**REFERENCE ONLY**

# UNIVERSITY OF LONDON THESIS

Degree PhD Year 2006 Name of Author LEE SHH

# Use of the two-stage procedure for analysis of cross-over trials in four aspects of medical statistics

Sally Hung-Hsin Lee

A thesis for the degree of doctor of philosophy of the University of London

Department of Epidemiology and Public Health

University College London

August 2005





ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

## Abstract

The two-stage procedure had been used as a standard method to analyse cross-over trials for many years before its major deficiency was found. The inflated type I error rate of the two-stage procedure indicates that there have been more trials which have produced false positive results than originally believed.

Because of the 24 years gap between the introduction of the method and the publishing of its deficiency, it is conceivable that the impact of the change in the perception about the validity of the two-stage procedure might not have taken effect overnight. The objective of this thesis is to examine the impact of the change in the perception about the validity of the two-stage procedure on four different aspects of medical statistics. The areas of medical statistics include both applications of, and references to, the analysis of cross-over trials to give a full picture of the use of the two-stage procedure. Methods used are citation analysis for all scientific journals, systematic review for medical journals, comprehensive review for general medical statistics books and questionnaire survey for the pharmaceutical industry.

The results have been inconclusive in terms of the estimation of how prevalently the two-stage procedure has been used. However, the four studies demonstrated that the analysis of cross-over trials is often associated with the two-stage procedure while the deficiency of the two-stage procedure has not been generally acknowledged. It can be concluded that further understanding of

the two-stage procedure and better references in the analysis of cross-over trials are needed in all four areas of medical statistics.

## Acknowledgement

It has been a very long way and I would like to thank everyone who has helped me and supported me during this period of time. I am most grateful towards:

Professor Stephen Senn for his guidance, constant support and incredible patience.

Professor Aubrey Sheiham for his advice in academic and other matters.

Martin Shipley for his comments and final corrections.

My parents and sister for their love and faith in me.

Raphaël for his love and encouragements.

I would also like to thank PSI committees who have helped with the distribution of the survey questionnaire and PSI members who have responded to the survey.

Without them, it would have been impossible to produce this thesis.

# Chapter 1 - Literature Review

## 1.1. Introduction

The 'cross-over' trial is a particular design of clinical trial for which no consensus about statistical analysis has been reached. Since 1965, a debate has been going on as to what constitutes a valid method for statistical analysis of such trials. The method that was introduced to target the major weakness of cross-over trials, the 'carry-over effect', was proven to be "too potentially misleading to be of practical use" [8] 24 years after it was first introduced.

This thesis looks into the applications of analysis of cross-over trials in four realms of medical statistics (medical journals, statistical journals, medical statistics books and the pharmaceutical industry), using four different research methods (citation analysis, systematic review, comprehensive review and questionnaire survey).

## 1.2. The cross-over trial

The cross-over trial is one of the most frequently used designs of clinical trials. The main difference between a cross-over trial and a parallel trial (the most commonly used clinical trial design) is the number of treatments given to each patient. In a parallel trial, each patient is randomised to a treatment group where only one treatment is given. Treatment effects can only be evaluated by between-patient differences (i.e., comparing results from different treatment groups). By contrast, in a cross-over trial, each treatment

group involves sequences of two or more treatments (depending on the design of the cross-over trial). Treatment effects of a cross-over trial can then be compared using within-patient difference. In other words, each patient acts as his/her own control in comparing the treatments.

### 1.2.1. Definition

"A cross-over trial is one in which subjects are given sequences of treatments with the object of studying differences between individual treatments (or sub-sequences of treatments)" [1].

### 1.2.2. Advantages of the cross-over design

One of the most important advantages of the cross-over design is that the treatment effect can be evaluated using within-patient comparisons. The design of cross-over trials allows each participant to receive more than one treatment or, sometimes, all the treatments to be evaluated by the trial. It gives information on the effects of different treatments in each participant, therefore providing statistics for within-patient comparison. Such a feature is especially valuable for diseases for which the between-patient variation is substantial.

Another related advantage of the cross-over design is that the sample size required to achieve the same statistical power is much smaller than what would be required for a parallel trial. As clinical trials are very costly and recruiting patients is very time

consuming, saving on sample size is, without a doubt, a very attractive feature for clinical researchers.

In addition to the advantages above that benefit the researchers, patients may also benefit from a cross-over trial. In contrast to a parallel trial where patients may have little or no benefit if they are randomised to the treatment group where only the placebo is given, patients of a cross-over trial receive at least one active treatment. Furthermore, because participants may have the chance to take different treatments they may be able to find their optimum treatment. Experimenting with different treatments may also help patients to gain a better understanding of their disease.

### 1.2.3. Disadvantages of the cross-over design

The main disadvantage of cross-over trials is known as the 'carry-over' effect (sometimes referred to as the 'residual effect'). It is defined as "the persistence (whether physically or in terms of effect) of a treatment applied in one period in a subsequent period of treatment" [1]. The United States Food and Drug Administration (FDA) considers the carry-over effect "the chief difficulty" [31] of such designs.

With the interference of the carry-over effect, a difference in the effect of treatments is very difficult to evaluate. The practical solution is to introduce a 'wash-out period', a period of time in which no active treatment is given to the patients between any two treatment periods. However, a wash-out period is less appropriate

for disease types for which patients require constant treatments (such as epilepsy, Alzheimer and other psychotic disorders).

Another disadvantage of the cross-over design is that it is not universally applicable: it is not appropriate for all therapeutic areas, especially for diseases which are curable, fatal or acute. The reason is that the baseline conditions of patients should be similar for all treatment periods. Therefore, therapeutic areas where patients might be cured or die before the second or further treatment period commences are not suitable for cross-over designs.

An additional disadvantage of cross-over designs affects both patients and clinical researchers. The fact that cross-over trials involve multi-treatment periods usually implies a longer trial period than for a parallel design. Patient compliance is therefore likely to be lower and they are more likely to leave the trial before it is completed. With similar drop out rates (i.e., the percentage of patients who drop out before the trial is complete amongst all patients who were enrolled into the trial), the problem is more serious within a cross-over trial than a parallel trial. The reason is that, in a cross-over trial, each patient is given more than one treatment and the analysis of the treatment effect is based on the measurements of all treatments they were given. For a parallel trial, each patient is given one treatment and therefore accounted for measurements of one treatment only. Furthermore, a longer

duration of the trial for both investigators and participants may cause more inconvenience to patients [1].

### 1.2.4. Therapeutic areas of the cross-over design

The therapeutic areas where cross-over designs have been used extensively include asthma, migraine, hypertension, insomnia, diabetes, angina and bioequivalence studies. Other therapeutic areas where cross-over trials are sometimes used but are more debatable (as endpoints of these treatments are not reversible) include the treatments of infertility [39] and dental care.

## 1.3. Analyses of cross-over trials

The main concern about analyses of cross-over trials is the issue with the carry-over effect. Although the carry-over effect is a design problem which involves pharmacokinetics and pharmacodynamics, statisticians have attempted to find a solution to 'prove' that the carry-over effect has not occurred in the trial. Analyses of cross-over trials before this 'solution' was introduced were straightforward but with some uncertainty. A statistical solution to the carry-over effect was first introduced by Grizzle in 1965. The method is known as the two-stage procedure and is also often called the Hills-Armitage analysis [8], especially in the UK. The context of the two-stage procedure and its developments are described below.

### 1.3.1. The two-stage procedure of cross-over trials

The first stage of this method is to determine whether there was an unequal carry-over effect between the two treatments. The evaluation of the treatment effect is determined by the significance level of the test for the unequal carry-over effect. If the differential carry-over effects is found to be significant at the 10% level as suggested, the treatments effect would be evaluated using the first period data only. As a result, the advantage of the within-patient comparison cannot be utilised and the sample size which is calculated based on the cross-over design will not be sufficient to achieve the statistical power it was designed for. If the unequal carry-over effects were found to be non-significant at the 10% level, data from both periods would be used to evaluate the treatment effect, in the same way the treatment effect was analysed before the two-stage procedure was introduced.

### 1.3.2. History of the two-stage procedure

The two-stage procedure was first introduced by Grizzle (1965) [4]. He demonstrated how unequal carry-over effects could be tested. It was suggested in this paper that only if the null hypothesis of the carry-over effects of both treatments being equal was not rejected at the 10% level, was analysis of variance using data from both periods appropriate. He summarised this: "the two-period changeover design is preferable when the residual effects of the treatments are equal and the correlation between the responses to

the two treatments is positive. Otherwise the design in which there is random assignment to a single treatment is preferable."

However, prior to Grizzle's paper, it had not been suggested how treatment effects should be analysed if unequal carry-over effects were found to be significant. Grizzle's approach was popularised by Hills and Armitage in 1979 [7], who suggested that "We should therefore place much more reliability on the comparison of treatments in period 1, and it seems sensible to study these alone when a treatment by period interaction is present." They further added that, "If the results of the trial suggest a definite interaction between treatments and periods then the chance has not come off and the treatment comparison should be based on the first period alone. This might seem to be a great loss since this is just what would have happened in a parallel group study but all too often the size of the trial, chosen with cross-over in mind, is too small for a proper comparison between groups." Regarding the issue of sample size and achieving adequate power, Hills and Armitage [7] suggested that a cross-over trial should always be large enough for it to be analysed as a parallel group study, and also for any appreciable interaction to be detected on a between-subject basis.

### 1.3.3. The deficiencies of the two-stage procedure

Hills and Armitage [7] mentioned that the two-stage procedure was not without flaws. They showed that the statistical power of the two-stage procedure was especially low when the ratio of true treatment difference to standard error was low. However, it was not

until 10 years later (in 1989) that the more serious deficiency of testing for the carry-over effect was demonstrated in a paper by Freeman in *Statistics in Medicine* [8]. The key finding of this paper was that the type I error rate (the probability of falsely declaring there to be a difference between the treatments) of the two-stage procedure is much higher than the nominal 5% claimed for it. Freeman concluded that "the two-stage analysis is too potentially misleading to be of practical use".

The overall type I error rate of the two-stage procedure is the probability weighted sum of the type I error rate of both arms of the results of significance of the carry-over effect. The key to such a result was that the correlation between the test for the carry-over effect and the test of evaluating first period data only is actually as high as 0.866. Senn [22] demonstrated that instead of the 5% claimed, the actual type I error rate of the two-stage procedure is between 7% and 9.5%.

### 1.3.4. Change in validity of statistical methods for the analysis of cross-over trials

The development of the analysis of cross-over trials has been mainly marked by the development of the two-stage procedure which was marked by two major events (the introduction of the two-stage procedure and the finding of its major deficiency) and can therefore be classified into three periods: the pre-two-stage procedure period, the two-stage procedure period and the post-two-stage procedure period. The optimal method for analysing cross-

over trials was different in those three periods. Such a division provides the background to possible methods used in the analysis of cross-over trials of research objects.

**The pre-two-stage procedure period** (before 1965): before the two-stage procedure was introduced by Grizzle [4] in the *British Journal of Pharmacology.* References on analyses of cross-over trials published before then recommended using cross-over differences without taking carry-over effect into consideration.

**The two-stage procedure period** (1965-1989): Although Hills and Armitage (1979)[7] had ascertained one of the main deficiencies of the two-stage procedure, its validity was only challenged by Freeman [8] in 1989. He concluded that the two-stage procedure was too potentially misleading for practical use. One might then describe the analysis of cross-over trials as having entered a **post-two-stage procedure era.** Although the finding of the deficiency of the two-stage procedure by Freeman (1989) [8] was an important breakthrough, the two-stage procedure has not been completely eliminated from the analysis of cross-over trials. Evidence of its applications can often be found in general medical statistics textbooks and trial reports in medical journals. The two-stage procedure remains one of the most recommended methods of analyses of cross-over trials in general medical statistics books. Despite its lack of statistical validity, testing for carry-over effect can also be found in statistical analyses sections of cross-over trial reports in medical journals.

## 1.4. Objective of the thesis

With the change in accepted validity of the statistical methods for analyses of cross-over trials, the objective of this thesis is to investigate methods used to analyse the two-period two treatment (AB/BA) designs, with the main focus on applications and recommendations of the two-stage procedure in four domains of medical statistics (medical journals, statistical journals, general medical statistics books and pharmaceutical companies) where analyses of cross-over trials are applied and advised. The core question to be answered is whether the finding of the high type I error rate had any impact on the use of the two-stage procedure in different realms of medical statistics. This question can be answered by looking at:

- how frequently the two-stage procedure has been used and recommended in different realms of medical statistics at different periods of its development;
- how up to date the information on methods of analyses of cross-over trials is in different realms of medical statistics.

# Chapter 2 - Statistical Analyses for AB/BA Trials

## 2.1. Introduction

This chapter describes the statistical methods that have been used to analyse cross-over trials. Part A of this chapter presents potential analyses which do not include the carry-over effect in the statistical model. For the purpose of referencing the results of the systematic review of the medical literature and the comprehensive review on general medical statistics books, the method of testing for the carry-over effect and the two-stage procedure will be illustrated in part B. Relevant literature on the performance of the two-stage procedure will be reviewed in part C.

## 2.A The correct analysis of cross-over trials

This section illustrates how cross-over trials can be correctly analysed. The effects that are included in these analyses include the treatment effect, the period effect and patient effects. This does not include any test on the carry-over effect as the potential of the carry-over effect should have been dealt with by design.

### 2.A.1 Statistical model

For the patient $j$ of the treatment sequence $i$, the observation during the $k$-th period can be denoted as $y_{ijk}$.

A commonly used statistical model for an AB/BA trial is as follows:

$$y_{ijk} = \mu + \pi_k + \varphi_l + \varepsilon_{ijk} \qquad (2.A.1)$$

where

$y_{ijk}$ = the measurement of $j$-th patient in the $i$-th sequence at the $k$-th period,

$\mu$ = general mean

$\pi_k$ = the effect of the $k$-th period ($k$ =1 or 2),

$\varphi_l$ = the direct effect of the drug $l$ [For patients randomised to the treatment sequence 1 ($i$ =1), $l$ =A when $k$ =1, $l$ =B when $k$ =2. For patients randomised to the treatment sequence 2 ($i$ =2), $l$ =B when $k$ =1, $l$ =A when $k$ =2],

$\xi_{ij}$ = the effect of patient $j$ of the $i$-th sequence. This can be treated as a fixed or random effect. It is assumed to be fixed at this stage. Patient effect as a random effect will be discussed in Section 2.A.7.

$\varepsilon_{ijk}$ = the random fluctuation which is normally distributed with mean 0 and variance $\sigma_\varepsilon^2$, and is independent of the patient effect $\xi$.

For an AB/BA trial, patients ($N=n_1$) randomised to sequence 1 receive treatment A during the first period and treatment B during the second period. The order of the treatments is reversed for patients ($N=n_2$) randomised to sequence 2.

The design and the parameters of the statistical model for each period in each sequence are as follows:

**Table 2.A.2**

| $n = n_1 + n_2$ | **Period 1** | **Period 2** |
|---|---|---|
| **Sequence 1** ($N=n_1$) | Receiving Treatment **A** $y_{1j1} = \mu + \pi_1 + \varphi_a + \varepsilon_{1j1}$ | Receiving Treatment **B** $y_{1j2} = \mu + \pi_2 + \varphi_b + \varepsilon_{1j2}$ |
| **Sequence 2** ($N=n_2$) | Receiving Treatment **B** $y_{2j1} = \mu + \pi_1 + \varphi_b + \varepsilon_{2j1}$ | Receiving Treatment **A** $y_{2j2} = \mu + \pi_2 + \varphi_a + \varepsilon_{2j2}$ |

## 2.A.2 Evaluate the treatment effect using simple cross-over differences

The most straightforward way to analyse a cross-over trial is to use a matched-pairs t test on the treatment difference. To estimate the difference $\varphi_b - \varphi_a$ between treatment A and treatment B, differences $y_{1j2} - y_{1j1}$ are calculated for patients randomised to sequence 1 and $y_{2j1} - y_{2j2}$ for patients randomised to sequence 2.

The mean of treatment differences $\dfrac{\sum_{j=1}^{n_1}(y_{1j2} - y_{1j1}) + \sum_{j=1}^{n_2}(y_{2j1} - y_{2j2})}{n_1 + n_2}$ and their standard deviation are then used to test against the null hypothesis that there are no treatment differences using the matched-pairs t test. The underlying assumptions of this analysis are that the cross-over differences are distributed at random about the true treatment effect and that any differences in the

observations between themselves are random. This was also the analysis commonly adopted before the method to test for the carry-over effect (the two-stage procedure) was introduced.

### 2.A.3 Testing for the period effect

It was demonstrated by Hills and Armitage [7] that the period difference $\pi_2 - \pi_1$ can be estimated by $\frac{1}{2}\left[\frac{\sum_{j=1}^{n_1}(y_{1j2} - y_{1j1})}{n_1} - \frac{\sum_{j=1}^{n_2}(y_{2j1} - y_{2j2})}{n_2}\right]$. To test the hypothesis that $\pi_2 = \pi_1$, an independent t-test is used to compare $(y_{1j2} - y_{1j1})$ of patients randomised to sequence 1 and $(y_{2j1} - y_{2j2})$ of patients randomised to sequence 2.

### 2.A.4 Adjusting for the period effect

The period effect can also be adjusted for when testing for the treatment effect. This can be done by using an independent t-test to compare the period difference, i.e., period 2 – period 1 for both sequence groups, $\frac{\sum_{j=1}^{n_1}(y_{1j2} - y_{1j1})}{n_1}$, $\frac{\sum_{j=1}^{n_2}(y_{2j2} - y_{2j1})}{n_2}$, with $(n_1 + n_2 - 2)$ degrees of freedom[19].

## 2.A.5 An example of analysis

An example of AB/BA trial dataset from Hills and Armitage [7] is used to illustrate the three types of analyses discussed in Sections 2.A.2, 2.A.3 and 2.A.4. The trial was conducted on patients with enuresis (bed-wetting), the numbers shown in the columns of period 1 and period 2 are the numbers of dry nights out of 14 days that they were given the treatment or placebo. The analysis will first be calculated step by step such as might be done using a scientific calculator or spreadsheet, followed by the SAS® code and output.

Example: Twenty-nine patients are randomised to either group 1 or group 2. Patients ($n_1$=17) randomised to the sequence 1 receive the active treatment in the period 1 and placebo in the second period. The order of drug/placebo is in the reverse order for patients ($n_2$=12) randomised to sequence 2. The observations of each patient from each treatment period are listed in Table 2.A.3. The sums, means and standard deviations were calculated using Microsoft Excel®.

Table 2.A.3

| Sequence 1 (drug/placebo) | | | | |
|---|---|---|---|---|
| **Patient Number** | **Period 1 (drug)** | **Period 2 (placebo)** | **Treatment Difference (drug-placebo)** | **Period Difference** |
| $n_1$=17 | $Y_1$ | $Y_2$ | $Y_1 - Y_2$ | $Y_1 - Y_2$ |
| 1 | 8 | 5 | 3 | 3 |
| 3 | 14 | 10 | 4 | 4 |
| 4 | 8 | 0 | 8 | 8 |
| 6 | 9 | 7 | 2 | 2 |
| 7 | 11 | 6 | 5 | 5 |
| 9 | 3 | 5 | -2 | -2 |
| 11 | 6 | 0 | 6 | 6 |
| 13 | 0 | 0 | 0 | 0 |
| 16 | 13 | 12 | 1 | 1 |
| 18 | 10 | 2 | 8 | 8 |
| 19 | 7 | 5 | 2 | 2 |
| 21 | 13 | 13 | 0 | 0 |
| 22 | 8 | 10 | -2 | -2 |
| 24 | 7 | 7 | 0 | 0 |
| 25 | 9 | 0 | 9 | 9 |
| 27 | 10 | 6 | 4 | 4 |
| 28 | 2 | 2 | 0 | 0 |
| **Sum** | 138 | 90 | 48 | 48 |
| **Mean** | 8.12 | 5.29 | 2.82 | 2.82 |
| **Standard Deviation** | 3.84 | 4.25 | 3.47 | 3.47 |

| Sequence 2 (Placebo/drug) | | | | |
|---|---|---|---|---|
| **Patient Number** | **Period 1 (placebo)** | **Period 2 (drug)** | **Treatment Difference (drug – placebo)** | **Period difference** |
| $n_2$=12 | $Y_1$ | $Y_2$ | $Y_2 - Y_1$ | $Y_1 - Y_2$ |
| 2 | 12 | 11 | -1 | 1 |
| 5 | 6 | 8 | 2 | -2 |
| 8 | 13 | 9 | -4 | 4 |
| 10 | 8 | 8 | 0 | 0 |
| 12 | 8 | 9 | 1 | -1 |
| 14 | 4 | 8 | 4 | -4 |
| 15 | 8 | 14 | 6 | -6 |
| 17 | 2 | 4 | 2 | -2 |
| 20 | 8 | 13 | 5 | -5 |
| 23 | 9 | 7 | -2 | 2 |
| 26 | 7 | 10 | 3 | -3 |
| 29 | 7 | 6 | -1 | 1 |
| **Sum** | 92 | 107 | 15 | -15 |
| **Mean** | 7.67 | 8.92 | 1.25 | -1.25 |
| **Standard Deviation** | 2.99 | 2.81 | 2.99 | 2.99 |

### 2.A.5.1 Testing the cross-over difference using paired t-test

If the period effect is ignored, the treatment effect is evaluated by the cross-over differences as described in Section 2.A.2. The t-statistics for a paired t-test is defined as: $t = \frac{(\bar{x} - \mu)}{SD/\sqrt{N}}$

where:

$\bar{x}$ are the within-patient difference between treatments A and B,

$\mu$ is the population mean under the null hypothesis (usually =0 and this is the value adopted here)

$N = n_1 + n_2$,

and SD denotes the standard deviation of the mean

Mean of cross-over difference= sum(cross-over difference) / $N$ = 63 / 29 = 2.17

Degrees of freedom= $(n_1-1) + (n_2-1)$ = (17-1) + (12-1) = 27

SD (Standard deviation) = 3.32

SE (Standard error of the mean) = SD / $\sqrt{N}$ = 3.32 / $\sqrt{29}$ = 0.61

t statistics = mean/SE = 3.53

p-value = 0.0015

The p value of the paired t-test shows that the difference between the placebo and the treatment is significant at the 5% level.

