evidence for a treatment effect.

Estimate 2.163925/2 = 1.1 (95% CI –0.9 to 3.0, P=0.3).

Analysis for an AB/BA cross-over trial

Test the difference between difference between periods 1 and 2 for the two orders.

```
. ranksum diff1m2, by(order)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
       order |      obs    rank sum    expected
-------------+---------------------------------
     A first |       43        1949      1827.5
     B first |       41        1621      1742.5
-------------+---------------------------------
    combined |       84        3570        3570
unadjusted variance    12487.92
adjustment for ties     -100.77
                     ----------
adjusted variance      12387.15
Ho: diff1m2(order==A first) = diff1m2(order==B first)
             z =   1.092
    Prob > |z| =   0.2750
.
```

No evidence for a treatment effect, P=0.3.

Analysis for an AB/BA cross-over trial

Same analysis by analysis of variance:

```
. anova   score sub treat period

                          Number of obs =     170     R-squared     =  0.6490
                          Root MSE      = 6.40033     Adj R-squared =  0.2765

                  Source |  Partial SS    df       MS           F     Prob > F
             ------------+----------------------------------------------------
                   Model |  6210.08374    87  71.3802729       1.74     0.0059
                         |
                     sub |  5990.61699    85  70.4778469       1.72     0.0071
                   treat |  49.1391331     1  49.1391331       1.20     0.2766
                  period |  158.377228     1  158.377228       3.87     0.0527
                         |
                Residual |   3359.0692    82  40.9642585
             ------------+----------------------------------------------------
                   Total |  9569.15294   169  56.6222068
```

Analysis for an AB/BA cross-over trial

Is the treatment difference the same whatever order of treatments is given?

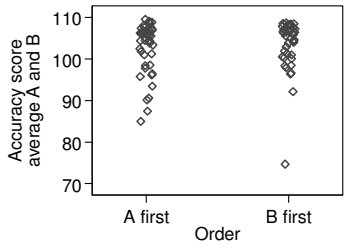
I.e., is there an interaction between period and treatment? Is there an order effect?

If not, the participant's average response should be the same whichever order treatments were given: is A1 + B2 = A2 + B1?

To test for period × treatment interaction, we compare the sum or the average of the scores between orders.

Analysis for an AB/BA cross-over trial

To test for period × treatment interaction, we compare the sum or the average of the scores between orders.



Analysis for an AB/BA cross-over trial

To test for period × treatment interaction, we compare the sum or the average of the scores between orders.

```
. ttest avland2, by(order)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 A first |      43     102.593    .9138191    5.992312    100.7489    104.4372
 B first |      41    103.2439    .9346191    5.984482     101.355    105.1328
---------+--------------------------------------------------------------------
combined |      84    102.9107    .6504313    5.961301     101.617    104.2044
---------+--------------------------------------------------------------------
    diff |           -.6508792    1.307167               -3.251251    1.949493
------------------------------------------------------------------------------
Degrees of freedom: 82
                 Ho: mean(A first) - mean(B first) = diff = 0
     Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
       t =  -0.4979                 t =  -0.4979                 t =  -0.4979
   P < t =   0.3099           P > |t| =   0.6199           P > t =   0.6901.
```

No evidence of an interaction, P=0.6.

Period × treatment interaction

In the mental fatigue trial, there could be an interaction because of the ceiling effect and practice.

One treatment could raise scores to the ceiling in the first period and all get near the ceiling in the second period.

Another possibility in cross-over trials is a carry-over effect, where the first treatment continues to have an effect in the second period.

Note that if the interaction is significant, there is a significant treatment effect.

Period × treatment interaction

Should we test it? Two views:

Grizzle (1965) says yes.

If significant, use period 1 data only.

Difference, A – B = 0.5 (95% CI –3.2 to 4.2, P=0.8).

Called the two-stage analysis.

Grizzle JE. (1965) The two-period change-over design and its use in clinical trials. *Biometrics*, **21**, 467-480.

Period × treatment interaction

Should we test it? Two views:

Grizzle (1965) says yes.

If significant, use period 1 data only.

Difference, A – B = 0.5 (95% CI –3.2 to 4.2, P=0.8).

Called the two-stage analysis.

Compare the full data estimate:

Difference, A – B = 1.1 (95% CI –0.9 to 3.0, P=0.3).

We lose power and precision.

Period × treatment interaction

Should we test it? Two views:

Senn (1989) says no.

> The average of the first and second periods is highly correlated with the first period.

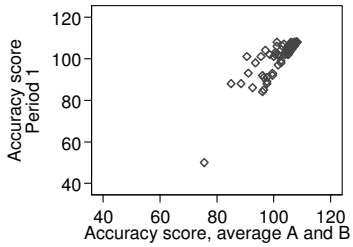
Senn S. *Cross-Over Trials in Clinical Research.* Chichester: Wiley, 1989.

Period × treatment interaction

Should we test it? Two views:

Senn (1989) says no.

> The average of the first and second periods is highly correlated with the first period.



Hence the correlated treatment test using first period only is highly correlated with the interaction test.

The alpha value conditional on the interaction test is >0.05.

Period × treatment interaction

The power of the test is low and alpha = 0.10 is often recommended as a decision point.

Example: Nicardipine against placebo in patients with Raynaud's phenomenon (Kahan *et al.*, 1987, given by Altman, 1991).

Kahan A, Amor B, Menkes CJ, *et al.*, Nicardipine in the treatment of Raynaud's phenomenon: a randomised doubleblind trial. *Angiology* 1987; **38**: 333-7.

