A graphical presentation of the results of crossover trials

T. A. SHELDON

Department of Community Health, University of Leicester, Leicester Royal Infirmary, P.O. Box 65, Leicester LE2 7LX

1 Crossover designs are widely used in clinical trials. A graphical method of describing the data as well as visually indicating the extent of treatment * period interaction and overall treatment effect is presented. This plots the treatment differences between the periods for each individual against the mean of their treatment outcomes. It is recommended that this plot accompany the analysis of crossover trials in order to aid summary and interpretation especially for the non-statistician.

Keywords crossover trial clinical trial Hills-Armitage

Introduction

The two period crossover design of clinical trial has become increasingly popular. It is useful when there is substantial variability among subjects and problems recruiting people to a trial. An important advantage of this design is that it requires less patients than a trial using separate groups for different treatments. 'Subjects provide information for two treatment periods and thus it enables assessment of treatment differences within the more sensitive framework of within-subject variability as opposed to between subject variability', Koch (1989).

The theoretical basis for the two-period crossover trial was reviewed in this journal by Hills & Armitage (1979) in a paper which presented the models of analysis in a convenient and rational form. This paper along with a further one extending the analysis represented a landmark in the analysis and thus use of the crossover design (Armitage & Hills (1982), also see the recent text by Jones & Kenward (1989)).

In this paper an effective and simple way of presenting the results from a 'Hills & Armitage' analysis in the form of a diagram is presented. This method was first described by Clayton & Hills (1988), and then by Jones & Kenward (1989) but unfortunately has not come to the attention of the more general research community.

The crossover design

The basic two period design is shown in Figure 1. Subjects in group A receive drug A in the first period followed by drug B. Subjects in group B get the treatments in the reverse order. At the end of each period the subjects' treatment outcomes (Y) are measured giving observed values Y_1 and Y_2 at the end of periods 1 and 2 respectively. The objective is to see whether there is a significant difference in treatment effect between treatments A and B, one of which may be a placebo.

The expected values for (Y_1, Y_2) in group A is denoted by (θ_{A1}, θ_{B2}) and the expected values for (Y_1, Y_2) in group B is denoted by (θ_{B1}, θ_{A2}) thus indicating which treatment is received in which period.

The difference in treatment effect in period 1 = $(\theta_{A1} - \theta_{B1})$ (1)

The difference in treatment effect in period 2 = $(\theta_{A2} - \theta_{B2})$ (2)

Figure 1 A two period crossover design.

Only if there is no significant change in the overall treatment difference from period 1 to period 2 can we pool these treatment effects to estimate the overall treatment effect. Hopefully these are the same and the treatment effect is not influenced by the period in which drugs A and B are received. To check this two new variables are constructed:

$Z_1 = (Y_1 - Y_2)/2$ the difference in outcome at the end of the two treatments.

$Z_2 = (Y_1 + Y_2)/2$ the mean outcome over the two treatment periods.

The expected values of Z_1 are given by:

$$E(Z_1) = (\theta_{A1} - \theta_{B2})/2 \text{ for period 1 and} \quad (3)$$

$$E(Z_1) = (\theta_{B1} - \theta_{A2})/2 \text{ for period 2.} \quad (4)$$

A shift in Z_1 between the two periods is given by (3) - (4):

$$\{(\theta_{A1} - \theta_{B2}) - (\theta_{B1} - \theta_{A2})\}/2 = \{(\theta_{A1} - \theta_{B1}) + (\theta_{A2} - \theta_{B2})\}/2 = \{(1) + (2)\}/2$$

This is the average or pooled estimate of the treatment effect in period 1 and period 2 and so a measure of the overall treatment effect.

The expected values of Z_2 are given by:

$$E(Z_2) = (\theta_{A1} + \theta_{B2})/2 \text{ for period 1 and} \quad (5)$$

$$E(Z_2) = (\theta_{B1} + \theta_{A2})/2 \text{ for period 2.} \quad (6)$$

A shift in Z_2 between the two periods is given by (5) - (6):

$$\{(\theta_{A1} + \theta_{B2}) - (\theta_{B1} + \theta_{A2})\}/2 = \{(\theta_{A1} - \theta_{B1}) - (\theta_{A2} - \theta_{B2})\}/2 = \{(1) - (2)\}/2$$

This is a measure of how much the treatment outcome changes between period 1 and period 2. If the shift in Z_2 is not significantly different from zero then we can assume there is no period * treatment interaction and so the shift in Z_1 between the periods is taken as the 'true' treatment effect. The tests for shift in location between Z_1 and Z_2 are the standard tests in the Hills-Armitage analysis.

It is proposed that these tests be accompanied by a plot of treatment differences (Z_1) against treatment means (Z_2) distinguishing subjects who received drug A first (group A) and those initially receiving drug B (group B). This is explained and illustrated below by means of the results from three studies.