## 2.A.5.2 Calculation of the independent t-test

The independent t-test is used to examine the treatment effect with an adjustment for the period effect (Sections 2.A.3 and 2.A.4). The t-statistics of an independent t-test is defined as:

$$t = \frac{(\overline{d_1} - \overline{d_2}) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{(n_1 - 1)SD_1^{\,2} + (n_2 - 1)SD_2^{\,2}}{n_1 - n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$\overline{d_i}$ are the within-patient difference between periods for sequence $i$ $(i = 1,2)$, and

$\mu_i$ is the population mean under the null hypothesis (both $\mu_1$ and $\mu_2$ are set at 0)

The formulae used to calculate the t-statistics for the tables of results (2.A.5.3 and 2.A.5.4) are listed in Table 2.A.4.

**Table 2.A.4**

| Independent t-test | | |
|---|---|---|
| $N = n_1 + n_2$ | **Sequence 1** $n_2$ | **Sequence 2** $n_1$ |
| Mean (of the period difference) | $\overline{d}_1$ | $\overline{d}_2$ |
| Standard Deviation | $SD_1$ | $SD_2$ |
| Correlated sums square | $SS_1 = (n_1 - 1) \times SD_1^2$ | $SS_2 = (n_2 - 1) \times SD_2^2$ |
| Pooled estimate of variance | $EV = (SS_1 + SS_2)/(N - 2)$ | |
| Difference in Mean | $D = \overline{d_1} - \overline{d_2}$ | |
| Standard error of difference in means | $S.E. = \sqrt{\{EV \times (1/n_1 + 1/n_2)\}}$ | |
| t statistics | $= D/S.E.$ | |
| Pr > \|t\| (p-value) | If the experiment is conducted an infinity of times, the probability of obtaining this t-statistic or more extreme. | |

## 2.A.5.3 Testing for the period effect

The period difference of each patient and its mean and standard deviation in each sequence were already provided in Table 2.A.3. Using the formulae shown on Table 2.A.4, the t-statistics for the period effect is 1.28. The correspondent p value at 27 degrees of freedom is 0.21. The period effect was therefore not significant at the 5% level.

**Table 2.A.5**

| Testing for the period difference using Independent t-test | | |
|---|---|---|
| | **Sequence 1 (Drug/placebo)** $n_1 = 17$ | **Sequence 2 (Placebo/drug)** $n_2 = 12$ |
| Mean | 2.82 | 1.25 |
| Standard Deviation | 3.47 | 2.99 |
| Correlated sums square | = 16 x $3.47^2$ = 192.47 | = 11 x $2.99^2$ = 98.25 |
| Pooled estimate of variance | = (192.47+98.25) / (29-2) = 10.77 | |
| Difference in Mean | = 2.82 – 1.25 = 1.57 | |
| Standard error of difference in means | = $\sqrt{\{10.77 \times (1/17 + 1/12)\}}$ = 1.23 | |
| t statistics | = 1.57/1.23 = 1.28 | |
| $Pr > \|t\|$ | 0.21 | |

## 2.A.5.4 Adjusting for the period effect

The statistics for the test of the treatment effect with adjustment for the period effect are similar to those for the test for the period effect except the mean difference. The t-statistics derived is 3.31 with a correspondent p-value of 0.0028. The treatment effect remains significant at the 5% level after adjustment for the period effect.

Table 2.A.6

| Examining the treatment effect with adjustment for the period effect using independent t-test | | |
|---|---|---|
| | **Sequence 1 (Drug/placebo)** $n_1 = 17$ | **Sequence 2 (Placebo/drug)** $n_2 = 12$ |
| Mean (of the period difference) | 2.82 | -1.25 |
| Standard Deviation | 3.47 | 2.99 |
| Correlated sums square | = 16 x $3.47^2$ = 192.47 | = 11 x $2.99^2$ = 98.25 |
| Pooled estimate of variance | = ( 192.47 + 98.25 ) / ( 29 - 2 ) = 10.77 | |
| Difference in Mean | = 2.82 – (-1.25) = 4.07 | |
| Standard error of difference in means | = √{10.77 x (1/17 + 1/12)} = 1.23 | |
| t statistics | = 4.07 / 1.23 = 3.31 | |
| $\Pr > \|t\|$ | 0.0028 | |

## 2.A.5.5 SAS programming for the analysis of cross-over trials

The SAS codes for the three types of analyses are as following:

```
input id group drug placebo;
/* specifying variables*/
/*id = patient's id number*/
/*group = randomised sequence; 1 = drug/placebo, 2 = place/drug*/
/*drug = measurement of the patient of the drug period*/
/*placebo = measurement of the patient of the placebo period*/
Datalines; /*data entry*/
1 1 8 5
2 2 11 12
3 1 14 10
4 1 8 0
5 2 8 6
6 1 9 7
7 1 11 6
8 2 9 13
9 1 3 5
10 2 8 8
11 1 6 0
12 2 9 8
13 1 0 0
14 2 8 4
15 2 14 8
16 1 13 12
17 2 4 2
18 1 10 2
19 1 7 5
20 2 13 8
21 1 13 13
22 1 8 10
23 2 7 9
24 1 7 7
25 1 9 0
26 2 10 7
27 1 10 6
28 1 2 2
29 2 6 7;
```

```
run;
proc print data=save.example;
run;

data exa;
set save.example;
If group =1 then pdiff= placebo - drug;
Else if group =2 then pdiff = drug- placebo; /*taking period
difference*/
run;

/*Testing for the CROSS-OVER DIFFERENCE using paired t-test*/
Proc ttest;
Paired drug*placebo;
Run;

/*Testing for the PERIOD EFFECT using independent t-test*/
proc ttest;
class group;
var pdiff;
run;

/*Adjusting for the PERIOD EFFECT*/
proc ttest;
Class group;
var tdiff;
run;
```

The code above produced the same results as the manual calculation demonstrated in Section 2.A.5. The outputs of the three analyses are as following:

## SAS Output:

**The TTEST Procedure**

**Statistics**

| Difference | N | Lower CL Mean | Mean | Upper CL Mean | Lower CL Std Dev | Std Dev | Upper CL Std Dev | Std Err |
|---|---|---|---|---|---|---|---|---|
| drug - placebo | 29 | 0.9106 | 2.1724 | 3.4343 | 2.6326 | 3.3174 | 4.4866 | 0.616 |

**T-Tests**

| Difference | DF | t Value | Pr > \|t\| |
|---|---|---|---|
| drug - placebo | 28 | 3.53 | 0.0015 |

**The TTEST Procedure**

**Statistics**

| Variable | GROUP | N | Lower CL Mean | Mean | Upper CL Mean | Lower CL Std Dev | Std Dev | Upper CL Std Dev | Std Err |
|---|---|---|---|---|---|---|---|---|---|
| pdiff | 1 | 17 | -4.607 | -2.824 | -1.04 | 2.5831 | 3.4683 | 5.2786 | 0.8412 |
| pdiff | 2 | 12 | -0.649 | 1.25 | 3.1489 | 2.1171 | 2.9886 | 5.0743 | 0.8627 |
| pdiff | Diff (1-2) | | -6.612 | -4.074 | -1.535 | 2.5943 | 3.2814 | 4.4664 | 1.2372 |

T-Tests

| Variable | Method | Variances | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| pdiff | Pooled | Equal | 27 | -3.29 | **0.0028** |
| pdiff | Satterthwaite | Unequal | 25.8 | -3.38 | 0.0023 |

Equality of Variances

| Variable | Method | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| pdiff | Folded F | 16 | 11 | 1.35 | 0.6260 |

Statistics

| Variable | GROUP | N | Lower CL Mean | Mean | Upper CL Mean | Lower CL Std Dev | Std Dev | Upper CL Std Dev | Std Err |
|---|---|---|---|---|---|---|---|---|---|
| tdiff | 1 | 17 | 1.0403 | 2.8235 | 4.6068 | 2.5831 | 3.4683 | 5.2786 | 0.8412 |
| tdiff | 2 | 12 | -0.649 | 1.25 | 3.1489 | 2.1171 | 2.9886 | 5.0743 | 0.8627 |
| tdiff | Diff (1-2) | | -0.965 | 1.5735 | 4.1121 | 2.5943 | 3.2814 | 4.4664 | 1.2372 |

T-Tests

| Variable | Method | Variances | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| tdiff | Pooled | Equal | 27 | 1.27 | **0.2143** |

Equality of Variances

| Variable | Method | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| tdiff | Folded F | 16 | 11 | 1.35 | 0.6260 |

## 2.A.6 Analyse of cross-over trials using general linear models

As the t-test is a special case of the Analysis of Variance (ANOVA) which is a type of general linear model, identical results can also be derived using the general linear model shown in 2.A.1. When patients are fitted as fixed effect in the general linear models, the results derived from the general linear models are identical to the t-test. In order to fit general linear models for cross-over trials, the dataset was re-structured in a way that each measurement was treated as an observation, the treatment received for the measurement and the treatment period (first or second) were specified for each observation.

The SAS code for the general linear models without and without adjustment to the period effect are as following:

```
data exa;
/* Re-arrange the data structure*/
/*group =randomized sequence; 1=drug/placebo, 2=placebo/drug*/
/*treatment=the treatment the patient received during that
period*/
/*obs=measurement of the patient of that period*/
/*period=treatment period*/
input group id treat$ obs period;
cards;
1 1 drug 8 1
1 1 placebo 5 2
1 3 drug 14 1
1 3 placebo 10 2
1 4 drug 8 1
1 4 placebo 0 2
1 6 drug 9 1
1 6 placebo 7 2
1 7 drug 11 1
1 7 placebo 6 2
1 9 drug 3 1
1 9 placebo 5 2
1 11 drug 6 1
1 11 placebo 0 2
1 13 drug 0 1
1 13 placebo 0 2
1 16 drug 13 1
1 16 placebo 12 2
1 18 drug 10 1
1 18 placebo 2 2
```

```
1 19 drug 7 1
1 19 placebo 5 2
1 21 drug 13 1
1 21 placebo 13 2
1 22 drug 8 1
1 22 placebo 10 2
1 24 drug 7 1
1 24 placebo 7 2
1 25 drug 9 1
1 25 placebo 0 2
1 27 drug 10 1
1 27 placebo 6 2
1 28 drug 2 1
1 28 placebo 2 2
2 2 placebo 12 1
2 2 drug 11 2
2 5 placebo 6 1
2 5 drug 8 2
2 8 placebo 13 1
2 8 drug 9 2
2 10 placebo 8 1
2 10 drug 8 2
2 12 placebo 8 1
2 12 drug 9 2
2 14 placebo 4 1
2 14 drug 8 2
2 15 placebo 8 1
2 15 drug 14 2
2 17 placebo 2 1
2 17 drug 4 2
2 20 placebo 8 1
2 20 drug 13 2
2 23 placebo 9 1
2 23 drug 7 2
2 26 placebo 7 1
2 26 drug 10 2
2 29 placebo 7 1
2 29 drug 6 2
;
run;

/*Evaluating the treatment effect using general linear model*/
proc glm;
class treat id;
model obs= treat id;
run;

/*Test and adjust for the period effect using a general linear
model*/
proc glm;
class treat period id;
model obs=treat id period;
run;
```

## SAS output:

The GLM Procedure

Dependent Variable: obs

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 29 | 669.3275862 | 23.0802616 | 4.19 | 0.0001 |
| Error | 28 | 154.0689655 | 5.5024631 | | |
| Corrected Total | 57 | 823.3965517 | | | |

| R-Square | Coeff Var | Root MSE | obs Mean |
|---|---|---|---|
| 0.812886 | 31.86241 | 2.345733 | 7.362069 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| treat | 1 | 68.4310345 | 68.4310345 | 12.44 | **0.0015** |
| id | 28 | 600.8965517 | 21.4605911 | 3.90 | 0.0003 |

The GLM Procedure

Class Level Information

Dependent Variable: obs

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 30 | 678.0362576 | 22.6012086 | 4.20 | 0.0002 |
| Error | 27 | 145.3602941 | 5.3837146 | | |
| Corrected Total | 57 | 823.3965517 | | | |

| R-Square | Coeff Var | Root MSE | obs Mean |
|---|---|---|---|
| 0.823463 | 31.51673 | 2.320283 | 7.362069 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| treat | 1 | 68.4310345 | 68.4310345 | 12.71 | **0.0014** |
| id | 28 | 600.8965517 | 21.4605911 | 3.99 | 0.0003 |
| period | 1 | 8.7086714 | 8.7086714 | 1.62 | **0.2143** |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| treat | 1 | 58.3638438 | 58.3638438 | 10.84 | **0.0028** |
| id | 28 | 600.8965517 | 21.4605911 | 3.99 | 0.0003 |
| period | 1 | 8.7086714 | 8.7086714 | 1.62 | **0.2143** |

## 2.A.7 Patient Effect and missing data using mixed models

The methods shown in Sections 2.A.2 to 2.A.6 can only be applied to datasets which are balanced. Therefore patients who did not complete both treatment periods would be excluded from the analysis. As the trial period is often longer for a cross-over trial than a parallel trial, patients are more likely to drop out and consequently the dataset may contain some/more missing values.

A mixed model that contains both fixed and random effects (also called multi-level analysis in some areas of research such as sociology and epidemiology) can be used to deal with the missing values in cross-over design.

Fitting patient effect as a random effect utilises the information on patients who have not completed the trial. The information of patients who have not participated in the second period can be used to pool between-patients differences. The results of fixed or random effects analyses are identical when there are no missing values.

A statistical model that incorporates the patient effect as random can be denoted as:

$$y_{ijk} = \mu + \xi_{ij} + \pi_k + \varphi_l + \varepsilon_{ijk}$$

where

$\xi_{ij}$ = the effect of patient $j$ of the $i$-th sequence and $\xi_{ij} \sim N(0, \sigma_\xi^2)$

Definitions of all other terms are the same as the equation (2.A.1)

## 2.A.8 Statistical programming for the analysis of cross-over trials using mixed model

When there is no period effect, the SAS® code below shows the statistical model that contains the treatment effect and patient effect as random.

```
proc mixed;
class treat patient;
model y = treat;
random patient;
run;
```

The period effect can also be tested and adjusted for using a similar procedure, the SAS® code below shows the statistical model that contains the treatment effect, the period effect and patient effect as random.

```
proc mixed;
class treat patient;
model y = treat period;
random patient;
run;
```

## 2.B Testing for the carry-over differences and the two-stage procedure

This section describes Grizzle's [4] and Hills and Armitages' [7] methods to examine the carry-over effect, followed by two examples of applications of the two-stage procedure to cover both cases (carry-over effect significant and insignificant). With the actual type I error rate of 7-9.5%, the two-stage procedure is the recommended method for analysis of cross-over trials. The examples only serve the purpose of referencing the results found in the systematic review on published cross-over trial reports and the comprehensive review of general medical statistics book.

### 2.B.1 Grizzle's model

Testing for the carry-over effect was first introduced by Grizzle (1965). In addition to the direct drug effect (treatment) effect, the residual (carry-over) effect, the period effect, and the patient effect were also included in the model. The statistical model he suggested was as follow:

$$y_{ijk} = \mu + \xi_{ij} + \pi_k + \varphi_l + \lambda_l + \varepsilon_{ijk} \quad (2.B.1)$$

where

$y_{ijk}$ = the measurement of $j$-th patient in the $i$-th sequence at the $k$-th period.

$\mu$ = general mean

$\xi_{ij}$ = the effect of $j$-th patient, within the $i$-th sequence

$\pi_k$ = the effect of the $k$-th period

$\varphi_l$ = the direct effect of the drug $l$ [For patients randomised to the treatment sequence 1 ($i$=1), $l$ =A when $k$=1, $l$ =B when $k$=2. For patients randomised to the treatment sequence 2 ($i$=2), $l$ =B when $k$=1, $l$ =A when $k$=2]

$\lambda_l$ = the residual effect of the $l$-th drug and

$\varepsilon_{ijk}$ = the random fluctuation which is normally distributed with mean 0 and variance $\sigma_\varepsilon^2$, and is independent of $\xi_{ij}$

For a AB/BA trial, the parameters for each treatment period would be as follow:

**Table 2.B.2**

| | **Period 1** | **Period 2** |
|---|---|---|
| **Sequence 1** $(N = n_1)$ | $y_{1j1} = \mu + \xi_{1j} + \pi_1 + \varphi_a + \varepsilon_{1j1}$ | $y_{1j2} = \mu + \xi_{1j} + \pi_2 + \varphi_b + \lambda_a + \varepsilon_{1j2}$ |
| **Sequence 2** $(N = n_2)$ | $y_{2j1} = \mu + \xi_{2j} + \pi_1 + \varphi_b + \varepsilon_{2j1}$ | $y_{2j2} = \mu + \xi_{2j} + \pi_2 + \varphi_a + \lambda_b + \varepsilon_{2j2}$ |

The main objective of this paper was to present the unbiased estimates of contrasts of treatment effects and the carry-over effects. Based on the assumption that patient effects are the same (in expectation) for both sequences

$E\{\xi_1\} = E\{\xi_2\}$, the unbiased estimates of contrasts of the carry-over effect are $\bar{y}_{1,1} + \bar{y}_{1.2} - \bar{y}_{2,1} - \bar{y}_{2,2}$ and default estimates of contrast for treatment effects were set to be based on the period 1 data only $(\bar{y}_{1,1} - \bar{y}_{2,1})$. The method used to test the hypothesis of equal carry-over effect was the one-way Analysis of Variance (ANOVA).

Grizzle [4] suggested that 'If the residual effects of the two drugs are equal we can delete residual effects from the model.' This would

then permit one to estimate the treatment effect using within patient differences. The decision of whether to delete the residual effect should be based on a preliminary test of high significance level. He then set $\alpha = .10$ as 'Larson and Bancroft (1963) demonstrated that few serious errors would be made if the preliminary test were made at this level'.

The significance level of the residual effect also determined how the trial would be analysed. If the hypothesis that $\lambda_1 = \lambda_2$ using $\alpha = .10$ was accepted, then the analysis of the treatment would be based on the analysis of variance with residual effect omitted. If such a hypothesis was rejected at the 10% level, the analysis of treatment effects would inevitably fall back to the default estimates of contrast which would base on the data from period 1 only.

### 2.B.2 Hills and Armitages' (1979) approach to the carry-over effect

A very similar approach with different presentation was adopted by Hills and Armitage [7]. To test the hypothesis $\lambda_1 = \lambda_2$, they used the independent t-test to compare the means of $(y_{1j1} + y_{1j2})$ from patients randomised to sequence 1 with the means of $(y_{2j1} + y_{2j2})$ from patients randomised to sequence 2. The mathematics of the independent t-test are however, equivalent to the one-way ANOVA suggested by Grizzle [4].

Testing for the carry-over effect and testing for the treatment effect using either period 1 only or the differences between the two

periods are the three tests that form the two-stage procedure (or Hills and Armitage method). Three labels, CARRY, CROS and PAR are commonly used to denote these three tests.

CARRY indicates the test for the carry-over difference.

CROS indicates the test for the treatment effect using cross-over differences.

PAR indicates the test for treatment effect using first period only.

### 2.B.3 Previous Example

The dataset used in part A is used here to demonstrate Hill and Armitage's approach to the carry-over effect. An independent t-test is used to examine the difference between sums of the 2 periods of both sequences. The p value for the t-statistics of -1.30 at 27 degrees of freedom is 0.20. Therefore the carry-over effect is not significant at the 10% level. According to the two-stage procedure proposed by Hills and Armitage [7], the treatment would be analysed using the cross-over difference as illustrated in Section 2.A.5.1.