Altman DG. *Practical Statistics for Medical Research* Chapman and Hall, London, 1991.

Attacks of Raynaud's phenomenon in two week periods.

Period 1 Nicardipine	Period 2 Placebo	Placebo – Nicardipine	Period 1 Placebo	Period 2 Nicardipine	Placebo – Nicardipine
16	12	–4	18	12	6
26	19	–7	12	4	8
8	20	12	46	37	9
37	44	7	51	58	–7
9	25	16	28	2	26
41	36	–5	29	18	11
52	36	–16	51	44	7
10	11	1	46	14	32
11	20	9	18	30	–12
30	27	–3	44	4	40
Mean		**1.0**			**12.0**

Looks like carry-over to me!

Period × treatment interaction

Nicardipine data, testing for interaction:

```
. ttest av , by(order)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
1st peri |      10        24.5    3.952496    12.49889    15.55883    33.44117
2nd peri |      10        28.3    4.782027     15.1221     17.4823     39.1177
---------+--------------------------------------------------------------------
combined |      20        26.4    3.050582    13.64262    20.01506    32.78494
---------+--------------------------------------------------------------------
    diff |              -3.8    6.204031               -16.83419    9.234185
------------------------------------------------------------------------------
    diff = mean(1st peri) - mean(2nd peri)                        t =  -0.6125
Ho: diff = 0                                     degrees of freedom =       18

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2739         Pr(|T| > |t|) = 0.5479          Pr(T > t) = 0.7261
```

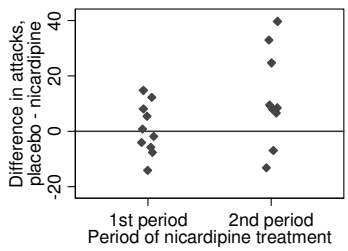
No evidence for an interaction, P = 0.5.

Period × treatment interaction

Nicardipine data, testing for interaction:

No evidence for an interaction, P = 0.5.

But there appears to be one!



Period × treatment interaction

Compare treatments:

```
. ttest diff1m2, by(order)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
1st peri |      10          -1    3.119829    9.865766   -8.057544    6.057544
2nd peri |      10          12    5.168279    16.34353    .3085399    23.69146
---------+--------------------------------------------------------------------
combined |      20         5.5    3.294733    14.73449   -1.395955    12.39595
---------+--------------------------------------------------------------------
    diff |                 -13    6.036923               -25.68311   -.3168945
------------------------------------------------------------------------------
    diff = mean(1st peri) - mean(2nd peri)                        t =  -2.1534
Ho: diff = 0                                     degrees of freedom =       18

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0225         Pr(|T| > |t|) = 0.0451          Pr(T > t) = 0.9775
```

Evidence for a treatment effect, P = 0.045.

But the estimate must be in doubt, due to the interaction.

An aside, CROS or simple paired t test?

CROS: Evidence for a treatment effect, P = 0.045.

Simple paired t test:

```
. ttest diffamb=0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 diffamb |      20         6.5     3.19745    14.29943    -.192339    13.19234
------------------------------------------------------------------------------
    mean = mean(diffamb)                                          t =   2.0329
Ho: mean = 0                                     degrees of freedom =       19

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9719         Pr(|T| > |t|) = 0.0563          Pr(T > t) = 0.0281
```

Evidence for a treatment effect, P = 0.056.

CROS adjusts for the period effect so reduces the effect of the (non-significant) period difference a bit. It is more powerful.

Period × treatment interaction

Should we test for a period × treatment interaction?

Grizzle (1965) proposed this.

Senn (1989) claims this is an error.

Jones and Kenward (1989) review the question but do not make a strong recommendation.

Jones B and Kenward MG. *Design and Analysis of Cross-Over Trials.* London: Chapman and Hall, 1989.

Period × treatment interaction

Should we test for a period × treatment interaction?

Grizzle (1965) proposed this.

Senn (1989) claims this is an error.

Jones and Kenward (1989) review the question but do not make a strong recommendation.

I find Senn's argument convincing. I would say do not test or do the two-stage analysis.

However, I think it is worth inspecting the data to see whether the assumption of no interaction required for the CROS estimate is plausible. If it is not, rely on the P value.

Period × treatment interaction

What should we do instead of Grizzle's approach using the first period only?

Senn (1989) suggests that if the estimate is needed, we should repeat the trial and design the carryover out of it, using washout periods described next.

Washout periods

A washout period is a time when the participants do not receive any active trial treatment.

Intended to prevent continuation of the effects of the trial treatment from one period to another.

Typical cross-over trial with washout periods:

washout / run-in	removes effects of pre-trial treatments
treatment 1	
washout	removes effects of treatment 1
treatment 2	
washout	removes effects of treatment 2
usual care	

Washout periods

A washout period is necessary if treatments might interact in an adverse way.

In a placebo controlled trial, we could simply make the treatment periods longer.

In drug trials, washout periods should be at least 3 × half life of drug in body (FDA).

If no washout periods are used, the treatment periods should be longer than would be required for washout and no measurements made in the time that would be needed for washout.

Two texts:

Senn S. *Cross-Over Trials in Clinical Research, 2nd ed.* Chichester: Wiley, 2002.

Jones B and Kenward MG. *Design and Analysis of Cross-Over Trials, 2nd ed.* London: Chapman and Hall, 2003.

For a brief introduction:

Altman DG. *Practical Statistics for Medical Research* Chapman and Hall, London, 1991.