Study I

This is a study quoted by Hills & Armitage (1979) where each of 29 patients took part in a randomised double blind two-period crossover study to examine the effects of a new drug for the treatment of enuresis. Group A with 17 patients received the drug in period 1 for 14 days followed by a placebo in period 2 for 14 days. Group B with 12 patients was treated in the reverse order. The outcome measure is the number of dry nights in the 2 weeks, groups are compared using a *t*-test.

The data are evaluated for period * treatment interaction. This tests whether the effects of the placebo and the supplement are the same in the two periods, i.e. whether there is a difference in the mean number of dry nights (Z_2) between the groups.

Table 1 indicates that there is no significant shift in Z_2 between the groups ($P = 0.19$) and thus no evidence for a treatment * period interaction. The overall treatment effect is evaluated by testing for a shift in the distribution of treatment differences (Z_1) between group A and B. Table 1 shows that there is a statistically significant reduction in the mean number of wet nights associated with the drug B compared to drug A of around 2 ($P = 0.0017$).

Plot of treatment difference vs treatment means (Z_1/Z_2) Figure 2 shows a graph of the difference in number of dry nights between periods (Z_1) against the mean number of dry nights across both periods (Z_2) for each patient, indicating membership of group A or B.

If there is a treatment * period interaction then the As and Bs will be shifted relative to one another along the Z_2 axis, horizontally. The size

Table 1 Number of dry nights

	Z_1	Z_2
Mean for group A ($n = 17$)	1.41	6.71
Mean for group B ($n = 12$)	-0.67	8.25
Estimated shift from group B to group A	2.08	-1.54
Approx. 95% confidence limits for the shift	(0.86, 3.3)	(-3.9, 0.8)
Significance of the shift (P)	0.0017	0.19

Figure 2 Plot of Z_1 vs Z_2 number of dry nights.

of overall treatment effect is shown by the degree to which the As and Bs are shifted relative to one another along the Z_1 axis, vertically.

The Z_1/Z_2 plot in Figure 2 shows no horizontal shift between the As and the Bs and thus no treatment * period interaction. However there is a sizeable shift in the treatment differences (Z_1) between the groups, the As being above the Bs, so confirming an overall treatment effect. The Z_1/Z_2 plot thus instantly conveys the results given in Table 1. The size of the overall treatment effect can be judged from vertical displacement between the centroids of the cloud of As and Bs.

The graph also conveys other useful information. The Z_2 value for each subject is the mean number of dry nights over the two periods. Thus the points can be used to identify the spread and centre of the distribution of dry nights and to identify outliers.

The tests for treatment effect and interaction are based on the expected values of Z_1 and Z_2 shown by the centre of the distributions, however the power of these tests to detect differences in Z_1 and Z_2 is affected by the variation of these values. The smaller the variance, the greater the power of the test to detect a difference at a given level of significance.

Because the graph plots the individual values of Z_1 and Z_2 on the same graph it clearly shows the degree of variability. In particular what is demonstrated is that the points are more spread out in a horizontal than a vertical direction, showing the greater variability of the Z_2 values (mean treatment outcomes) compared to the Z_1 values (treatment differences). This illustrates clearly that the two-period crossover trial has less power to detect period * treatment interaction than an overall treatment effect. By accompanying the table of analysis with the plot, the difference in variance and thus relative power of the tests is made open and explicit.

Study II

This example shows how the plot is also effective at conveying information in trials with small numbers of subjects and in indicating where there is poor power to detect treatment effects. Eleven subjects took part in a randomised double-blind two-period crossover study to examine the effect of dietary supplementation with fatty acids on blood pressure. Group A with five subjects received the active treatment (A) followed by a placebo (B) and the six subjects in group B were treated in the reverse order.

Since there were so few subjects, no distributional assumptions are made and a non-parametric method is used for comparing groups. This is the Wilcoxon rank sum test (Mann-Whitney U) and is calculated here using the statistical package MINITAB (Ryan *et al.*, 1985).

The method of analysis used in Study I is repeated here for standing diastolic, lying diastolic and systolic blood pressure. The Z_1/Z_2 plot in Figure 3 shows a clear vertical shift indicating a significant treatment effect for standing diastolic pressure. This is confirmed in Table 2 where a 5 mm Hg drop in standing diastolic blood pressure is associated with the active treatment A compared with placebo B ($P = 0.008$). The plot also makes clear that group A has a hypotensive and possibly a hypertensive subject, demonstrating its use in identifying possible outliers.

The analysis is repeated for lying diastolic blood pressure. Figure 4 shows that whilst there is no shift in Z_2 there is some evidence for separation vertically on the Z_1 scale, however this is not as clear as in the standing diastolic case (Figure 3). This is confirmed in Table 3 which shows no significant difference in Z_2 but some evidence for a shift in Z_1 ($P = 0.068$). It is clear from the variability of the treatment differences (spread of points in a vertical direction) that the

Figure 3 Plot of Z_1 vs Z_2 diastolic-standing (mm Hg).