**Table 2.B.3**

| **Testing for the carry-over effect using independent t-test** | | |
|---|---|---|
| | **Sequence 1 (Drug/placebo) $n_1 = 17$** | **Sequence 2 (Placebo/drug) $n_2 = 12$** |
| Mean (of period total) | 13.41 | 16.58 |
| Standard Deviation | 7.32 | 4.98 |
| Correlated sums square | $= 16 \times 7.32^2 = 857.32$ | $= 11 \times 4.98^2 = 272.80$ |
| Pooled estimate of variance | $= (857.32 + 272.80) / (29-2) = 41.86$ | |
| Difference in Mean | $= 13.41 - 16.58 = -3.17$ | |
| Standard error of difference in means | $= \sqrt{\{41.98 \times (1/17 + 1/12)\}} = 2.44$ | |
| t statistics | $= -3.17/2.44 = -1.30$ | |
| $Pr > \|t\|$ | 0.20 | |

## 2.B.4 Modified dataset

In order to illustrate the method used to analyse the treatment effect when the carry-over effect is found to be significant, the data were manipulated with the purpose of showing significance in the carry-over effect at the 10% level.

The measurements of patients were modified so that the carry-over effect would be tested to be significant while the number of patients in each group and the sequence each patient was randomised to remain the same.

The modified dataset is as following:

**Table 2.B.4**

| Sequence 1 (drug/placebo) | | | |
|---|---|---|---|
| **Patient Number** | **Period 1 (drug)** | **Period 2 (placebo)** | **Sum** |
| **$n_1$=17** | **$Y_1$** | **$Y_2$** | **$Y_1 + Y_2$** |
| 1 | 8 | 15 | 23 |
| 3 | 14 | 16 | 30 |
| 4 | 10 | 17 | 27 |
| 6 | 9 | 12 | 21 |
| 7 | 11 | 16 | 27 |
| 9 | 13 | 5 | 18 |
| 11 | 16 | 8 | 24 |
| 13 | 0 | 2 | 2 |
| 16 | 13 | 16 | 29 |
| 18 | 10 | 12 | 22 |
| 19 | 7 | 15 | 22 |
| 21 | 13 | 14 | 27 |
| 22 | 8 | 10 | 18 |
| 24 | 7 | 8 | 15 |
| 25 | 9 | 12 | 21 |
| 27 | 10 | 12 | 22 |
| 28 | 2 | 3 | 5 |
| **Sum** | 160 | 193 | 353 |
| **Mean** | 9.41 | 11.35 | 20.76 |
| **Standard Deviation** | 4.07 | 4.70 | 7.66 |

| Sequence 2 (Placebo/drug) | | | |
|---|---|---|---|
| **Patient Number** | **Period 1 (placebo)** | **Period 2 (drug)** | **Sum** |
| **$n_2 = 12$** | **$Y_1$** | **$Y_2$** | **$Y_1 + Y_2$** |
| 2 | 6 | 5 | 11 |
| 5 | 6 | 8 | 14 |
| 8 | 3 | 9 | 12 |
| 10 | 8 | 8 | 16 |
| 12 | 8 | 9 | 17 |
| 14 | 4 | 8 | 12 |
| 15 | 8 | 14 | 22 |
| 17 | 2 | 4 | 6 |
| 20 | 5 | 13 | 18 |
| 23 | 6 | 7 | 13 |
| 26 | 7 | 10 | 17 |
| 29 | 7 | 8 | 15 |
| **Sum** | 70 | 103 | 173 |
| **Mean** | 5.83 | 8.58 | 14.41 |
| **Standard Deviation** | 1.99 | 2.84 | 4.08 |

### 2.B.4.1 Testing for the carry-over effect

An independent t-test is used on the sum of the two treatment periods from each sequence. The t statistics derived from the two sequences is 2.43 and it gives a p value of 0.01 at 27 degrees of freedom. The carry-over effect is significant at the 10% level. Therefore the treatment effect will be evaluated using the first period only.

**Table 2.B.5**

| Testing for the carry-over effect using independent t-test | | |
|---|---|---|
| | **Sequence 1 (Drug/placebo)** $n_1 = 17$ | **Sequence 2 (Placebo/drug)** $n_2 = 12$ |
| Mean (sum of 2 periods) | = 20.76 | = 14.41 |
| Standard Deviation | = 7.66 | = 4.08 |
| Correlated sums square | $=16 \times 7.66^2 = 938.81$ | $= 11 \times 4.08^2 = 183.11$ |
| Difference in Mean | = 20.76 – 14.41 = 6.35 | |
| Pooled estimate of variance | =(938.81+183.11)/(29-2) = 41.55 | |
| Standard error of difference in means | $= \sqrt{\{41.55 \times (1/17 + 1/12)\}} = 2.43$ | |
| t statistics | = 6.35 / 2.43 = 2.61 | |
| $Pr > \|t\|$ | 0.01 | |

## 2.B.4.2 Examining the treatment effect using first period data only

To examine the treatment effect, an independent t-test is used on the first period data. Patients in sequence 1 received drug during period one while patients in sequence 2 received the placebo. The t-statistics of 2.82 at 27 is significant at 1% level.

**Table 2.B.6**

| Examining the treatment effect using first period data only by independent t-test | | |
|---|---|---|
| | **Sequence 1 (Drug/placebo)** $n_1 = 17$ | **Sequence 2 (Placebo/drug)** $n_2 = 12$ |
| Mean (of period 1) | = 9.41 | = 5.83 |
| Standard Deviation | = 4.07 | = 1.99 |
| Correlated sums square | = 16 x $4.07^2$ = 265.04 | = 11 x $1.99^2$ = 43.56 |
| Difference in Mean | = 9.41 – 5.83 = 3.58 | |
| Pooled estimate of variance | = ( 265.04 + 43.56 ) / ( 29 - 2 ) = 11.43 | |
| Standard error of difference in means | = $\sqrt{\{11.43 \times (1/17 + 1/12)\}}$ = 1.27 | |
| t statistics | = 3.58 / 1.27 = 2.82 | |
| $Pr > \|t\|$ | 0.0093 | |

## 2.B.4.3 SAS code

The following are the SAS code and outputs for the analyses in Sections 2.B.4.1 and 2.B.4.2.

```
libname save 'c:\';
data save.example;
input id group drug placebo;
Datalines;
1 1 8 15
2 2 5 6
3 1 14 16
4 1 10 17
5 2 8 6
6 1 9 12
7 1 11 16
8 2 9 3
9 1 13 5
10 2 8 8
11 1 16 8
12 2 9 8
13 1 0 2
14 2 8 4
15 2 14 8
16 1 13 16
17 2 4 2
18 1 10 12
```

```
19 1 7 15
20 2 13 5
21 1 13 14
22 1 8 10
23 2 7 6
24 1 7 8
25 1 9 12
26 2 10 7
27 1 10 12
28 1 2 3
29 2 8 7
;
run;

data exa;
set save.example;
If group =1 then pdiff= placebo - drug;
Else if group =2 then pdiff = drug- placebo; /*taking period
difference*/
tdiff=drug-placebo; /*taking treatment difference*/
carry=drug+placebo; /*taking the sum of the 2 periods*/
If group =1 then period1 = drug; /*specify the period 1 data*/
Else if group =2 then period1 = placebo;
run;

proc ttest; /*testing for the carry-over difference*/
class group;
run;

proc ttest; /*testing for the treatment difference using period
data only*/
class group;
var period1;
run;
```

## SAS output

The TTEST Procedure

Statistics

| Variable | GROUP | N | Lower CL Mean | Mean | Upper CL Mean | Lower CL Std Dev | Std Dev | Upper CL Std Dev | Std Err |
|---|---|---|---|---|---|---|---|---|---|
| carry | 1 | 17 | 16.826 | 20.765 | 24.704 | 5.7057 | 7.661 | 11.66 | 1.8581 |
| carry | 2 | 12 | 11.826 | 14.417 | 17.008 | 2.8887 | 4.0778 | 6.9237 | 1.1772 |
| carry | Diff (1-2) | | 1.3611 | 6.348 | 11.335 | 5.0966 | 6.4463 | 8.7743 | 2.4305 |

T-Tests

| Variable | Method | Variances | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| carry | Pooled | Equal | 27 | 2.61 | **0.0145** |

Equality of Variances

| Variable | Method | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| carry | Folded F | 16 | 11 | 3.53 | 0.0393 | |

The TTEST Procedure

Statistics

| Variable | GROUP | N | Lower CL Mean | Mean | Upper CL Mean | Lower CL Std Dev | Std Dev | Upper CL Std Dev | Std Err |
|---|---|---|---|---|---|---|---|---|---|
| period1 | | 17 | 7.3149 | 9.4118 | 11.509 | 3.0374 | 4.0783 | 6.2068 | 0.9891 |
| | 1 | | | | | | | | |
| period1 | | 12 | 4.5674 | 5.8333 | 7.0993 | 1.4114 | 1.9924 | 3.3829 | 0.5752 |
| | 2 | | | | | | | | |
| period1 | Diff (1-2) | | | 0.958 | 3.5784 | 6.1989 | 2.678 | 3.3873 | 4.6105 |
| | 1.2771 | | | | | | | | |

T-Tests

| Variable | Method | Variances | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| period1 | Pooled | Equal | 27 | 2.80 | **0.0093** |

Equality of Variances

| Variable | Method | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| period1 | Folded F | 16 | 11 | 4.19 | 0.0204 |

## 2.C Performance of the two-stage procedure

This section presents a literature review on the performance of the two-stage procedure, which has been discussed in terms of its power and the type I error rate.

The first paper that discussed the type I error rate of the two-stage procedure was Freeman [8]. It was presented by the coverage probability which is 1-type I error rate. One assumption made in this paper was to eliminate the between patient variations. The data were reduced to the within-patient variation only by subtracting the baseline value at the beginning of the experiment. With the same notations as the statistical model described in equation 2.A.1, the parameters and methods used by Freeman [8] are as follow:

$$s_{ij} = y_{ij1} + y_{ij2} - 2y_{ij0}, \qquad f_{ij} = y_{ij1} - y_{ij0}, \qquad d_{ij} = y_{ij1} - y_{ij2}$$

$$\overline{s_i} = \sum_j \frac{s_{ij}}{n}, \qquad \overline{f_i} = \sum_j \frac{f_{ij}}{n}, \qquad \overline{d_i} = \sum_j \frac{d_{ij}}{n}$$

$$S = \overline{s_1} - \overline{s_2}, \qquad F = \overline{f_1} - \overline{f_2}, \qquad D = \frac{1}{2}\left(\overline{d_1} - \overline{d_2}\right)$$

and $F = \frac{1}{2}S + D$

The sampling distributions of $S$, $F$ and $D$ are:

$$S \sim N\left(\lambda, \frac{12\sigma^2}{n}\right), \qquad F \sim N\left(\phi, \frac{4\sigma^2}{n}\right), \qquad D \sim N\left(\phi - \frac{\lambda}{2}, \frac{\sigma^2}{n}\right)$$

The correlation coefficient between two distributions with variances $S_1^2$ and $S_2^2$ is= $\frac{1}{\sqrt{1+\left(\frac{S_1^2}{S_2^2}\right)}}$

The correlation between $F$ and $S$ is therefore

$$\frac{1}{\sqrt{1+\frac{\frac{4\sigma^2}{n}}{\frac{12\sigma^2}{n}}}}=\frac{1}{\sqrt{1+\frac{1}{3}}}=\frac{\sqrt{3}}{2}\approx 0.866$$

Coverage probability (=1-type I error rate) of the two-stage procedure:

$$CP_{ts}=prob\left(\left|\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right|<Z_{\alpha_1}\ \ and\ \ \left|\frac{(D-\phi)\sqrt{n}}{\sigma}\right|<Z_{\alpha}\right)+prob\left(\left|\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right|>Z_{\alpha_1}\ \ and\ \ \left|\frac{(F-\phi)\sqrt{n}}{2\sigma}\right|<Z_{\alpha}\right)$$

(2.C.1)

Since CARRY was set at 10%, and CROS and PAR were both set at 5% $\alpha_1=0.10$, $\alpha=0.05\Rightarrow Z_{\alpha_1}=1.645$, $Z_{\alpha}=1.96$

Since $F$ and $S$ are normally distributed and correlated, the second term can be expressed as a Bivariate normal distribution. The probability under the Bivariate normal distribution can be evaluated using a double integration:

$$P(U\le u,V\le v;\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty-\infty}^{v}\int^{u}\exp\left[-\frac{(U^2-2\rho UV+V^2)}{2(1-\upsilon^2)}\right]dUdV$$

where $\rho$ is the correlation between the two distributions

Therefore

$$prob\left(\left|\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right| > Z_{\alpha_1} \text{ and } \left|\frac{(F-\phi)\sqrt{n}}{2\sigma}\right| < Z_{\alpha}\right)$$

$$= prob\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}} > Z_{\alpha_1} \text{ and } -Z_{\alpha} < \frac{(F-\phi)\sqrt{n}}{2\sigma} < Z_{\alpha}\right) +$$

$$prob\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}} < -Z_{\alpha_1} \text{ and } -Z_{\alpha} < \frac{(F-\phi)\sqrt{n}}{2\sigma} < Z_{\alpha}\right)$$

$$= \frac{1}{2\pi\sqrt{1-\sigma^2}} \int_{Z_{\alpha_1}}^{\infty} \int_{-Z_{\alpha}}^{Z_{\alpha}} \exp\left[-\frac{\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right)^2 - 2\times 0.866\times\frac{S}{\sigma}\sqrt{\frac{n}{12}}\times\frac{(F-\phi)\sqrt{n}}{2\sigma} + \left(\frac{(F-\phi)\sqrt{n}}{2\sigma}\right)^2}{2\times(1-0.866^2)}\right]$$

$$+ \frac{1}{2\pi\sqrt{1-\sigma^2}} \int_{-\infty}^{-Z_{\alpha_1}} \int_{-Z_{\alpha}}^{Z_{\alpha}} \exp\left[-\frac{\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right)^2 - 2\times 0.866\times\frac{S}{\sigma}\sqrt{\frac{n}{12}}\times\frac{(F-\phi)\sqrt{n}}{2\sigma} + \left(\frac{(F-\phi)\sqrt{n}}{2\sigma}\right)^2}{2\times(1-0.866^2)}\right]$$

Freeman [8] showed that the coverage probability was only 0.917 when there is no carryover ($\lambda = 0$) and the coverage probability drops to the lowest point (0.56) when $\lambda\left(\sqrt{n}\right)/\sigma$ is around 5. The calculations of these two examples are as following:

**Case 1:** $\lambda = 0$

$$prob\left(\left|\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right| > Z_{\alpha_1} \textbf{ and } \left|\frac{(D-\phi)\sqrt{n}}{2\sigma}\right| < Z_{\alpha}\right) + prob\left(\left|\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right| > Z_{\alpha_1} \textbf{ and } \left|\frac{(F-\phi)\sqrt{n}}{2\sigma}\right| < Z_{\alpha}\right)$$

$$= 0.90\times 0.95 + prob\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}} > 1.645 \textbf{ and } -1.96 < \frac{(F-\phi)\sqrt{n}}{2\sigma} < 1.96\right) +$$

$$prob\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}} < -1.645 \textbf{ and } -1.96 < \frac{(F-\phi)\sqrt{n}}{2\sigma} < 1.96\right)$$

$$= 0.8550 + 0.031 + 0.031^{A}$$

$$= 0.917$$

The double integral used to evaluate the probability under the Bivariate distribution was evaluated using a Maple program as shown in Figure A.

Maple Code: Bivariate Normal Distribution Function

```
>f:=(x,y,rho)->(1/(2*Pi*sqrt(1-rho^2)))*exp((-1/(2*(1-
rho^2)))*(x^2+y^2-2*rho*x*y));
```

$$f := (x, y, rho) \rightarrow \frac{e^{\left(-\frac{x^2+y^2-2\rho x y}{2-2\rho^2}\right)}}{2\pi\sqrt{1-\rho^2}}$$

```
>int(int(f(x,y,0.866),x=-infinity..-1.645),y=-1.96..1.96);
```

0.03147187089

```
> int(int(f(x,y,0.866),x=1.645..infinity),y=-1.96..1.96);
```

0.03147187089

The coverage probability of the two-stage procedure while there is no carryover is therefore 0.8550+0.062=0.917

**Case 2:** $\frac{\lambda\sqrt{n}}{\sigma} = 5$

The first part of the equation 2C is evaluated as the follows.

$$prob\left(\left|\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right| < Z_{\alpha_1} \textbf{ and } \left|\frac{(D-\phi)\sqrt{n}}{\sigma}\right| < Z_{\alpha}\right)$$

$$= prob\left(Z_{\alpha_1} < \frac{S}{\sigma}\sqrt{\frac{n}{12}} < Z_{\alpha_1}\right) \times prob\left(Z_{\alpha} < \frac{(D-\phi)\sqrt{n}}{\sigma} < Z_{\alpha}\right)$$

$$= \left[\Phi\left(Z_{\alpha_1} - \frac{\lambda}{\sigma}\sqrt{\frac{n}{12}}\right) - \Phi\left(-Z_{\alpha_1} - \frac{\lambda}{\sigma}\sqrt{\frac{n}{12}}\right)\right] \times \left[\Phi\left(Z_{\alpha} + \frac{1}{2}\frac{\lambda\sqrt{n}}{\sigma}\right) - \Phi\left(-Z_{\alpha} + \frac{1}{2}\frac{\lambda\sqrt{n}}{\sigma}\right)\right]$$

$$= \left[\Phi\left(1.645 - \frac{5}{\sqrt{12}}\right) - \Phi\left(-1.645 - \frac{5}{\sqrt{12}}\right)\right] \times \left[\Phi\left(1.96 + \frac{1}{2}\times 5\right) - \Phi\left(-1.96 + \frac{1}{2}\times 5\right)\right]$$

$$= (0.5798 - 0.001) \times (1 - 0.7054)$$

$$= 0.5788 \times 0.2946$$

$$= 0.1705$$

For the second part of the equation 2C, the value of the double integral was calculated using a Maple program as shown in Figure B.

Let $S'$ be a normal distribution.

Let $D'$ be a normal distribution.

Let $F'$ be a normal distribution.

$$prob\left(\left|\frac{S}{\sigma}\sqrt{\frac{n}{12}}\right| > Z_{\alpha_1} \text{ and } \left|\frac{(F-\phi)\sqrt{n}}{2\sigma}\right| < Z_{\varepsilon}\right)$$

$$= prob\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}} > Z_{\alpha_1} \text{ and } -Z_{\alpha} < \frac{(F-\phi)\sqrt{n}}{2\sigma} < Z_{\alpha}\right) + prob\left(\frac{S}{\sigma}\sqrt{\frac{n}{12}} < -Z_{\alpha_1} \text{ and } -Z_{\alpha} < \frac{(F-\phi)\sqrt{n}}{2\sigma} < Z_{\alpha}\right)$$

$$= prob(S' > 1.645 + 1.443 \textbf{ and } -1.96 - 2.5 < D' < 1.96 - 2.5) +$$

$$prob(S' < -1.645 + 1.443 \textbf{ and } -1.96 < F' < 1.96)$$

$$= prob(S' > 3.088 \textbf{ and } -4.46 < D' < -0.54) +$$

$$prob(S' < -0.202 \textbf{ and } -1.96 < F' < 1.96)$$

$$= 1.480 \times 10^{-14} + 0.3951$$

$$\approx 0.3951$$

Maple Code: Bivariate Normal Distribution Function

```
> f:=(x,y,rho)->(1/(2*Pi*sqrt(1-rho^2)))*exp((-1/(2*(1-
rho^2)))*(x^2+y^2-2*rho*x*y));
```

$$f := (x, y, rho) \rightarrow \frac{e^{\left(-\frac{x^2+y^2-2\rho xy}{2-2\rho^2}\right)}}{2\pi\sqrt{1-\rho^2}}$$

```
> int(int(f(x,y,0.866),x=3.08837567297406..infinity),y=-4.46..-0.54);
```

$$1.480240143\ 10^{-14}$$

```
> int(int(f(x,y,0.866),x=-infinity..-0.20162432702594),y=-1.96..1.96);
```

$$0.3951169099$$

The coverage probability of the two-stage procedure while $\frac{\lambda\sqrt{n}}{\sigma} = 5$

is therefore 0.1705 +0.3951=0.56

The conclusion of Freeman [8] was that the two-stage procedure

is slightly misleading even when there is no carryover whatsoever, and potentially very misleading when there is even a small differential carryover effect.

The deficiencies of the two-stage procedure in comparison with CROS and PAR were also discussed by Jones and Lewis [45] and Grieve and Senn [46].

Jones and Lewis [45] used a simulation to demonstrate that the power and bias of the two-stage procedure were only slightly misleading under a moderate level of carry-over effect.

The assumptions they used for the simulation were:

Treatment effect $\phi = 0$

Within-patient variance $\sigma_w^2 = 48$

Between-patient variance $\sigma_b^2 = 96$

Number of patients in both groups $n_1 = n_2 = 22$

Although Jones and Lewis [45] considered that using the CROS procedure under the assumption that no carry-over effect is the best strategy, they showed that the power of the two-stage procedure is only 10% below the target level when $\lambda$ (the carry-over effect) is less than 1.5. They also showed that the CROS procedure has a higher bias level than the two-stage procedure.

Grieve and Senn [46] considered the false-positive rate is the more serious deficiency of the two-stage procedure by calculating the p-values of PAR and CROS conditional on the value of CARRY. They

explained the reason that the two-stage procedure appeared to be less problematic in Jones and Lewis [45] was that CROS was used most of the time. They showed that the PAR procedure is also vulnerable to the CARRY as

$$PAR = CROS + CARRY/2$$

$$E[PAR \mid CARRY] = E[CROS] = \phi - \frac{\lambda}{2}$$

and $E[PAR \mid CARRY] - \phi = (CARRY - \lambda)/2$

which lead to $E[PAR \mid CARRY] = \phi - \frac{\lambda}{2} + \frac{CARRY}{2}$

If the two-stage procedure is used, large absolute values of CARRY would lead one to use the PAR procedure. When absolute values of CARRY are low, the two-stage procedure would be very similar to CROS. However when PAR is used, it is subject to high level of bias. Grieve and Senn [46] suggested that the two-stage procedure should be examined especially when it differs from the CROS. They showed that the p-value of the treatment effect is only reduced to the claimed 5% level or lower when the p-value of the CARRY test is around 15% or above. However, the PAR procedure would not have been used when the p-value is 10% or higher. In other words, when the PAR procedure is used, it is usually very misleading. Therefore the two-stage procedure would perform extremely badly when p-value of the CARRY test is low (when PAR is used). They therefore concluded that power comparisons are misleading.

One way of correcting the size of the two-stage procedure to the level that was originally claimed, was suggested by Senn [22]. He showed that by setting PAR at 0.5%, the size of the two-stage procedure would be between 4.75% and 5%. However, such an adjustment is not recommended as the two-stage procedure loses its advantage in power.

# Chapter 3 - Methodology

## 3.1. Introduction

Applications of and references to the two-stage procedure can easily be found in various domains of medical statistics where cross-over trials are introduced or analysed. The four main areas of medical statistics that involve analyses of cross-over trials are medical journals, statistical journals, medical statistics books and the pharmaceutical industry. These four domains can be categorised into two: applications of, and references to, statistics. Applications of analyses of cross-over trials are found in medical journals and the pharmaceutical industry. Trials reported in medical journals and analysed in the pharmaceutical industry are usually mutually exclusive and present different phases of clinical trials. References to analyses of cross-over trials can be found in general medical statistics books and statistical journals. Medical statistics books provide references to basic analyses for clinicians or health practitioners while statistical journals publish statistical methods at a more advanced level.

Although the methods used in each area of medical statistics may have an impact on other areas, the contexts are mutually exclusive. Therefore, research on references and actual analyses of cross-over trials will be conducted in all four areas using appropriate research methods for each area of medical statistics.

## 3.2. Research methods

### 3.2.1. Periodicals: All subjects

All three major articles (Grizzle 1965 [4], Hills and Armitage 1979 [7] and Freeman 1989 [8]) about the two-stage procedure were published in periodicals (statistical and medical journals). Although the main issue with the method is its validity in theory rather than difficulties in applications, the development of the two-stage procedure has sparked some discussion in both types of journals. Applications of the two-stage procedure are usually found in medical journals and discussions of such a statistical procedure are usually found in statistical journals. This part of the thesis uses the number of citations given to the 3 publications of the two-stage procedure to quantify the influence of the two-stage procedure and the finding of its deficiencies in journals covering both applications and methodologies of the different subjects.

**Objective**: The main objective of this study is to highlight the seriousness of the issue by quantifying the overall impact of the two-stage procedure and its deficiencies on both theories and applications of analyses of cross-over trials.

**Research method**: Citation analysis.

Citation analysis is a bibliometric method that uses reference citations found in scientific papers as the primary analytical tool [28].

**Strengths:** The main strength of the citation analysis is that all papers that contained any reference to the two-stage procedure (regardless of the subject of journals) are included in the study. The impact of the two-stage procedure and its major deficiency in different fields can be discussed in terms of their fields of research.

**Weaknesses:** Citation analysis is based on the assumption that citations indicate use which is in fact not confirmed. Furthermore, it cannot be established whether the absence of citations of the two-stage procedure implies that it is not in use. Therefore, the study may not be able to give a precise estimate of what percentage of cross-over trials have been analysed with the two-stage procedure. In order to overcome this weakness, a systematic review is made of a number of randomly selected cross-over trial reports published in medical journals.

### 3.2.2. Cross-over trials published in medical journals

The term "medical journals' is used here to mean all periodicals that publish medical advances. These are published at regular intervals which may be weekly, fortnightly, monthly, bi-monthly or quarterly. Their target readers are medical professionals. Clinical trials are one of the most common topics in medical journals. The typical cross-over trial published in medical journals is conducted in a hospital (often a teaching hospital) with the objective of comparing an existing popular treatment with a new one. Results of such cross-over trials can often affect clinicians' choices of treatments.

In addition to the recruitment of patients and assigning of treatments, the physician who is in charge of the trial is often also responsible for designing and analysing it. There might be some occasional consultation with statisticians if resources permit. Otherwise the most common references used for statistical analyses are medical statistics books and journals. Some description of the statistical method used to analyse the trial is usually part of the report, as it is required by most medical journals. This systematic review uses the descriptions of statistics in cross-over trial reports to study methods used for analysing cross-over trials in medical journals.

**Objective:** The objective of this study is to investigate actual analyses of cross-over trials published in medical journals, including impact for the validity of the two-stage procedure and the change in methods used to analyse cross-over trials across time.

**Research method:** Systematic review.

Systematic review is a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyse and summarise the results of the included studies [35].

**Relation to other realms of medical statistics**: Results of the systematic review on analyses of cross-over trials in the medical literature will be compared with methods used in the

pharmaceutical industry to illustrate differences (or similarities) in analyses by researchers of different professions.

Results of the systematic review of cross-over trials published in medical journals will also be compared with recommendations for analyses of cross-over trials given in medical statistics books. Results of such comparisons will give an indication of the impact of general medical statistics books on actual analyses published in medical journals.

**Strengths:** The systematic review allows a close examination of statistical methods used for cross-over trials published in medical journals. By comparing the description of the statistics used in the trial report with references cited, this method also complements the shortage of the citation analysis by establishing the relationship between citations and actual applications.

**Weaknesses:** There are two main weaknesses of this method. The first is that the number of reports reviewed in the systematic review is limited. Due to time limitation, it is not feasible to review every cross-over trial report. The second weakness is that information given in medical journals regarding to statistical methods used can be inadequate [32].

### 3.2.3. General medical statistics books

General medical statistics books are one of the most commonly used references of statistics for non-statisticians. Such books usually include basic statistical concepts and tests for simple

analysis. Methods for the analysis of cross-over trials are sometimes covered. Their target readers are usually clinicians with limited statistics backgrounds and medical students. Because cross-over trials are often reported by the former population in medical journals, it is hypothesised that statistical methods used to analyse cross-over trials are to some extent influenced by recommendations given in general medical statistics books.

Methods of analyses of cross-over trials commonly used before the two-stage procedure was introduced were simple and straightforward, often in the form of matched paired t-tests. Although the two-stage procedure involves an additional preliminary t-test, the statistics involved are similar. The difference between the statistical methods used to analyse a cross-over trial with and without testing the carry-over effect (the two-stage procedure) is the type of t-test being used. For a simple cross-over difference, the t-test used is one sample, while the two-sample t-test is used to test the carry-over effect and to evaluate treatment effect using the first period only (when the hypothesis that there is no differential carry-over effects is rejected). As the statistical methods recommended in the general medical statistics books rarely involve complicated and specific topics, it is partly the objective of this thesis to discover how general medical statistics books have dealt with the more complicated side of the two-stage procedure: the low power and high type I error rate.

As the statistics required to understand the concept are beyond the scope of general medical statistics books, the absence of mentions of the deficiencies of the two-stage procedure is conceivable. On the other hand, it cannot be verified whether the void is due to authors' concern for not confusing the reader or because authors are not up to date with the change in validity of the two-stage procedure. Assuming that the author is aware and agrees with the deficiencies of the two-stage procedure, he/she still has the options of dismissing the two-stage procedure altogether or mentioning the two-stage procedure and explaining why it should not be used anymore.

**Objective:** The objective of this study is to investigate statistical methods recommended for analysing cross-over trials in general medical statistics books and to examine how up to date the general medical statistics books are in the analysis of cross-over trials.

**Research method:** Comprehensive review.

**Strengths:** The comprehensive review includes all general medical statistics books published in English. Therefore, all recommendations of statistical methods to which clinical researchers may be exposed are captured.

**Weaknesses**: The relationship between actual analyses of cross-over trials and the general medical statistics books is not confirmed. The extent of the influence of these books is unknown.

### 3.2.4. Pharmaceutical industry

The pharmaceutical industry conducts hundreds of clinical trials each year and their results often have direct influence on whether a new treatment will be licensed and brought out to the general public. In contrast to clinical trials published in medical journals, which are usually conducted and analysed by clinicians, trials conducted by pharmaceutical companies are designed, planned and analysed by statisticians. Results of such trials are rarely published in medical journals and the only way to gain information on such methods and statisticians' views on analyses of cross-over trials is by contacting the companies directly. A questionnaire survey is considered the most appropriate method for this purpose as it can be designed to target the objective of the research which includes questions that cannot be answered by any other research method.

Due to the professional background of those who are responsible for the analyses of cross-over trials, it is assumed that statistical methods used by the pharmaceutical industry provide higher levels of statistical analyses than cross-over trials published in medical journals. Therefore, results drawn from the questionnaire survey amongst the pharmaceutical statisticians should not be used to generalise to other realms of medical statistics.

**Objective:** There are two main objectives of this study. First, to gain some insights about statistical methods used to analyse cross-over trials by the pharmaceutical industry. Second, to understand pharmaceutical statisticians' rationales behind the choice of

statistical methods used to analyse cross-over trials as well as their views on the two deficiencies of the two-stage procedure.

**Research method:** Questionnaire survey.

**Strengths:** Gaining insights on methods and attitudes towards design and analyses of cross-over trials which can not be obtained elsewhere.

**Weaknesses:** Results are not representative if the response rate is low. To maximise the response rate, follow-up emails/letters are sent a few weeks after the first round of questionnaire is distributed.

## 3.3. Plan of this thesis

Chapters 4 to 7 summarise methodologies and the results of the four independent studies. Conclusions of results from all four studies will be presented in chapter 7 to give a complete picture of analyses of cross-over trials.

To give an overall indication of the influence and impact of the two-stage procedure on analyses of cross-over trials, the citation analysis (Chapter 4 - Citation analysis on three major papers about the two-stage procedure) is the first study to be reported in this thesis. It also acts as a confirmation of the necessity of investigating the use of the two-stage procedure for analyses of cross-over trials and to determine the importance of the three main publications of the two-stage procedure.

Following the citation analysis is the systematic review on published cross-over trials (Chapter 5 - Cross-over trials in medical journals – a systematic review) in medical journals. This includes cross-over trial reports which did or did not cite any of the two-stage procedure papers. The focus is on statistical methods used to analyse cross-over trials.

To track references and sources of influences on analyses of cross-over trials published in medical journals, the systematic review is followed by a comprehensive review of general medical statistics books (Chapter 6 - Analyses of cross-over trials in general medical statistics books – a comprehensive review).

Finally, a questionnaire survey of the pharmaceutical industry (Chapter 7 - Analysis of cross-over trials in the pharmaceutical industry - A questionnaire survey) is reported in order to fill the gap between cross-over trials conducted and published by different research populations.

# Chapter 4 - Citation analysis on three major papers about the two-stage procedure

## 4.1. Introduction

The first research method and domain of medical statistics used to investigate the two-stage procedure of cross-over trials is the citation analysis on published periodicals. Despite the problems of low power and high type I error rate, the two-stage procedure can still be seen in recent medical literature and is recommended by general medical statistics books without mentioning the statistical implications of applying it. The objective of this chapter is to establish the impact of the two-stage procedure on published research papers in terms of quantities of citations they received. It also serves the purpose of illustrating the background of this thesis.

## 4.2. Objective

The objective of this study is to evaluate the impact of the three major papers about the two-stage procedure (Grizzle 1965 [4], Hills and Armitage 1979 [7] and Freeman 1989 [8]) on both the theoretical and actual analysis of cross-over trials using citation analysis.

## 4.3. Methodology

Citation analysis is a bibliometric method that uses reference citations found in scientific papers as the primary analytical tool

[28]. It has been used to quantify the influences of research articles on scientific development [23]. Citation analysis has also been used to search for fundamental articles and identify key contributors in many research areas, including biomedicine, economics, information systems, computing and chemistry.

This chapter is the first report of a citation analysis which follows the development of a statistical application. It contains a detailed bibliometric study of the major publications of the two-stage procedure regarding citation rates, journal types and citation types.

### 4.3.1. The Web of Science

Data for the citation analysis of the two-stage procedure was generated on the Science Citation Index expanded on the Web of Science [40] provided by the Institute of Science (ISI) which is hosted and supported by the Manchester Information and Associated Services (MIMAS). The Science Citation Index expanded is a source database of articles published in journals and contains all the end-of-article citations in papers published by more than 5,700 scientific journals across 164 scientific disciplines. The database includes journal titles of all research fields in which cross-over trials are applied. With an average of 17,750 new records added every week, results of this chapter were based on a search performed in February 2002.

It was learnt that an introduction to MIMAS bibliographic services including 'cited reference search' was offered by MIMAS after the

citation analysis was completed. Therefore instead of learning the method of citation search, the purpose of attending the one day course was to confirm that the search on the Web of Science for the citation analysis had been conducted properly.

The course took place in the University of London Computer Centre on the 19th of March 2002. Strategies used for searching citing papers on the three key papers on the two-stage procedure were consulted with the Web of Science Support officers on the day of the course. Thereafter, several emails regarding further details and developments of the Web of Science were exchanged with the MIMAS helpdesk before the citation rates for the analysis were finalised.

### 4.3.2. The search strategy

The three most important publications which marked the history of the two-stage procedure in analyses of two-period two-treatment cross-over trials are: Grizzle (1965) [4], Hills and Armitage (1979) [7] and Freeman (1989) [8] (as described in Sections 1.3.1. 1.3.2. ). The database was searched for papers citing any of these three papers. The results searched on the Web of Science gave details of the citing paper, including the title, author, publishing journals, year of publication, and abstract in most cases.

The analyses carried out on these bibliographic data included (a) citation rates of each paper over time, (b) journal types in which citing papers are published (c) the citation type and (d) the

development of statistical models to compare the differences between the three cited papers.

## 4.4. Results

Results are first presented by citation rates of each individual paper, followed by the characteristics of citations including citation years, citation types and types of citing journals, in order to investigate areas of impact of each paper and to seek explanations for these impacts.

### 4.4.1. Citation rates

The two papers describing the two-stage procedure were:

**(i) Grizzle (1965)[4], The two-period change-over design and its use in clinical trials, *Biometrics***

**(ii) Hills and Armitage (1979)[7], The Two-Period Cross-Over Clinical Trial, *British Journal of Clinical Pharmacology***

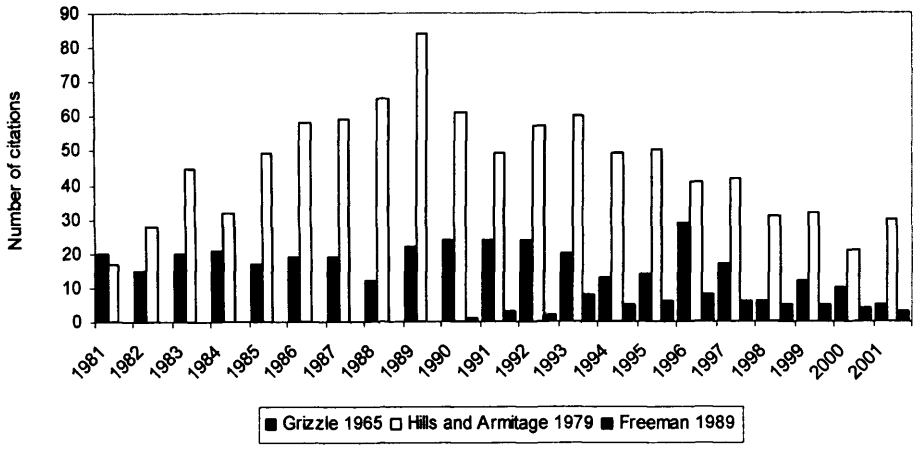
The two papers describing the two-stage procedure together received 1328 citations over the period of 22 years from 1981 to February 2002. Hills and Armitage (1979)[7] contributed 961 citations while Grizzle (1965)[4] accounted for the other 367. On average for the past two decades, there have been around 60 citations on the two-stage procedure every year, with about 44 citations on Hills and Armitage (1979)[7] and 17 on Grizzle (1965)[4].

The significance of these citation rates were weighed against Garfield's conclusion in 1971 that only 1% of all published items were cited more than 15 times a year [Garfied, Current Contents, #18, 1971]. The above figures indicate that both Grizzle (1965)[4] and Hills and Armitage (1979)[7] on the two-stage procedures were among the top 1 per cent of all cited publications.

High citation rates of the two-stage procedure also made these two papers citation 'classics'. Garfield [43] suggested that *'in general, a publication cited more than 400 times should be considered a classic; but in some fields with fewer researchers, 100 citations might qualify a work'*. Hills and Armitage (1979)[7] qualifies all for a 'citation classic' by all standards while Grizzle (1965)[4] would be considered a citation classic if analysis of cross-over trials is considered to be a less researched discipline.

Changes in citation rates on Grizzle [4](1965) and Hills and Armitage [7](1979) over the years have not been substantial. The publication year of each citing article was recorded so as to plot the frequency of citation over time (Figure 4.4-1). Although Grizzle [4](1965) was the original paper of the two-stage procedure, citation rates on Hills and Armitage [7](1979) soon superseded Grizzle [4](1965) two years after it was published. Hills and Armitage [7](1979) had received more citations than any other papers on analyses of cross-over trials ever since.

**Figure 4.4-1 Citation rate by year and author (1981-2002)**




**Freeman (1989), The performance of the two-stage analysis of two-treatment, two-period crossover trials. Statistics in Medicine**

In contrast to the two papers of the two-stage procedure, Freeman [8](1989), the latest development on the two-stage procedure which showed the major deficiency of the two-stage procedure, had noticeably lower citation rates. This paper was cited only 56 times for the 12 year period from 1990 to 2001. Although it could be qualified as a 'classic' when total citations over the years are averaged, there were a few years when Freeman [8](1989) was cited less than four times.

The results of the search showed that none of these publications have generated as much impact as Grizzle [4](1965) and Hills and Armitage [7](1979) in terms of the citation rates. It can therefore be concluded that these two papers dominated the citations on the analysis of cross-over trials and were appropriate to be used to

indicate the use of the two-stage procedure in the medical literature.

## 4.5. Discussion

Despite the similarity in context, the differences in citation rates of Grizzle [4](1965) and Hills and Armitage [7](1979) have been substantial. Citation rates are often influenced by a few factors. It has been suggested that "*papers containing important ideas will not necessarily continue to be highly cited for all time. Eventually an idea or paper may become so widely known that citing its original version is unnecessary. Or a new paper will supersede the original one by reformulating the idea in more up-to-date terms so the newer paper then receives all the citations to the idea.*" [Garfield, Citation data as science indicator]

These statements are applicable to the Grizzle [4](1965) and Hills and Armitage [7](1979), since the later paper provided a complete version of the two-stage procedure while Grizzle [4](1965) did not recommend methods of analysis if the carry-over effect was found to be significant.

The disparity between the citation rates for these two papers is partly explained by this. The fact that the two-stage procedure has been proven invalid may imply that application of the method may not be the only reason that these two papers were cited. Other possible causes will be investigated when other citation characteristics are discussed in following sections.

The differences between Freeman [8](1989) and the two papers which promoted the two-stage procedure were more complicated. These differences will be further discussed in this chapter after characteristics of the cited and citing papers are analysed.