Table 2 Diastolic blood pressure—standing (mm Hg)

	Z_1	Z_2
Mean for group A ($n = 5$)	-2	80.5
Median for group B ($n = 6$)	3	81.75
Estimated shift from group B to group A	-5	-2.25
Approx. 95% confidence limits for the shift	(-10, -2.5)	(-36.5, 19.5)
Significance of the shift (P)	0.008	0.72

Table 3 Diastolic blood pressure—lying (mm Hg)

	Z_1	Z_2
Median for group A ($n = 5$)	-3	68
Median for group B ($n = 6$)	4.3	69.5
Estimated shift from group B to group A	-5	-1.5
Approx. 95% confidence limits for the shift	(-9.5, 1.5)	(-15.0, 11.5)
Significance of the shift (P)	0.068	1.0

small size of the study does not give enough power to detect a 5 mm Hg treatment effect.

Lastly, a look at the results for standing systolic indicates neither shift between the groups in Z_2 nor in Z_1 (Figure 5). Thus there is no treatment effect on systolic blood pressure. This is confirmed in Table 4. It should be noted how the graph clearly identifies the hypertensive in group A.

Study III

This example shows the use of the plot in demonstrating a period * treatment interaction and is based on data reported by Grizzle (1965) which compares the effect of drugs A and B on haemoglobin levels.

The plot in Figure 6 shows that there is a difference in treatment means (Z_2) between

Figure 4 Plot of Z_1 vs Z_2 diastolic—lying (mm Hg). **Figure 5** Plot of Z_1 vs Z_2 systolic—standing (mm Hg).**Table 4** Systolic blood pressure—standing (mm Hg)

	Z_1	Z_2
Median for group A ($n = 5$)	-0.5	102.5
Median for group B ($n = 6$)	2.0	107.0
Estimated shift from group B to group A	-2.5	-1.0
Approx. 95% confidence limits for the shift	(-5.5, 5.0)	(-17.0, 28.0)
Significance of the shift (P)	0.24	0.78

Table 5 Difference between pre- and post-treatment Hb

	Z_1	Z_2
Median for group A ($n = 6$)	-0.15	0.48
Median for group B ($n = 8$)	-0.55	-0.63
Estimated shift from group B to group A	0.5	0.93
Approx. 95% confidence limits for the shift	(-0.25, 1.85)	(0.20, 1.7)
Significance of the shift (P)	0.25	0.03

Figure 6 Difference between pre- and post-treatment Hb.

group A and B as shown by a horizontal shift. This implies a treatment * period interaction. The plot also appears to show that whilst there is a treatment effect for patients in group B, who received drug B first, there is none in group A.

In other words Figure 6 demonstrates that the order in which the treatments are received does make a difference. Presumably, treatment A has a strong carry-over effect relative to B and thus patients who receive A first, experience the effect over both periods thus showing no treatment difference. On the other hand, patients receiving B first only experience the effect in the second period thus showing a treatment difference between the period. This is confirmed by Table 5 which shows a significant period * treatment interaction ($P = 0.03$). Clearly there is a treatment effect but differential carry-over

means that the overall estimate of its effect given in Table 5 is wrong.

Conclusion

The Hills and Armitage analysis of a two period crossover trial tests for shifts in location between the distribution of treatment differences (Z_1) and treatment means (Z_2) between the two groups. By plotting Z_1 against Z_2 we get an immediate visual picture of the degree of treatment * period interaction and overall treatment effect.

In addition, because the value of Z_2 is given for each subject the distribution of the response variable is clear and outliers can be identified. The variance of the values and so the relative power of the tests to detect the interaction and overall treatment effects is displayed and the way the treatment affects individuals differently may also be observed.

The Z_1/Z_2 plot presents, in an easily interpretable graph, descriptive data on an individual basis as well as the values of interest in the analysis of the trial. It may also be extended to present the results of a three-period crossover design (Clayton & Hills, 1988). Its routine use in the presentation of crossover trials is therefore highly recommended.

I thank Drs A. Heagerty, D. Maciver, P. McNally and J. Ollerenshaw of the Department of Medicine, Leicester Royal Infirmary for permission to use their study data and David Clayton for his helpful comments during the preparation of this paper.

References

- Armitage, P. & Hills, M. (1982). The two-period crossover trial. *Statistician*, **31**, 119-131.
- Clayton, D. & Hills, M. (1988). A two period cross-over trial, In *The statistical consultant in action*, eds Hand, D. & Everitt, B. Cambridge: Cambridge University Press.
- Grizzle, J. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, **21**, 467-80.
- Hills, M. & Armitage, P. (1979). The two period cross-over trial. *Br. J. clin. Pharmacol.*, **8**, 7-20.
- Jones, B. & Kenward, M. (1989). *Design and analysis of crossover trials*. London: Chapman and Hall.
- Koch, G. (1989). A two period cross-over design for the comparison of two active treatments and a placebo. *Stats. in Med.*, **8**, 487-504.
- Ryan, B., Joiner, B. & Ryan, T. (1985). *MiniTab Handbook*, 2nd edition. Boston: Duxbury Press.

(Received 2 March 1990,
accepted 14 May 1990)