**Table 4.5-1: Total number of citations on the three papers from 1981 to 2001**

| | Grizzle 1965 | Hills and Armitage 1979 | *Freeman 1989* |
|---|---|---|---|
| Total Number of Years | 22 | 22 | 12 |
| Total Number of Citations | 367 | 961 | 56 |

**Table 4.5-2: Citation rate by year for the period 1981-2001**

| Year | Grizzle 1965 | Hills and Armitage 1979 | Freeman 1989 | Total Number of Cross-over Trial Reports from Medline Search |
|---|---|---|---|---|
| 1981 | 20 | 17 | N/A | 95 |
| 1982 | 15 | 28 | N/A | 126 |
| 1983 | 20 | 45 | N/A | 121 |
| 1984 | 21 | 32 | N/A | 129 |
| 1985 | 17 | 49 | N/A | 134 |
| 1986 | 19 | 58 | N/A | 168 |
| 1987 | 19 | 59 | N/A | 160 |
| 1988 | 12 | 65 | N/A | 156 |
| 1989 | 22 | 84 | N/A | 165 |
| 1990 | 24 | 61 | 1 | 148 |
| 1991 | 24 | 49 | 3 | 182 |
| 1992 | 24 | 57 | 2 | 181 |
| 1993 | 20 | 60 | 8 | 177 |
| 1994 | 13 | 49 | 5 | 166 |
| 1995 | 14 | 50 | 6 | 166 |
| 1996 | 29 | 41 | 8 | 167 |
| 1997 | 17 | 42 | 6 | 161 |
| 1998 | 6 | 31 | 5 | 126 |
| 1999 | 12 | 32 | 5 | 122 |
| 2000 | 10 | 21 | 4 | 124 |
| 2001 | 5 | 30 | 3 | 188 |
| **Total** | **367** | **961** | **56** | **3,162** |

### 4.5.1. Citation types

Citation type is a classification of a citation by its purpose. It is used in this section as an indicator to identify areas of impact of the three papers. Although one assumption made for citation analysis is similarity of content {Smith (1981 and Garfield, 1979)}, it does not apply to citations on the two-stage procedure due to its

controversial history. Citation types cannot be assumed to be similar in this citation analysis but are treated as an indicator of areas of impact.

Papers citing Grizzle [4](1965), Hills and Armitage [7](1979) and Freeman [8](1989) can be categorised into two broad citation types according to the purpose of the citation. One is to use the citation as a reference of the statistical method applied in the paper. For example, the citation received from a cross-over trial report was assumed to be a reference to the statistical analysis. The other citation type was methodological discussion on designs or statistical methods of the two-stage procedure. This type of citation included Freeman [8](1989), which cited both Grizzle [4](1965) and Hills and Armitage (1979). Implications of these two citation types were not the same and sometimes contrasted.

The three papers to be investigated represented two very different viewpoints of analyses of cross-over trials. Depending on the paper being studied, implications of influences were sometimes different within the same citation type. Cross-over trial reports which cited any of these three papers usually used it as a reference for their statistical analyses. When Grizzle [4](1965) or Hills and Armitage [7](1979) received a citation from a cross-over trial report, the implication is that the two-stage procedure had been used to analyse the trial. In contrast, a citation given to Freeman [8](1989) from cross-over trial would imply that the two-stage procedure was not used. In other words, Freeman [8](1989) was more likely to

provoke discussions on methodological publications than Grizzle [4](1965) or Hills and Armitage [7](1979). There are also publications which cited two or all three papers. This is called co-citation and will be discussed in Section 4.5.4.

### 4.5.2. The classification of citation type

To classify the citation types of papers giving citations to Grizzle (1965), Hills and Armitage [7](1979) and Freeman [8](1989), titles of citing papers were first screened to identify whether they were trial reports (abstracts were searched where the classification could not be made from the title alone). All papers citing any of the three publications were categorized into two citation types: cross-over trial reports and non-trial papers. The results are shown in Table 4.5-3.

The distribution of citation types amongst papers citing Grizzle [4](1965), Hills and Armitage [7](1979) and Freeman [8](1989) were notably different. Both papers describing the two-stage procedure, Grizzle [4] (1965) and Hills and Armitage [7](1979) were more often cited in cross-over trial reports than in any other type of publication.

Among all 961 citations Hills and Armitage (1979) received between 1981 and 2002, 89 per cent were given by trial reports. In comparison, Grizzle [4](1965) received a lower percentage of citations from cross-over trial reports (62 per cent) and a higher percentage from non-trial reports.

Based on the assumption that citations given to Grizzle (1965)[4] and Hills and Armitage (1979)[7] by cross-over trial reports indicated adopting the method, there had been at least one thousand cross-over trials analysed with the two-stage procedure. With the actual type I error rate of the two-stage procedure being between 7-9.5%, there may have been between 20 to 45 more cross-over trials with false positive results than expected.

Although Grizzle (1965)[4] and Hills and Armitage (1979)[7] described the same statistical method in their papers, the distributions of their citations were different. There were two characteristics that differentiated these two papers: one was the type of journal they were published in; the other was the description of the method. Hills and Armitage (1979)[7] was easier to follow, giving a more carefully illustrated demonstration of how to apply the two-stage procedure. This paper clearly indicated that the analysis should be based on period 1 data solely if the carry-over effects were statistically significant while Grizzle (1965)[4] only suggested that "the change-over design should be avoided unless residual effects are of interest in their own right and are not regarded as a nuisance when residual effects are thought to be unequal".

**Citation types for Freeman (1989)**

In contrast, Freeman (1989)[8], showing the deficiency of the methodology applied in the previous two papers, was cited in a very different context. Ninety-five per cent of its citations were given by

non-trial papers, such as discussions of statistical methods and design issues of cross-over trials. Freeman (1989)[8] was rarely cited in cross-over trial reports. Only 3 of the 56 citations given to Freeman (1989)[8] were from cross-over trial reports, including two papers published by the same group of authors.

The distribution of citation types of papers citing Freeman (1989)[8] implied this paper was more of theoretical value than practical use. It raised awareness of the deficiency of the two-stage procedure but did not succeed in putting an end to its use.

**Table 4.5-3: Number of citation rates by citation type**

| Citation Type | Grizzle 1965 | | Hills and Armitage 1979 | | Freeman 1989 | |
|---|---|---|---|---|---|---|
| Trial reports | 229 | 62% | 860 | 89% | 3 | 5% |
| Non-trial reports | 138 | 38% | 101 | 11% | 53 | 95% |
| **Total** | **367** | 100% | **961** | 100% | **56** | 100% |

**Possible weaknesses of the citation analysis on the two-stage procedure of cross-over trials:** First, the assumption that citation indicates use was not verified for this analysis. The citation analysis can only give an approximation of the prevalence of use of the two-stage procedure. There might be occasions where the method was not applied even when Grizzle (1965)[4] or Hills and Armitage (1979)[7] was cited.

Secondly, the citation analysis did not cover cross-over trial reports that did not cite any of the 3 papers of the two-stage procedure. Therefore, methods of analysis for cross-over trials which did not cite any of these three papers remained unknown. Both weaknesses will be overcome in the next chapter by the systematic review which

will give a closer examination of statistical methods used in cross-over trial reports.

The high citation rates on Grizzle (1965)[4] and Hills and Armitage (1979)[7] by cross-over trial reports and the low citation rate of Freeman (1989)[8] among cross-over trials implied that the deficiency of the two-stage procedure was virtually unknown to clinical researchers or the research community who published cross-over trials. In order to investigate statistical methods recommended to researchers who analysed these papers, another common reference of medical statistics, general medical statistics books will be reviewed and discussed in Chapter 5.

### 4.5.3. Journal types

There has been some considerable differences in the distribution of citation types between Grizzle (1965)[5] and Hills and Armitage (1979)[7] despite their similarity in content. One of the differences between these two papers was the area of research covered by the journals in which they were published. Grizzle (1965)[4] published the original paper of the two-stage procedure in *Biometrics*, a publication of the International Biometrics Society which focuses on the development of new methods in the biological sciences. Hills and Armitage (1979)[7], the most frequently cited paper of the two-stage procedure, was published in the *British Journal of Clinical Pharmacology*, which contains reports on all aspects of drug action and is published on behalf of the British Pharmacology Society.

These two journals have different target audiences, which is likely to affect the distribution of citation types.

The titles of journals in which these three papers are cited were checked manually according to their research areas. The results of the analysis of journal types are shown in Table 4.5-4.

Differences in citation types were noticeable amongst the three papers. There were however, some similarities between the two papers promoting the two-stage procedure, which were both most cited in medical journals. Hills and Armitage (1979)[7] received nearly 90 per cent of their citations from medical journals while three quarters of Grizzle's (1965)[4] were given by papers published in medical journals.

The difference in citation types seen previously between Grizzle (1965)[4] and Hills and Armitage (1979)[7] could probably be explained by the type of journal in which the cited papers were published. Grizzle (1965)[4] was published in a statistical journal (*Biometrics*), while Hills and Armitage (1979)[7] was published in a medical journal (*British Journal of Clinical Pharmacology*).

All three papers were more cited in medical journals than journals of other disciplines. Statistics in Medicine, where Freeman (1989)[8] was published, has a similar objective to the journal where Grizzle (1965)[4] was published. Although Freeman (1989)[8] was more cited in medical journals than statistical journals, the contents of the citation papers were mostly methodology issues rather than practical applications (see Table 4.5-3).

**Table 4.5-4: Number of citations by citing journal type**

| | Grizzle 1965 | | Hills and Armitage 1979 | | Freeman 1989 | |
|---|---|---|---|---|---|---|
| | Biometrics | | British Journal of Clinical Pharmacology | | Statistics in Medicine | |
| **Medical Journals** | 267 | 73.5% | 856 | 89.3% | 22 | 39% |
| **Statistical Journals** | 83 | 23% | 51 | 5.3% | 34 | 61% |
| **Journals of Other Disciplines** | 13 | 3.5% | 52 | 5.4% | 0 | 0% |
| **Total** | 363 | 100% | 959 | 100% | 56 | 100% |

**Table 4.5-5: Impact factors of the original papers**

| Author | Journal Title | Impact Factor* | 2000 Total Cites | 2000 Articles | Immediacy Index** | Citation Half-Life*** |
|---|---|---|---|---|---|---|
| **Grizzle 1965** | **Biometrics** | 1.170 | 6,816 | 171 | 0.129 | >10.0 |
| **Hills and Armitage 1979** | **British Journal of Clinical Pharmacology** | 2.151 | 5,725 | 167 | 0.623 | 7.8 |
| **Freeman 1989** | **Statistics in Medicine** | 1.717 | 4,088 | 231 | 0.190 | 6.7 |

*Impact Factor measures how often articles in a specific journal have been cited. The total number of quotes during a year of the two immediately preceding years' issues, for example quotations in 1994 of the journals published in 1992 and 1993, are weighed against the number of articles published in 1992 and 1993 in that journal.

**Immediacy Index measures the average number of times that an article, published in a specific year within a specific journal, is cited over the course of the same year.

***Cited Half-life measures the number of years, going back from the current year, that account for half the total citations received by the cited journal in the current year.

**Table 4.5-6: Journal Distribution of Citations to Freeman 1989**

| Statistical Journals N=34 | Number of Citations |
|---|---|
| Statistics in Medicine | 15 |
| Biometrics | 7 |
| Journal of Statistical Planning and Inference | 3 |
| Journal of Royal Statistics Society D-Stat | 3 |
| Biometrical Journal | 1 |
| Biometrika | 1 |
| Journal of American Statistics Association | 1 |
| Journal of Applied Statistics | 1 |
| Statistical Methods in Medical Research | 1 |
| Communication in Statistics-Theory and Methods | 1 |
| **Non-Statistical Journals N=22** | |
| Journal of Clinical Pharmacology | 4 |
| British Journal of Clinical Pharmacology | 3 |
| European Journal of Clinical Chemistry and Clinical Biochemistry | 2 |
| International Journal of Clinical Pharmacology and Therapeutics | 2 |
| Archives of General Psychiatry | 1 |
| Controlled Clinical Trials | 1 |
| BMJ | 1 |
| Pharmacology World and Science | 1 |
| European Journal of Clinical Pharmacology | 1 |
| Drug and Alcohol Dependence | 1 |
| Psychopharmacology | 1 |
| Nephrology Dialysis Transplantation | 1 |
| Angiology | 1 |
| Clinical Pharmacology Therapeutics | 1 |
| Postgraduate Medicine | 1 |

### 4.5.4. Co-citation analysis

Due to the contrasting implications of citations given to Freeman (1989)[8] and the other two papers describing the two-stage procedure, the co-citation was investigated. Co-citation reflects the association between highly cited papers as perceived by the current population of specialists who have themselves published papers. [Garfield, Citation data as science indicator]

The focus of this section is on co-citation of Freeman (1989)[8] with Grizzle (1965)[4] or Hills and Armitage (1979)[7]. All papers citing any of the three papers were checked to see whether there was any co-citation with the other papers. Results are shown in Table 4.5-7.

Freeman (1989)[8] was often cited with Grizzle (1965)[4] or Hills and Armitage (1979)[7] in methodological publications. While there is usually only one statistical referenced cited in cross-over trial reports, co-citation between these three papers rarely occurred in cross-over trials reports.

Since Freeman (1989)[8] only received three citations from cross-over trial reports and none of them co-cited either Grizzle (1965)[4] or Hills and Armitage (1979)[7], it indicates that all co-citations among Freeman (1989)[8] with Grizzle (1965)[4] or/and Hills and Armitage (1979)[7] were by methodology papers.

Freeman (1989)[8] and Grizzle (1965)[4] had been co-cited 23 times in a period of 12 years since Freeman (1989)[8] was published. Eighteen of these co-citations were published in statistical journals and five of them in medical journals (one in the *European Journal of Clinical Chemistry and Clinical Biochemistry*, two in the *International Journal of Clinical Pharmacology and Therapeutics*, two in the *British Journal of Clinical Pharmacology*).

There were 19 publications which co-cited Freeman 1989 with Hills and Armitage 1979. Among these 19 papers, 9 were published in non-statistical journals and were contributed by 4 individuals. One author contributed 6 co-citations in non-statistical journals.

The titles and authors of the 9 papers are listed below:

1. *Journal of Clinical Pharmacology* – statistical issue, binary data - Cleophas
2. *International Journal of clinical Pharmacology and Therapeutics* – a simple analysis of crossover studies with one-group interaction - Cleophas
3. *Nephrology Dialysis Transplantation* – Statistical aspects of the design and analysis of studies to compare hemodialysis membranes – Matthews JNS
4. *Journal of Clinical Pharmacology* – Interaction in crossover studies - Cleophas
5. *Journal of Clinical Pharmacology* – Carryover bias in clinical investigations - Cleophas
6. *Clinical Pharmacology and Therapeutics* – Crossover studies – a modified analysis with more power - Cleophas
7. *Angiology* – Interaction in Cardiovascular crossover studies – the standard and the clinical analysis – Cleophas
8. *British Journal of Clinical Pharmacology* – Should we cross off the crossover – Armitage
9. *British Journal of Clinical Pharmacology* – Problems with the 2 stage analysis of crossover trials – Senn

There were also eight papers which cited all three papers and all were on theoretical publications on the design and statistics of cross-over trials.

The titles are as follows:

1. Cross-over trials in drug development: theory and practice – *Journal of Statistical Planning and Inference*
2. Multivariate AB-BA crossover trial – *Journal of Applied Statistics*
3. A comment on interim analysis in crossover trials - *Biometrics*
4. Design considerations in crossover trials with a single interim analysis and serial patient entry - *Biometrics*
5. Interim analysis in 2X2 crossover trials – *Biometrics*
6. A simple analysis of crossover studies with one-group interaction – *International Journal of Clinical Pharmacology Theory*
7. Should we cross off the crossover – *British Journal of Clinical Pharmacology*

8. Problems with the 2 stage analysis of crossover trials – *British Journal of Clinical Pharmacology*

**Table 4.5-7: Co-citation rates**

| Number of co-citations | Grizzle 1965 | Hills and Armitage 1979 | Both Grizzle 1965 and Hills and Armitage 1979 |
|---|---|---|---|
| **Freeman 1989** | 23 | 19 | 8 |
| Total number of citations | 367 | 961 | 56 |

**Co-citation Strength**

Two formulas: Jaccard's index and Salton's cosine formula, have been used to express the degree of similarity in citation analysis. Salton's cosine formula is used here to quantify co-citation strength as *"Salton's cosine formula deals more effectively with links between high- and low-cited papers than does Jaccard's."* [25]

Salton's cosine formula for co-citation strength between document $i$ and $j$ is defined as:

$$S_s(i,j) = \frac{COC_{i,j}}{\sqrt{cit_i \times cit_j}}$$

where

$COC_{i,j}$ = number of citations between documents $i$ and $j$,
$cit_i$ = number of citations for document $I$, and
$cit_j$ = number of citations for document $j$

Co-citation strength between Freeman and Grizzle was $\frac{23+8}{\sqrt{367 \times 56}} = 0.2162$.

Co-citation strength between Freeman and Hills and Armitage was $\frac{19+8}{\sqrt{961 \times 56}} = 0.1164$.

The coefficients showed that co-citation was stronger between Freeman (1989)[8] with Grizzle (1965)[5] than Freeman (1989)[8] with Hills and Armitage (1979)[7]. As 53 of the 56 citations on

Freeman (1989)[8] were non-trial reports, most co-citations between Freeman (1989)[8] and the other two papers were methodological or theoretical papers. It can be concluded that the original paper of a new method was more likely to be cited in theoretical papers than the follow-up paper. Furthermore, papers which were easier to follow were more likely to be cited in the papers of practical applications.

### 4.5.5. Citation age

Citation age is the amount of time (in years) between the publication of the cited document and the publication of the citing document. The citation rate often decreases with the citation age as new and more advanced references may be conducted in the field. Citation age is used instead of citation years in this section to compare citation rates. As the three papers were published in three different decades, there was no common citation age for all three papers.

Papers were only available in the Science Index Expanded if they were published after 1981. The database provided citation ages of the citing papers from 16 to 36 for Grizzle (1965)[4], 2 to 22 for Hills and Armitage (1979)[7], 1 to 13 for Freeman (1989)[8] (as shown in Table 4.5-8). The only common years that allowed direct comparisons were the seven years from 16 to 22 for both Grizzle (1965)[4] and Hills and Armitage (1979)[7], and 2 to 12 for Hills and Armitage (1979)[7] with Freeman (1989)[8].

**Table 4.5-8: Number of citations by citation age**

| Citation age (years) | Grizzle 1965 | Hills and Armitage 1979 | Freeman 1989 |
|---|---|---|---|
| *1* | | | 1 |
| *2* | | 17 | 3 |
| *3* | | 28 | 2 |
| *4* | | 45 | 8 |
| *5* | | 32 | 5 |
| *6* | | 49 | 6 |
| *7* | | 58 | 8 |
| *8* | | 59 | 6 |
| *9* | | 65 | 5 |
| *10* | | 84 | 5 |
| *11* | | 61 | 4 |
| *12* | | 49 | 2 |
| *13* | | 57 | |
| *14* | | 60 | |
| *15* | | 49 | |
| *16* | 20 | 50 | |
| *17* | 15 | 41 | |
| *18* | 20 | 42 | |
| *19* | 21 | 31 | |
| *20* | 17 | 32 | |
| *21* | 19 | 20 | |
| *22* | 19 | 16 | |
| *23* | 12 | | |
| *24* | 22 | | |
| *25* | 24 | | |
| *26* | 24 | | |
| *27* | 24 | | |
| *28* | 20 | | |
| *29* | 13 | | |
| *30* | 14 | | |
| *31* | 29 | | |
| *32* | 17 | | |
| *33* | 6 | | |
| *34* | 12 | | |
| *35* | 10 | | |
| *36* | 3 | | |

There has not been an obvious decline in numbers of citations on all three papers as the age of cited documents increased. Citation rates did not seem to be affected by the length of time between cited and citing papers.

Citation rates on Hills and Armitage (1979)[7] superseded Grizzle (1965)[4] in all calendar years except 1981, the results were similar when the citation age was used for comparison instead of calendar years. Hills and Armitage (1979)[7] received more citations than

Grizzle (1965)[4] for all citation ages they had in common except year 22. The difference is twice as much for citation ages 16 to 18, the difference in their citation rates got smaller as the cited paper got older. Eventually, at citation age 22, there were more citations on Grizzle (1965)[4] than on Hills and Armitage (1979)[7]. It can be concluded that Hills and Armitage (1979)[7] had a bigger impact than Grizzle (1965)[4] in terms of the total number of citations.

Hills and Armitage (1979)[7] and Freeman (1989)[8] could be compared from citation age 2 to 12 years. For these 11 years, citations to Hills and Armitage (1979)[7] had far exceeded citations to Freeman (1989)[8]. The difference in citing rate has been substantial and there was no indication of any change as times passes.

There was no common citation age to compare citation rates between Grizzle (1965)[4] and Freeman (1989)[8]. However, it can be predicted that Freeman (1989)[8] is unlikely to exceed Grizzle (1965)[4] in the near future unless awareness of the deficiency of the two-stage procedure is raised among the researchers involved with cross-over trials.

### 4.5.6. The statistical model of the citation analysis

**Generalised Linear Model for Count Data:**

**The Poisson Regression Model**

The analysis is completed with Poisson regressions with the objective of determining whether the citation rates were influenced

by the paper cited, citation age and numbers of cross-over trials published in that year. For citation rate on each of these papers per year, Poisson regression models estimated the number of citations and a 95% confidence interval as a function of citation age and total number of cross-over trials. The GENMOD procedure in the SAS system was used to fit the regression models by maximum likelihood.

The variables fitted in this citation analysis included number of citations for each author (21 observations for both Grizzle and Hills & Armitage, 12 for Freeman), citation age (number of years since publication), author (as a categorical variable) and tt (total number of cross-over trials found in Medline of the year, used as a proxy measure). An over-dispersion parameter was fitted in order to correct for the effects of the larger variance on the p-values.

The SAS code is:

```
Proc genmod data=com;
Class author;
Model citation = author citage /offset=ltt dist=Poisson link=log
obstats residuals dscale Type1 Type3;
Estimate 'Grizzle-Freeman' author 1 0 -1;
Estimate 'H&A-Freeman' author 0 1 -1;
Estimate 'H&A-Grizzle' author -1 1 0;
```

## Results

The conclusion can be summarised as follows. Hill and Armitage (1979)[7] has the highest rate of citation, followed by Grizzle (1965)[4]. The total number of citations is in proportion to the total number of cross-over trials in that given year. The number of citations decreases as the citation ages. The overdispersion has

been adjusted for by a parameter of 1.5.

**Table 4.5-9: Criteria for assessing goodness of fit**

| Criterion | DF | Value | Value/DF |
|---|---|---|---|
| Deviance | 50 | 117.6588 | 2.3532 |
| Scaled deviance | 50 | 50.0000 | 1.0000 |
| Pearson Chi-Square | 50 | 111.2199 | 2.2244 |
| Scaled Pearson X2 | 50 | 47.2638 | 0.9453 |
| Log likelihood | - | 1462.2133 | |

**Table 4.5-10: Analysis of parameter estimates**

| Parameter | DF | Estimate | Standard Error | Chi Square | Pr>Chi |
|---|---|---|---|---|---|
| Intercept | 1 | -3.3751 | 0.2096 | 259.26 | <.0001 |
| Grizzle | 1 | 1.8574 | 0.2623 | 50.15 | <.0001 |
| Hill&Armitage | 1 | 2.4800 | 0.2147 | 133.10 | <.0001 |
| Freeman | 0 | 0.0000 | 0.0000 | | |
| Citation age | 1 | -0.0249 | 0.0073 | 11.68 | 0.0006 |
| Scale | 0 | 1.5340 | 0.0000 | | |

## 4.6. Results and Discussion

It has been shown in Section 4.4.1. that there were clear differences in terms of citation rates between any two of the three papers.

Despite the similarity in context, the difference in citation rates of Grizzle (1965)[4] and Hills and Armitage (1979)[7] has been substantial. Such disparities were partly explained in Section 4.4.1. However, when looking at citation types and journal types of citing papers, there were also some differences. In addition to the reasons exposed in Section 4.4.1. , other factors could explain the difference in citation rates.

It was shown in Table 4.5-4 that there were differences between types of journals that published citations on Grizzle (1965)[4] and Hills and Armitage (1979)[7]. Seventy-four per cent of citations given to Grizzle (1965)[4] were from medical journals while the

percentage was as high as 89 per cent for Hills and Armitage (1979)[7]. Grizzle (1965)[4] was published in a statistical journal that is often not so reader friendly to those who conducted and analysed cross-over trials which were published in medical journals, while Hills and Armitage (1979)[7] was published in a medical journal. Both papers describing the two-stage procedure received most citations from cross-over trial reports which were published in medical journals. Therefore, it could be that the difference in citation rates between Grizzle (1965)[4] and Hills and Armitage (1979) was a result of the type of journal they were originally published in.

Another difference that could result in the disparity between citation rates of Grizzle (1965)[4] and Hills and Armitage (1979)[7] was their description of the two-stage procedure. The two-stage procedure described by Hills and Armitage was a more complete version than Grizzle (1965)[4]. Grizzle (1965)[4] did not propose a solution for analysing the trial when the carry-over effect was found to be significant. Hills and Armitage (1979)[7] suggested that the treatment difference should be evaluated using first period data only if the carry-over effect was found to be significant at the 10% nominal level. It is possible that Grizzle (1965)[4] was less likely to be applied as there was a risk that the treatment effect could not be evaluated. However, it is also possible that most researchers who analysed and published the paper were not aware that there were two versions [Grizzle (1965)[4] and Hills and Armitage (1979)[7]] of

the two-stage procedure. Hence, Hills and Armitage (1979)[7] was more cited simply because it was better known. The reason could be explained by a snowball effect. Clinical researchers often seek other trial reports of similar design or treatment for references. It is a potential path of learning statistical analyses of cross-over trials and to generate a higher citation rate for Hills and Armitage (1979)[7].

### 4.6.1. The difference in citation rates between papers describing the two-stage procedure and papers exposing its deficiencies

The reason for the disparity in citation rates between the two-stage procedure and Freeman (1989)[8] is more complicated and controversial.

The most common explanation of the low citation rate of a publication was that it was simply not as well-known as other highly cited papers. There were a few reasons for its anonymity. The two-stage procedure had been established for 24 years before Freeman (1989)[8] found that it had a high type I error rate. The two-stage procedure was a 'classic' and was so often encountered in cross-over trial reports that those who analysed cross-over trials may have considered it a standard procedure. In addition, the two-stage procedure was also suggested by many medical statistics books (see chapter 5) and clinical trial courses in medical school.

As Freeman (1989)[8] is a paper which exposed the major deficiency of a widely known and applied statistical procedure, anonymity

might not be the only reason for its relatively low citation rates. There were a few reasons that could result in citation rates not meeting up to awareness of such a deficiency.

First, the citations on Grizzle (1965)[4] and Hills and Armitage (1979)[7] originated from a very homogenous research population (i.e., medical researchers), while Freeman (1989)[8] was better known by statisticians. As the population of medical researchers (doctors) was bigger than the population of statisticians, the relatively low citation rates may be the result of the difference in the size of the papers' respective readerships. Freeman (1989)[8] was better known amongst statisticians than medical researchers who analysed and published the cross-over trial reports. Medical researchers may have been more likely to seek their references from analyses of similar study designs that have already been published in medical journals rather than search for methods in statistical journals

Secondly, the impact of Freeman (1989)[8] may only be on a superficial level. It might have raised some awareness of the deficiency but may not have been agreed with (examples of this can be found in the questionnaire survey of the pharmaceutical statisticians in Chapter 7 - ). In addition, testing for the carry-over effect was an extremely attractive feature. Despite its deficiencies, many authors of cross-over trial reports feel that only by using it could they ensure that the treatment effect has not been biased due to the carry-over effect.

In conclusion, the impact of Grizzle (1965)[4] and Hills and Armitage (1979)[7] on the two-stage procedure were mainly on actual analysis of the two-stage published in periodicals. Hills and Armitage (1979)[7] who completed the method described by Grizzle (1965)[4] had most influence on analysis of cross-over trials. The high type I error rate of the two-stage procedure discovered by Freeman (1989)[8] received far fewer citations in comparison to Grizzle (1965)[4] and Hills and Armitage (1979)[7]. Freeman's impact was mainly on theoretical discussion and the awareness of the deficiency of the two-stage procedure has not been raised amongst researchers.

(For further discussion, please see the survey conducted among pharmaceutical statisticians regarding the issue of the two-stage procedure (in Chapter 7 - below), and a comprehensive review of medical statistics books on what clinicians are told (in Chapter 6 - below).

## 4.7. Cited books

In addition to the three papers, there were three books dedicated to cross-over trials and all three discuss the statistical issues of such designs. As a rule, about 20% of the references in science journals cite books and other non-journal items [26]. In order to provide other references used for analyses of cross-over trials published in medical journals, citations given to the three books of cross-over trials were also searched using the same database and options.

**Table 4.7-1: Citation rates of cross-over trial books**

| | Jones and Kenward 1989 | Senn 1993 | Ratkowsky 1993 |
|---|---|---|---|
| **Title** | The Design and Analysis of Cross-over Trials | Cross-over Trials in Clinical Research | Cross-over Experiments, Design, Analysis and Application |
| **Number of Citations** | 411 | 174 | 36 |

## 4.8. The updated search

A new search on the citation rates for the three key papers was conducted on the 5th of November 2004. The searching strategy for the new search is identical to the one used for the original search.

**Table 4.8.1: The total numbers of citations on the three papers between Feb 2002 and Nov 2004**

| | Grizzle (1965) | Hills and Armitage (1979) | Freeman (1989) |
|---|---|---|---|
| Search conducted in Feb 2002 | 367 | 961 | 56 |
| Search conducted on the 5th Nov 2004 | 468 | 1043 | 66 |
| Total Number of Citation Years | 39 | 25 | 15 |

The Web of Science has been updated since the original search was conducted in February 2002. Papers that were published between 1945 and 1980 have been added to the Web of Science. The results of the updated citation rates were 468 citations for Grizzle (1965), 1043 citations for Hills and Armitage (1979) and 66 citations for Freeman (1989).

The number of citation years for each of the three papers has also changed. The newly added citations on Freeman (1989) were papers that published between February 2002 and Nov 2004 while the newly added citations on Grizzle (1965) contained papers that were published during this period of time as well as papers published between 1945 and 1980.

In terms of the citation types, the new citations given to Freeman (1989) contained only 1 cross-over trial report while new citations given to Grizzle (1965) and Hills and Armitage (1979) were mainly

trial reports. The results of the new citation searches confirmed the previous prediction that it is unlikely that the citation rates on Freeman (1989) would catch up with Grizzle (1965) or Hills and Armitage (1979).

# Chapter 5 - Cross-over trials in medical journals – a systematic review

## 5.1. Introduction

The overall objective of this thesis is to investigate the use of the two-stage procedure for the analyses of cross-over trials in four different realms of medical statistics. It was shown in Chapter 4 - that both papers [4][7] that described the two-stage procedure of cross-over trials had been widely cited in medical journals while the impact of the deficiencies was mainly on the discussion of analyses of cross-over trials. There were two questions which could not be answered by the citation analysis. First, whether the two-stage procedure had been applied when a reference to this procedure was not cited. Second, whether citation of the two-stage procedure always indicated the usage of this procedure. These two questions will be investigated in this chapter.

## 5.2. Literature review of the use of the statistics in medical journals

The credibility of statistics in medical journals has often been questioned. Several studies have shown that it is not uncommon to find statistical errors in published articles. A review of 62 articles in British Medical Journal (BMJ) from January to March 1976 by Gore et al [47] showed that "*32 of them had statistical errors of one kind or another and 18 articles contained fairly serious faults.*"

Felson et al [33] compared misuses of statistical methods in the journal Arthritis and Rheumatism in 1982 and between 1967 and 1968. They found that methodological errors were contained in 66% of the articles published in 1982 and 60% in those published in 1967-1968. They also found that there was a growth in the use of statistics in medical journals from 1967/68 to 1982.

Another study, conducted by Cruses [35], reviewed 201 articles published in the American Journal of Tropical Medicine and Hygiene and found that 73.5% of the papers had at least 1 detectable statistical error.

It was noted that two of the most commonly used statistical methods: Student's t-test and Chi-squared test also had the highest rate of misuse. The most common mistake in using the Student t-test occurred when the data were not (approximately) normally distributed, and the misuse of the one and two sample t-tests.

In addition to the statistical errors, a common problem in the reporting of trials was the publication bias (publication bias is a tendency *on average* to produce results that appear significant, because negative or near neutral results are almost never published). Pocock et al [44] reviewed 45 reports of comparative trials published in British Medical Journal, the Lancet, or the New England Journal of Medicine and concluded that, overall, the reporting of clinical trials appeared to be biased toward an exaggeration of treatment differences.

## 5.3. Statistics for the two-stage procedure and its deficiency in relation to general medical statistics books

The statistical concepts used to evidence the deficiencies of the two-stage procedure were beyond the average coverage of general medical statistics books. It would be unrealistic to expect clinical researchers who conducted cross-over trials to understand the consequences in terms of statistics if the two-stage procedure was used. Nevertheless, it would be difficult to determine whether not using the two-stage procedure was a sign of not knowing the existence of the method or a full understanding of the deficiencies. On the other hand, considering the misuse of basic statistical tests published in medical journals, the two-stage procedure of cross-over trials which involves three t-tests including one and two-sample type may contribute to more misuse in medical journals if it was adopted. It is, however, not the intention of this study to examine or criticise the level of statistical knowledge of authors of cross-over trials reports. The objective is to evaluate and illustrate the impact of the two-stage procedure and its deficiencies.

In addition to medical statistics books and statistics journals, analyses of a cross-over trial may also be based on a trial that is similar in the patient conditions, the trial design or the treatment. The author of the report may review other cross-over trials to ensure that the statistical method he/she used was similar to what had been published. Therefore, a snowball effect may be created in this context, especially when authors, research centres or medical

journals with more credits may have some impact on analyses of cross-over trials in medical journals. It is not the purpose of this thesis to analyse the snowball effect. It has therefore been left out of the systematic review.

## 5.4. Objective

The objective of the review is to investigate statistical methods used to analyse cross-over trials in medical journals, with emphasis on the two-stage procedure and whether the carry-over effect has been tested for by statistics. Other information provided in cross-over trial reports will be used to investigate design and application aspects.

## 5.5. Methodology

Systematic review (as defined in Section 3.2.2. )

As a general rule, a second reviewer is required for a publication of a systematic review. One was, however, not available for this thesis. As an alternative for the validity of results, original quotes of information used for drawing the results of this chapter are given in an appendix for reference.

### 5.5.1. Sampling population

The sample population was all cross-over trial reports available on Medline. It was searched using the key words 'cross-over trial', 'cross-over trials', 'crossover trial', 'crossover trials', 'change-over trial', 'change-over trials', 'changeover trial' and changeover trials'.

The total number of cross-over trials published in medical journals between 1981 and 2001 was 3,162, an average of 151 cross-over trials every year. See Appendices Appendix A

### 5.5.2. Sampling method

A stratified, simple random sampling method was used to choose reports to be reviewed. Four publication years were used in order to study statistical methods used to analyse cross-over trials at different years. The publication years (1988, 1992, 1996 and 2000) were chosen at even intervals of 4 years, including the year before the major deficiencies of the two-stage procedure was published, in order to study the impact of the finding of the deficiency.

### 5.5.3. Sample size

The database was stored in Reference Manager® 9 and each report was given an identity number for sampling purposes.

It was estimated that a sample size of 60 reports was a feasible number as part of the PhD project and a sufficient number comparing to systematic reviews published in medical journals. There were 4 publication years to be reviewed and it was decided to stratify by year with equal numbers per year, therefore the total number of papers being reviewed had to be a multiple of four. According to the assessment of time from the pilot study, it would take approximately 3 months to complete the first round of the review. Once the first review was completed, the second review

would be conducted in order to ensure that all relevant information regarding the design and analysis of trial would be extracted from the report. Time taken for second review was about half of the first review. In addition to the time taken for the review, time required for obtaining the trial reports was also taken into account. Therefore a sample size of sixty-four reports was considered a feasible quantity to review in 6 months.

Similar sample sizes were commonly used for systematic reviews published in medical journals, which included 62 reports in the British Medical Journal (BMJ) by Gore et al (1977), 67 clinical trials by DerSimonian et al [48] and 45 trials by Pocock et al [44].

### 5.5.4. Sample selection

The list of cross-over trial reports to be reviewed was generated randomly on the SAS® programme. To choose reports from the sampling population, sixteen numbers with two decimal places between 0 and 1 were drawn and the procedure was repeated four times for the four publication years. The figures were multiplied by the total number of trials of that year and round to the nearest integer. Reports of identity numbers which met the number derived from the simulation were the cross-over trials to be reviewed.

Due to the use of simple random sampling for each of the selected four years, samples of cross-over trial reports chosen for analysis came from publications across all types of journals. The selected reports were collected from university libraries and the British

Library (if not available in university libraries). All of the selected reports drawn by simple random sampling were obtained. Because the sample selection was chosen at random, the samples cover various types of diseases, treatments, size of trial, type of cross-over designs and geographical locations. See Appendix B for the list of trial reports reviewed.

## 5.6. Pilot of the systematic review

A checklist of items to be reviewed was tested over a pilot study in order to examine its completeness and practicality. A pilot study of 20 cross-over trial reports was conducted using the exact method that was planned for the systematic review. By assessing the length of time taken to review each paper, the pilot study also served the purpose of estimating a practical number of cross-over trial reports to be reviewed. Time taken for each review depended on the complexity and length of the paper. On average, each review took approximately one working day. It was also gradually reduced as more papers were reviewed.

## 5.7. Checklist - references to guidelines

As there are no existing guidelines for reporting cross-over trials, the checklist of this systematic review took references from various guidelines (the CONSORT statement, ICH-E9 Statistical Principles for Clinical Trials, and BMJ checklist). The CONSORT statement [42] was the main framework for the checklist as it provided the standard report of parallel clinical trials in medical journals. It was

introduced in 1996, and was intended to improve the reporting of the RCT (randomised controlled trials), enabling readers to understand how a trial had been conducted and to assess the validity of results. It has been endorsed by 175 medical journals [37] and 32 organisations, including pharmaceutical companies, medical associations, societies and government bodies [38]. Therefore a checklist that covers equivalent items for cross-over trials could be used as a standard criterion of what could be expected from cross-over trial reports. In addition to the items recommended by the CONSORT statement (Appendix E), the checklist for the systematic review on cross-over trial reports included a few items which only applied to cross-over design. The main objective of the checklist was to make sure that readers would be able to identify the design of trial from the report.

## 5.8. How to report a cross-over trial

How cross-over trials should be reported can be discussed in terms of two aspects: the design and the analysis.

### 5.8.1. Design

The main objective of the systematic review of cross-over trial reports published in medical journals was to investigate statistical methods used for such a design. However, without knowing the design of the trial, it is impossible to judge whether the statistical analysis has been applied appropriately. Therefore the reporting of

the design was also included in the systematic review in order to assess the statistical analysis.

The CONSORT statement was intended for reports of parallel-group randomised trials, therefore some information on design of cross-over trials would be missing from the report even if the list was strictly followed. As a consequence, readers would not be provided with sufficient information regarding the design of cross-over trials. For example, the CONSORT statement requires the number of interventions for the trial to be reported. However, unlike parallel trials, numbers of treatments are not always equivalent to numbers of treatment groups or treatment sequences for cross-over trials (For example, Balaam's design: AA/BB/AB/BA and AABB/BBAA/ABBA/BAAB). While parallel trials involve only one treatment period, there are at least two treatment periods in cross-over trials. Numbers of periods are not always equivalent to number of treatments (For example, three period design with four sequences: ABB/BAA/BAB/ABA, four periods with four sequences: AABB/BBAA/ABBA/BAAB, and four period design with six sequences: AABB/BBAA/ABBA/BAAB/ABBB/BAAA).

In addition to the items in the CONSORT statement, cross-over trial reports should specify the number of treatment groups (sequences) and the number of treatments given for each patient. It is not sufficient to just state the number of treatments or sequences. Items which should be included in cross-over trial reports are:

**Numbers of randomised groups.** This is usually equivalent to numbers of sequences.

**Sequences of treatments for each randomised group**. This should include numbers of treatment periods and which treatment is given at each treatment period.

As incorporating a wash-out period is the most reliable method to eliminate the possibility of the carry-over effect, whether the wash-out period was incorporated between treatment periods should always be reported. In addition, the length of the wash-out period should also be included in the trial report. The half-life of the treatment can also be used to justify the length of the wash-out period.

### 5.8.2. Randomisation

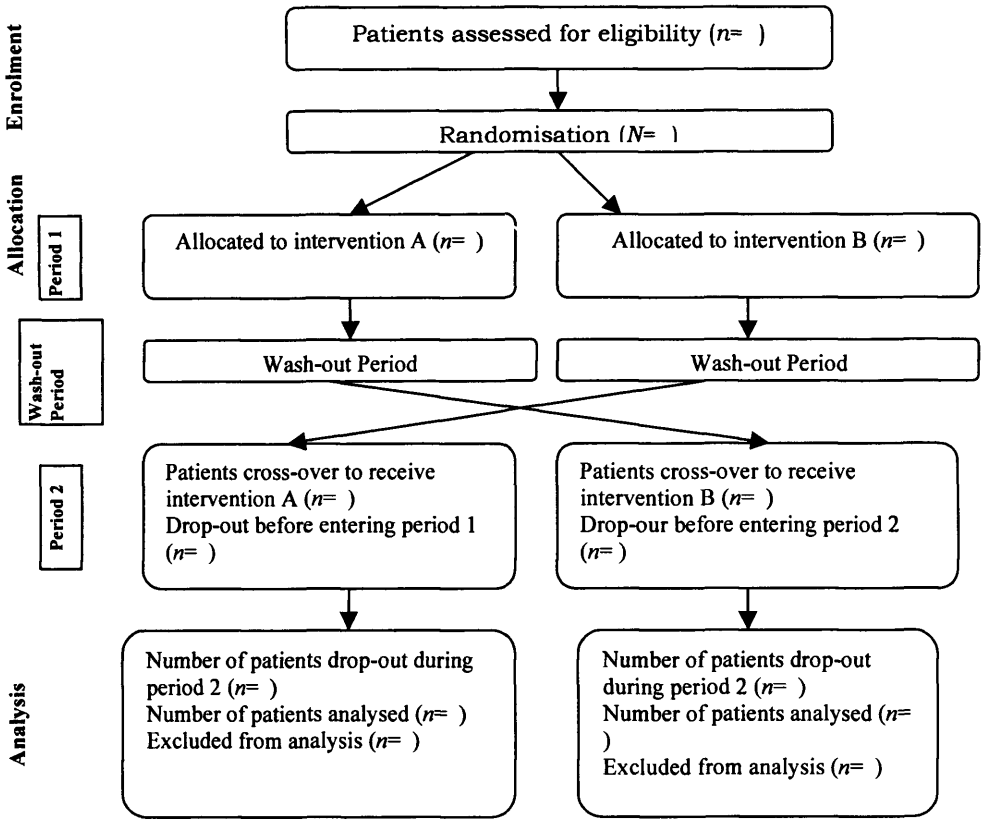
A difference between randomisations of a parallel trial and a cross-over trial is the unit that patients are randomised to. Patients participate in cross-over trials are randomised to sequences of treatments while patients are randomised to treatments in a parallel trial. In a randomised cross-over trial, the number of randomised groups is equivalent to the number of sequences, but is not necessarily the same as the number of treatments. The sequence that a patient is randomised to would determine the order of treatments he/she is given. All randomisation methods (e.g., simple randomisation, block randomisation, stratified randomisation and minimisation) should be based on sequences

not treatments. For example, the block size would be a multiple of the number of sequences rather than the number of treatments. However, the reporting of randomisation for a cross-over trial is very similar to what is suggested in the CONSORT, except the sequence of each randomisation group should be specified.

### 5.8.3. Flowchart

A flowchart that indicates all stages of the cross-over trial is probably the most comprehensible tool for readers to identify the design. The recommendation for the flow diagram of the progress through the phases of a randomised parallel trial was made in the CONSORT statement. Below is an example of the diagram for an AB/BA trial. Numbers of patient drop-outs at different stages of the trial can also be specified using the flowchart.

**Figure 5.8.3: A suggested flow diagram of subjects progress through the phases of a AB/BA trial**



### 5.8.4. Analysis

A few factors such as the period effect, the carry-over effect and the patient effect, which are not generally relevant to parallel trials are often associated with the cross-over design.

The main outcomes of a randomised controlled trial (RCT) are usually the treatment effects. However, there are a few different methods for evaluating the treatment for cross-over trials (as shown in Chapter 2 - ). Therefore cross-over trial reports should be especially clear about how treatment effects are analysed. For example, whether direct treatment differences are assessed or period effects have been adjusted for and perhaps also whether patient effects are treated as fixed or random.

In addition to the results of treatment effects, the carry-over effect and the period effect are often mentioned in trial reports. The carry-over effect is considered to be the 'chief difficulty' of the cross-over design and an approach to testing for its presence was described by Grizzle (1965)[4] and Hills and Armitage (1979)[7]. Although the carry-over effect should be designed out by allowing sufficient wash-out period (see previous section) and therefore not tested for, it is evident that the two-stage procedure was often cited in the cross-over trial reports.

If one chooses to report the results of the statistical test for the carry-over effect, one should be clear about the method used and its significance level. This should be followed by whether the results

of the test have affected how treatment effects are analysed. Similar rules should apply to the period effect.

## 5.9. Results

The results of the systematic review of cross-over trials will be discussed in two parts: designs and analyses. Designs of the cross-over trials include the general background of the trial, such as types of disease, size of the trial, type of cross-over trial and where the trial was conducted. The discussion of analyses of cross-over trials includes statistical methods applied and references used.

Although what has been reported in the cross-over trials published in medical journals may not cover every statistical procedure which had been carried out by the authors, the discussion of the results is based on information given in the reports. It was sometimes feasible to contact the authors for further details of the report. This was, however, not considered. The reasons are, first, that only what has been published would have influence on analyses of cross-over trials for other researchers. Second, omission of information is a common problem that is sometimes considered to be as serious as giving false information. Hence, the analysis presented in this chapter refers only to published information about trials.

### 5.9.1. Diseases

Cross-over designs are only appropriate when the disease of interest is not fatal or curable. Table 5.9-1 shows the distribution of the disease types among the 64 selected reports. More than a

quarter (27%) of the trials were conducted on healthy volunteers. Cross-over trials were most popular amongst trials for behavioural disorders (14%). Other areas of diseases amongst the cross-over trials were endocrine, nutritional and metabolic diseases (11%) and diseases of the circulatory system (10%).

**Table 5.9-1: Disease type of cross-over trial reports reviewed (using International classification of Diseases)**

| | ICD-10 Code | Number of cross-over trials | Percentage of cross-over trials |
|---|---|---|---|
| Neoplasms | C00-D48 | 1 | 2% |
| Endocrine, nutritional and metabolic diseases | E00-E90 | 7 | 11% |
| Mental and behavioural disorders | F00-F99 | 9 | 14% |
| Diseases of the nervous system | G00-G99 | 4 | 6% |
| Diseases of the ear and mastoid process | H60-H95 | 1 | 2% |
| Diseases of the circulatory system | I00-I99 | 4 | 6% |
| Diseases of the respiratory system | J00-J99 | 6 | 10% |
| Diseases of the digestive system | K00-K93 | 2 | 3% |
| Diseases of the musculoskeletal system and connective tissue | M06-M99 | 2 | 3% |
| Pregnancy, childbirth and the puerperium | O00-O99 | 3 | 5% |
| Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified | R00-R99 | 2 | 3% |
| Injury, poisoning and certain other consequences of external causes | S00-T98 | 2 | 3% |
| Factors influencing health status and contact with health services | Z00Z99 | 1 | 2% |
| Healthy (normal ) | | 17 | 27% |
| Cigarette smoker | | 1 | 2% |
| Animal | | 1 | 2% |
| TOTAL | | 63 | 100% |

### 5.9.2. Regions

Seventy-eight percent (n=49) of the cross-over trials reviewed were conducted in Europe or North America. Twenty-eight of the 64 cross-over trials were conducted in Europe and half of the 28 cross-over trials in the UK. Seventeen of the 21 trials conducted in North America were in the US. There were 6 cross-over trials in Asia and 4 in Oceania. One of the trials was an international collaboration between Austria and UK in 1988.

**Table 5.9-2: Regions of cross-over trials reviewed**

| | 1988 | 1992 | 1996 | 2000 | TOTAL |
|---|---|---|---|---|---|
| **Africa** | 1 | 0 | 0 | 0 | **1** |
| **Asia** | 0 | 2 | 2 | 2 | **6** |
| **Europe** | 11 | 6 | 6 | 5 | **28** |
| **Middle East** | 0 | 1 | 1 | 1 | **3** |
| **North America** | 3 | 5 | 6 | 7 | **21** |
| **Oceania** | 1 | 2 | 1 | 0 | **4** |
| **TOTAL** | **16** | **16** | **16** | **15*** | **63** |

** Location of one trial (Appendix B, 2002-2) was unknown.*

### 5.9.3. Design of cross-over trials

Unlike parallel trials, there are many types of design for cross-over trials. Although the type of design is an essential feature of a trial report, it was noticed that the numbers of periods and treatments of the trial were often ambiguous and the design of the trial was rarely specified in the abstract and could only be determined when the methodology was carefully examined. None of the trial reports included the number of treatments and trial periods in the design.

For a large percentage of the reports, designs could only be identified by reviewing number of treatments given to each patient. The most commonly used design amongst cross-over trial reports reviewed was the AB/BA design which had been used by 45 of the 64 reports reviewed. Other designs applied were 3 period 3 treatments and 4 periods 4 treatments.

There were also 12 trials for which the design could not be determined by reading the trial report. For example, the design of a trial that involves two treatments, both taken with or without alcohol was described as 6-way (1992-9). No further information was given on whether all participants were enrolled for 6 periods and how many randomisation groups were used.

There were considerable percentages of cross-over trials which did not specify number of treatments or treatment period. It is possible that many considered AB/BA as a standard design of cross-over trials. Therefore, unless it was stated, it was most likely to be two-period two-treatment design.

**Table 5.9-3: Designs of cross-over trials reviewed**

| Design | 1988 | 1992 | 1996 | 2000 | TOTAL |
|---|---|---|---|---|---|
| 2 treatments 2 periods | 9 | 12 | 11 | 13 | 45 |
| 3 treatments 3 periods | 2 | 1 | 1[c] | 1 | 4 |
| 4 treatments 4 periods | 1 | 1 | | | 2 |
| Not clear | 4[a] | 2[b] | 4[d] | 2[e] | 12 |
| TOTAL | 16 | 16 | 16 | 16 | 64 |

[a] (Letter to the editor 1988-5), (1988-6) (3 treatment groups, but number of treatments and period unknown1988-9) (3 sequences 1988-12)

[b] (1996-4 4 treatments) (1996-9 6 way)

[c] 3 treatment sequence by latin square

[d] (four types of computer assisted tutorial, periods unknown 1996-2)(1996-5)(1996-9)(1996-13)

[e] (2000-12) (2000-13)

### 5.9.4. Length of treatment period

The length of one treatment period varied from a few hours to 12 weeks. The shortest treatment period was 5 hours. The longest trial period was 28 weeks including 4 weeks of wash-out period (1988-11) on 22 diabetes patients and the number of patients who dropped out was 4.

### 5.9.5. Wash-out period

Table 5.9-4 shows that the wash-out period was incorporated between treatment periods by 24 (38%) of the 64 cross-over trial reports. Forty trial reports did not specify whether there was a wash-out period. Eleven (46%) of the 24 reports which gave length of the wash-out period were published in 2000. This may have indicated that there had been an improvement in design or report of cross-over trials or both.

Forty of the 64 trials did not mention the wash-out period. It was not known whether it was for clinical reasons that the carry-over effect was unlikely to occur, or whether it was not feasible to incorporate (e.g. patients needed to be treated at all times), or for practical reasons that adding a wash-out period would lengthen the trial period and cause higher drop out rate, or simply that the authors were unaware of the potential problem of carry-over.

The longest wash-out period was 6 weeks (2000-3) while the shortest one was reported to be 4 days (1992-8). The decision on length of the wash-out period was rarely justified. One report, after the trial was completed, concluded that their wash-out period was too short and that the carry-over effect could not be completely eliminated. Another report did not incorporate a wash-out period in the trial and only found that it should have been incorporated when the results were analysed. They said *"This has implications for the design of the crossover intervention trial and suggests that a wash-out period should be incorporated between treatment periods to minimize carry-over effects"(1996-5).*

**Table 5.9-4: Use of wash-out period in cross-over trials reviewed**

| | 1988 | 1992 | 1996 | 2000 | Total |
|---|---|---|---|---|---|
| Number of trials that incorporate a **wash-out period** | 5 | 4 | 4 | 11 | 24 |
| Wash-out period **NOT** mentioned in the report | 11 | 12 | 12 | 5 | 40 |
| Total Number of Reports | 16 | 16 | 16 | 16 | 64 |
| Proportion | 5/16 | 4/16 | 4/16 | 11/16 | 24/64 |
| CI wash-out period | (0.1416, 0.556) | (0.1018, 0.495) | (0.1018, 0.495) | (0.444, 0.8584 | (0.2667,0. 4975) |

### 5.9.6. Drop-out rate

A cross-over trial is more likely to have a higher drop-out rate than a parallel trial. The problem with a high drop out rate is that it may result in inadequate power.

The number of patients who dropped out was specified in only 9 of the 64 trials. It is not known whether it was because there were no drop outs in the trial or whether this information had been omitted. The number of patient drop outs was compared with the number of patients included in the final analysis, and one case (1992-12) showed that only 6 of the 14 patients completed the trial (such a high drop out rate was not mentioned in the text in the report). A clinical trial exhibiting such a low completion rate raises questions about the design of the trial. Since less than half of the participants completed the trial, it is debatable whether a statistical analysis was worth carrying out and whether results of the analysis (or the trial) were validated or published without discussing its drop out rate.

The second highest drop-out rate occurred in a multiple sclerosis trial which lasted 8 weeks for both periods with a two week wash-out period in between. Five of the 26 patients randomised to trial dropped out before it was completed.

Although it seems plausible that a longer period for the trial can cause a higher drop out rate, it is not possible to establish such a relationship with such a small sample size. Other possible causes

of patient drop out include treatments which are less tolerable or treatments which produce effects that were not noticeable to the patients. This might happen in the case of diseases which do not necessarily require treatment or treatments which are not effective. In addition, trials which require regular visits to the hospital are more likely to cause patients drop-out than trials which require short period in-patient hospital stay.

### 5.9.7. Statistical analysis

The method used for the statistical analysis is a required disclosure in most medical journals. Results of this section are based on the description of statistics in the cross-over trial reports published in medical journals.

The most commonly used methods described in cross-over trial reports are listed in Table 5.9-5. It is regrettable that no further conclusion can be drawn from the review on statistical analyses of cross-over trials published in medical journals other than from the statistical methods mentioned in the reports. It was impossible to judge whether the statistical methods applied were appropriate for the trial as statistical tests were often given without mentioning to which variable they were applied to. For example:

> *"The statistics used for paired comparison were the McNemar test with Yates correction, and the paired Wilcoxon test. The Mann-Whitney U-test was used for comparison of unpaired data. Exact 95% confidence intervals (CI) are given in parentheses in the text." (1996-1)*
>
> *"All analyses were performed using either SAS software programme (SAS Institute, Cary, NC) or Microsoft 4.1 (Ecosoft Inc., Indianapolis, IN). Analysis of*

*variance and paired t tests, or chi-squared and Fisher exact tests, were used to assess continuous or categorical variables respectively. All tests were two-tailed and alpha levels of $<0.05$ were considered significant. Mean ± SE are presented." (1996-6)*

Despite the requirement of journals, there was one report each in year 1992, 1996 and 2000 where no other statistics were given beyond mean and standard error.

The most commonly used methods were t-test and ANOVA. Other methods which have been used include two-sample t-test, Wilcoxon signed ranked test, Mann-Whitney, Fisher's exact test, Chi-squared, MANOVA, regression analysis and general linear model.

**Table 5.9-5: Most used methods of statistical analyses for cross-over trials reviewed**

| | 1988 | 1992 | 1996 | 2000 | Total |
|---|---|---|---|---|---|
| **ANOVA** | 4 | 4 | 5 | 4 | **13** |
| **Paired t-test** | 10 | 3 | 5 | 3 | **21** |
| **Wilcoxon signed rank test** | 6 | 2 | 3 | 3 | **14** |
| **Hills and Armitage** | 1 | 1 | 1 | 1 | **4** |

**Testing for the carry-over effect**

Although testing for the carry-over effect is the key feature of the two-stage procedure, it is not confirmed that it necessarily implies the two-stage procedure had been used. It has been stated in many reports that the carry-over effect was not significant or did not exist without giving the test used. The followings are quotes from the trial reports which mentioned the carry-over effect without giving the method of the test.

*"In each study pattern, similar values for the clinical and metabolic findings, vitamin E plasma levels, platelet aggregation tests, and TXB2 production were observed after each vitamin E and placebo period, thus excluding any carryover effect." (1988-8)*

*"There was no evidence of either a carryover effect or of a period-treatment interaction" (1992-10)*

*"No sequence effect was demonstrated with respect to the different dialytic therapies (RHSD and SD) in any of the outcome measures" (1996-6)*
*"No carryover effects were demonstrated, and baseline values did not differ significantly" (2000-8)*

*"Analysis is difficult, as adjustment has to be made for both period effect and treatment period interaction" (2000-10)*

*"This has implications for the design of the crossover intervention trial and suggests that a wash-out period should be incorporated between treatment periods to minimize carry-over effects"(1996-5)*

**The two-stage procedure**

The 'two-stage procedure' was only mentioned in two of the 64 reports. One report stated that the statistical method used for analysis of the cross-over trial followed the method published by Hills and Armitage (1979) while the other report followed the book by Pocock (1984), which recommended the two-stage procedure, see Chapter 6 - .

Although testing for the carry-over effect was an important feature of the two-stage procedure, it has been mentioned more frequently than the method itself. It seems that testing for the carry-over effect had a high priority in analyses of cross-over trials. Below is an example in which, despite the wash-out period between two treatment periods, testing for the carry-over effect was considered to be more important than the deficiencies of such a test.

The trial was conducted by Harvard Medical School with one of the authors from the department of biostatistics. This report cited

Freeman (1989) and used a model suggested by Senn (1994), which excluded the carry-over effect from the analysis. Despite acknowledging that models including the carry-over effect "*are biased and possess poor power*" and incorporate a wash-out period between two treatment period, it was stated that "*in any event, a carryover effect seemed unlikely with our design because evidence from kinetic studies with testosterone cypionate suggest that 6-week washout period after 6-week treatment period allows neuroendocrine function to return to baseline. The observed carry-over effect raised some concerns despite the 6-week washout period which was designed to 'allow neuroendocrine function to return to baseline*". The statistical analysis was conducted on the first treatment period only. Their rationale for such an analysis was "*free of concern about carryover effect*". This example illustrates the fear of the carry-over effect for clinicians and may explain why the two-stage procedure has been widely cited (Chapter 4 - ).

In contrast to testing for the carry-over effect without mentioning the method used or whether it was used as part of the two-stage procedure, the two-stage procedure has been cited without giving the result of the test for the carry-over effect. For example, 2000-16 cited Hills and Armitage (1979) without mentioning the result of the carry-over effect in discussion.

Although the carry-over effect was rarely mentioned in the description of analysis, it was often mentioned that the carry-over was not significant without indicating methods of tests. Results of

this review indicated that the use of the two-stage procedure cannot be estimated and it is possible that the two-stage procedure has not been fully understood or applied for analyses of cross-over trials.

### 5.9.8. Cited references

In addition to the statistics methods described in the report, citations and references were used as indicators of sources of statistics. Reference served the purpose of confirming the method used for analysis and establishing the relationship between citation and actual analysis.

Fifteen of the 64 papers provided references to statistics. There were 4 citations on Hills and Armitage [7] while Grizzle [4] was not cited by any of these reports. Similar to the previous conclusion that the carry-over effect could still have been tested even it had not been stated in the statistics description, not citing any of the documents relating to the two-stage procedure does not necessarily imply that it had not been used. It was found that the carry-over effect was reported to be non-significant while no statistical method was given and none of the two-stage procedure references was cited. It can be confirmed that the two-stage procedure was more commonly used than what was found in this systematic review (Table 5.9-5). Using citations as an indicator of whether the two-stage procedure had been used may have underestimated the prevalence of its use.

However, citation does not necessarily indicate usage. There were papers which cited Hills and Armitage [7] and did not report if the carry-over effect had been tested, or at what level or whether it was insignificant. None of the trials reported that treatment effects were evaluated using the first period data only because the hypothesis of equal carry-over effects was not accepted

**Table 5.9-6: Citations of cross-over trial reports reviewed**

| | 1988 | 1992 | 1996 | 2000 | TOTAL |
|---|---|---|---|---|---|
| **Number of cross-over trial reports that cited statistical references** | 5 | 1 | 4 | 5 | 15 |
| Number of citations to Hills and Armitage | 0 | 1 | 1 | 2 | 4 |
| Citations give to Freeman | 0 | 0 | 0 | 1 | 1 |
| Citations give to Senn | 0 | 0 | 1 | 1 | 2 |

**Table 5.9-7: References for statistical analyses of cross-over trial reports reviewed**

| Reference | 1988 | 1992 | 1996 | 2000 |
|---|---|---|---|---|
| Cohen, Statistical power analysis for the behavioural science. New York: Academic Press, 1969 | 1 | | | |
| Miller RG. Simultaneous statistical interference. New York: McGraw Hill, 1966:62 | 1 | | | |
| Altman DG, Gore SM, Gardner MJ, Pocock SJ: Statistical guidelines for contributors to medical journals, Br Med J 286: 1489-93, 1983 | 1 | | | |
| Hollander, M., Wolf, D. A., Nonparametric Statistical Methods, J. Wiley Sons, New York (1973) | 1 | | | |
| Pocock SJ. Clinical trials: a practical approach. New York: Wiley, 1983 | 1 | | | |
| Altman DG. Clinical trials; In: anonymouse. Practical statistics for medical research. London, Chapman & Hall, 1991: 440-75 | | | 1 | |
| Hills and Armitage 1979 | | 1 | 1 | 2 |
| Wallenstein S, Fisher AC. The analysis of the two-period repeated measurements crossover design with application to clinical trials. Biometrics 1977; 33: 261-9 | | | 1 | |
| Senn S. In cross-over trials in Clinical Research. New York: John Wiley & Sons, 1993 | | | 1 | |
| Steinijans VW, Hauschka D: Update on the statistical analysis of bioequivalence studies. Int. J Clin Pharmacol Ther Toxicol 1990; 28: 105-110 Hauschke D, Steinijans VW, Diletti E: A distribution-free procedure for the statistical analysis of bioequivalence studies. Int J Clin Pharmacol Ther Toxicol 1992; 30 (suppl. 1):S37-S43 | | | | 1 |
| Stolley PD, Strom BL. Sample size calculations for clinical pharmacology studies. Clin Pharmacol Ther 1986; 39: 489-90 | | | | 1 |
| 1. Senn S. The AB/BA crossover: past, present | | | | 1 |

| | | | | |
|---|---|---|---|---|
| and future? Stat Methods Med Res. 1994; 3:303-324<br>2. Freeman PR. The performace of the two-stage analysis of two-treatment, two-period crossover trials. Stat Med. 1989; 8:1421-1432<br>3. SAS Institute Inc. SAS/GIS Software [computer program]: release 6.12. Cary. NC: SAS Institute Inc; 1997 | | | | |

However, the fact that a trial report cites the papers describing the two-stage procedure and the books which recommended that method can be used to check the consistency of the statistical methods used in the report. Citations can only be used as an additional confirmation that the method was used.

It can be concluded that the use of the two-stage procedure had in fact exceeded the number of its citations. Whether the two-stage procedure was more commonly used than the original simple cross-over difference remains unknown. It is also impossible to conclude whether not mentioning the two-stage procedure was a result of Freeman's [8] paper on its deficiencies. Such a possibility was, however, rather slim as the difference in citation rate between Freeman [8] and other references of the two-stage procedure is substantial.

The results of this systematic review also show that Freeman's [8] influence has not extended to the analysis of cross-over trials published in medical journals. Despite the high citation rates of the two-stage procedure in medical journals, the method was likely to be unknown to many clinical researchers. Therefore the change in the validity of the two-stage procedure would not have had any

impact on those who were not aware of this method in the first place.

### 5.9.9. Sample size calculation

Sample size calculation has always been a common practice of experimental design and a basic requirement in medical journals. It has, however, become a controversial issue in recent years [27].

Sample size calculation was originally planned to be included in the results of the systematic review. However, in order to reflect the personal view on recent discussion of the sample size calculation, it has been removed from the discussion of results.

### 5.9.10. Discussion on individual papers

Despite the inconclusive results regarding the use of the two-stage procedure in analysis of cross-over trials, this systematic review has provided an opportunity to examine published cross-over trial reports. Although it is not the focus of this thesis, it was found that a few papers have raised some concerns regarding different issues of clinical trials while reviewing the reports. Below are three papers which raised issues with ethics, randomisation and design respectively.

### 5.9.11. Ethics

As clinical trials are experiments on human beings, the ethical issues should be carefully reviewed so that any potential harm to the patients' physical and mental health may be avoided. It was

found that there was a cross-over trial in which patients' health was at risk. This report, which caused serious concern and should never have been allowed by the relevant 'ethical committee', was 1992-6 on salt intake. It was a trial with the objective of examining whether the intake of salt would worsen the condition of asthma patients, although it has been acknowledged in the paper that a high sodium intake had been found to increase bronchial reactivity with asthma patients. Participants in the trial were asked to add as much salt as possible to their regular diet and to take four tablets of sodium chloride (500mg) daily. The trial consisted of three treatment periods of different levels of salt intake. The paper was accepted by *Thorax*, a journal with a high impact factor of 4.09.

This trial raised a few concerns. First, it is a trial without a justified objective. There was no benefit for human beings to increase sodium chloride intake and a reduction in its intake has been widely encouraged. Whether a high consumption of sodium chloride would worsen the condition of asthma was irrelevant as it may cause other problems such as high blood pressure and was certainly not worth exploring by putting patients at risk. The harmful effect of sodium has long been established and the upper limit of UK daily recommended nutrition intake (RNI) is 1,600 mg (2,400 mg for US sodium RDA). The most common problem associated with high sodium diet is raising blood pressure [41]. In addition to the trial design that required patients to *"add as much*

*salt in their food as possible"*, participants were asked to take four tablets of 500mg of sodium chloride everyday.

Another issue with this trial is that blinding/masking had not been introduced. Instead, patients were deceived with the intention of blinding/masking (*"To minimise bias, patients were told that the response of an individual to a change in dietary salt intake was unpredictable."*). They were told that *"a low and high dietary salt intake for two weeks has no effect on peak expiratory flow in patients with mild asthma"*.

This trial was not only unethical but unnecessary. There were other possibilities to study whether reduced sodium chloride consumption would help with the asthma condition. A suggested design of this trial would include two treatment periods only. One treatment period would involve patients' usual diet and one treatment period with a lower sodium diet. A diet in high sodium chloride should have never been encouraged anyway.

### 5.9.12. Randomisation

Randomisation is generally agreed to be highly desirable for clinical trials [31][32].

It was found that one cross-over trial (2000-2) did not incorporate randomisation into the trial of which the objective was to determine the effect of the consumption of a fat substitute (olestra) in measurements of fecal fat excretion. The two treatments to be compared were conventional potato chips and treatment chips

containing olestra. The trial was carried out in two centres with 5 participants each. All 10 patients were assigned conventional chips first, followed by chips containing olestra. This is a serious design flaw as treatment was confounded with period, even though there does not seem to be any practical difficulty in incorporating randomisation into the trial.

### 5.9.13. Design

Cross-over trials are special designs requiring that the condition of patients stay stable throughout the trial periods. However, it was found that some trial reports did not meet this requirement and the cross-over design was not the optimum type of clinical trial on such an occasion.

The example was taken from (1996-11). The trial compared two types of oral analgesia in relieving the intensity of postoperative pain. The two treatment periods took place in the first and second day after the operation. Each treatment lasted for 5 hours. Postoperative pain intensity is unlikely to remain stable throughout both treatment periods. Such a design is very vulnerable to period effect.

## 5.10. Conclusion

The original objective of the systematic review was to investigate the use of the two-stage procedure in medical journals. However, results of this systematic review showed that the reporting and the description of statistics methods used in the cross-over trial reports

reviewed was insufficient to assess the prevalence of the two-stage procedure in medical journals. It can only be concluded that the method was likely to be used more often than the number of citations given to the papers describing the two-stage procedure would suggest.

The reason for this was that testing for the carry-over effect was often mentioned without giving details of the method and results. It can therefore not be confirmed that the two-stage procedure was carried out fully.

This systematic review has shown that the design and analyses of cross-over trials has not been fully understood by all the clinical researchers who conducted the trials. The cross-over trial should only be used when its strength can be fully utilised and weakness carefully avoided.

It also worth mentioning that the quality of reporting often affects results of designs and analyses in the review. Therefore it is important that methods and results of statistical tests should always be reported in medical journals.

The inconclusive result of the systematic review implies that more concise reports, better understanding of cross-over trials and analyses are needed to achieve good quality science. One probable solution to this would be more collaboration between clinicians and statisticians. (Recommendations of information that should be included in cross-over trial reports is given in Chapter 8 - ).

# Chapter 6 - Analyses of cross-over trials in general medical statistics books – a comprehensive review

## 6.1. Introduction

General medical statistics books have been criticised for their errors in fundamental concepts of statistics, for example the interpretation of p values, the confusion of the two sample t test and the presentation of ordered data in chi-squared contingency tables to name just a few [33]. The analysis of cross-over trials has a more turbulent history than other statistics topics covered in general medical statistics books. It has been shown in the previous chapters that the two-stage procedure has been used by medical researchers while its deficiencies were rarely mentioned. Methods used to analyse cross-over trials are often influenced by recommendations given in medical statistics books. This chapter focuses on the statistical methods recommended to analyse cross-over trials in general medical statistics books.

## 6.2. Objective

The objective of this chapter is to investigate statistical methods recommended for analyses of cross-over trials in general medical statistics books.

## 6.3. Methodology

The methodology used was a comprehensive review of all general medical statistics books published in English. First, a list of a

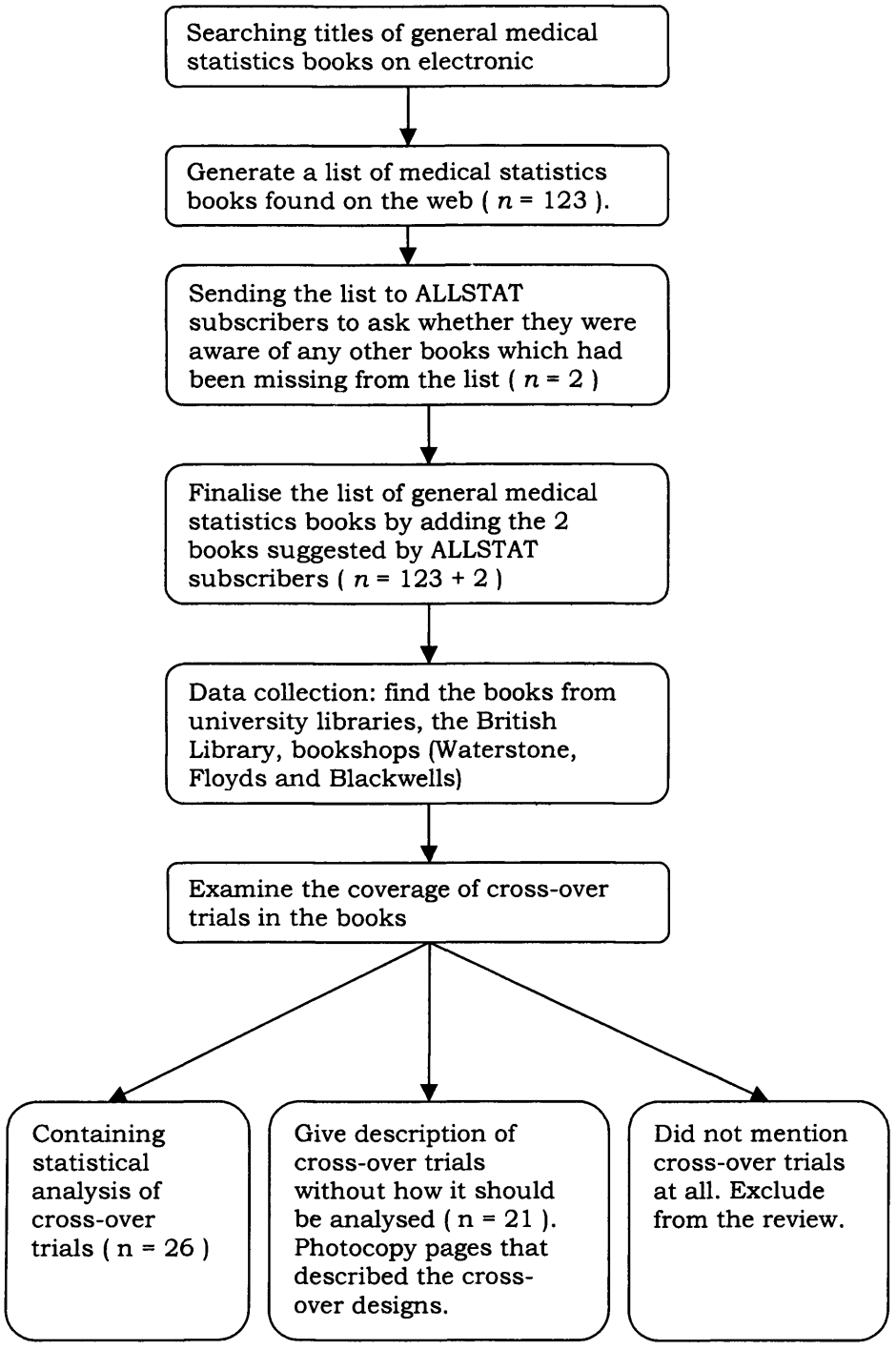
complete selection of textbooks was gathered by searching the key words 'statistics' and 'medical' on several web-sites, such as online bookshop www.amazon.com and university libraries. The list was sent to statisticians on ALLSTAT (an UK-based worldwide e-mail broadcast system for the statistical community) with a request to add any relevant titles omitted from the list. The final list consisted of 125 general medical statistics books. All of them were reviewed for their statistical methods recommended for analysing cross-over trials.

Publication dates of the 125 medical statistics books reviewed ranged from 1963 to 2000, and covered the whole duration of the development of the two-stage procedure (from the first paper that published such method in 1965 to the finding of its major deficiency in 1989). Although two books published three decades apart can hardly be compared, the concepts and the basic statistical methods remained the same during this period. Therefore, many of the early published medical statistics books remained popular and can usually be found in book shops and libraries. Analysis of cross-over trials is an exception amongst other commonly covered topics in general medical statistics books, the standard operation having changed three times in the previous four decades (detailed in Section 1.2.2).

All 125 books were reviewed for their coverage of the cross-over design. They were categorised into three types: books with statistical analyses of cross-over trials, books with a description of

the cross-over design but without analyses and books not mentioning cross-over trials at all. This comprehensive review included all books which covered statistical methods of cross-over trials (See Appendix C List of medical statistics books reviewed). The cross-over design was described in 47 books and 26 of them gave recommendations of statistical methods to analyse cross-over trials. The results given below are based on those 26 books.

**Figure 6.3: Flow diagram to show how the 26 books finally examined were obtained**



## 6.4. Results

It was found that some descriptions/statements in general medical statistics books were ambiguous. As a result, they may be perceived differently by different readers. In order to illustrate an objective discussion of the books, results of the comprehensive review are presented in their original quotes.

### 6.4.1. The cross-over design - General background

Many authors of general medical statistics books held a sceptical view on both the design and the analysis of cross-over trials. The two main causes of concern were the carry-over effect and the potential of higher drop-out rates. For example, "Because of the problems described, crossover studies are probably overused." (Altman 1994). The problem can in fact, be dealt with in most cases if the cross-over trial is an appropriate design for the research to be undertaken. The solution to the former problem is to allow for a sufficient wash-out period, and for the latter one it is to fit the patient as a random effect. The solutions, especially the latter one is rarely offered to readers. In addition, analysis of cross-over trials was thought to be less straight-forward.

> *"The actual design and analysis are best carried out by a professional statistical advisor, but nevertheless some of the main types of design available will be described in general terms" (Goldstone 1983)*

Lack of consensus in analysing cross-over trials could be the reason why general medical statistics books merely gave definitions of cross-over trials without going into details of their analyses.

There was also a tendency to prefer parallel over cross-over design. Bolton, (1997) reflected this:

> *"Properties of a particular design should be carefully considered before the final choice is made. For example, the presence of carryover effects will negate the advantage of a crossover design as presented ..However, caution should be used when considering this design in studies where carryover effects or other interactions are anticipated. Under this circumstance, a parallel design may be more appropriate."*

The key advantage of cross-over designs, the within-patient comparison, was rarely mentioned. The only exception was in reduction of sample size. *"Therefore the only advantage of such an intra-subject comparison is reduction of sample size."* (Mainland and Sanders, 1963)

### 6.4.2. The carry-over effect

The term 'treatment-period interaction' was also used to refer to the carry-over effect. The main question in analyses of cross-over trials was whether or not to test for the carry-over effect. The key feature of the two-stage procedure was to test for the carry-over effect. It was not uncommon to see books that insisted on the carry-over effect tested by statistics, and it often led to the use of the two-stage procedure. In addition to the preliminary t-test, it was shown how carry-over effects can be identified without carrying out a proper statistical test:

> *'The assessment of order effects must be based on a statistical model..Therefore, unless order effects are known to be negligible, the crossover design loses its advantages' (Bailer, Frederick, Mosteller, pp 90, 1992).*

In addition to testing for the carry-over effect as the first step of the two-stage procedure, a common recommendation to identify the carry-over effect was to compare whether the responses from both first and second periods were the same. Here are some examples of suggestions of how that is done.

> *"The subjects showing the same response in both periods give no information about the difference between drugs but they do give information about whether the order or administration matters" (Colquhoun, 1971).*
>
> *"If the patient is improved by both treatments or not improved by either he contributes no information about a possible difference between the two treatments, and consequently, in examining the results of a crossover trial of this kind, we must restrict ourselves to considering those pairs of responses in which one treatment produces improvement and the other does not" (Smart, 1963).*
> *"A carry-over effect is present when the true treatment effect is different for subjects in group A than for subjects in group B" (Rosner, 1995)*

These statements were vague and most importantly, only applied when endpoints were binary data. All three books quoted above failed to identify the type of the data that applied to these recommendations.

### 6.4.3. The two-stage procedure

In order to offer readers a solution to the problem of the carry-over effect, the two-stage procedure was often recommended. Nine of the 26 books reviewed recommended the two-stage procedure for analysing cross-over trials. Only one (Pocock, 1987) of these 9 books was published before the high type I error rate was found by Freeman in 1989; the other 8 books were all published between 1990 and 1998.

**Use of the two-stage procedure: N=9**

Books which recommended the two-stage procedure usually considered it to be the correct and optimum method for analysing cross-over trials. The original quotes of the 9 books which recommended the two-stage procedure were extracted and presented below.

> *"It is incorrect to ignore the design of the study and just perform a simple comparison of treatments. The correct analysis consists of three two sample t tests or Mann-Whitney tests. For categorical data we use Chi-squared tests." (Altman, Practical statistics for medical research 1990)*

> *"A test for the existence of a TP interaction is obtained as follows. If there is no TP interaction, the sum of the two responses for subject j in group I, $e_{ij}=y_{ij1}+y_{ij2}$, should (apart from random error) be the same whether the subject is in group I or group II If there is a TP interaction, there will be little point in testing and estimating the treatment effect from the whole set of data". (Armitage and Berry, Statistical Methods in Medical Research, 1994)*

> *"The general procedure for comparing direct treatment effects is to first perform a preliminary test for the equality of the carryover effects." (Buncher, Tsay, Statistics in the Pharmaceitical Industry, 1994)*

> *"If present (the carry-over effect), using only data from period 1 to estimate treatment effect. ..Suppose however, there had been evidence of a non-zero carryover effect. Then the previously described estimators for treatment and period effect would be biased and so could not be used. In such cases, Hills and Armitage (1979) advise that the treatment effect now be estimated from the first period observations only." (Everitt. Statistical Methods for Medical Investigations, 1994)*

> *"What can we do if we identify a significant carry-over effect using ..? In this case, the second-period data are not useful to use since they provide a biased estimate of treatment effects usually for subjects who were on active drug in the first period and on placebo in the second period, and we must base our*

*comparison of treatment efficacy on period-1 data only" (Rosner. Fundamentals of Biostatistics, 1991)*

*"If a significant interaction is found, one's best policy is to abandon the above within-patient analysis and resort to a between-patient comparison of treatment using the first period only." (Pocock. Clinical Trials – A Practical Approach 1987)*

*"If there is reason to suspect a carry-over effect, the only way to salvage some of the data is to analyse the results in parallel groups for the initial period of the trial only, disregarding the subsequent 'contaminated' periods" (Spriet, Dupin-Spriet, Simon. Methodology of Clinical Drug Trials, 1998)*

*"Thus, it is of interest to perform some preliminary tests for the presence of the period effect and/or the carry-over effects before the comparison of bioavailability between formulations is made. ..To increase the test power, however, Grizzle (1965) suggested testing the null hypothesis at the $\alpha$=10% level instead of the traditional 5% level." (Chow, Liu. Design and Analysis of Bioavailability and Bioequivalence Studies 1992)*

*"Grizzle has published an analysis to detect carryover (residual) effects. When differential carryover effects are present, the usual interpretation and statistical analysis of crossover studies are invalid. Only the first period results can be used, resulting in a smaller, less sensitive experiment. An example of Grizzle's analysis is shown in this chapter in the discussion of bioavailability studies (sec. 11.4.2). Brown concludes that most of the time, in these cases, the parallel design is probably more efficient. Therefore, if carryover effects are suspected prior to implementation of the study, an alternative to the crossover design should be considered." (Bolton. Pharmaceutical Statistics – Practical and Clinical Applications, 1997)*

## Versions of the two-stage procedure

Despite the similarity in methods of analysis and description of the problem, the completeness of the two-stage procedure differed from book to book. Some books gave clear indications that the analysis should be based on the first period data only. Others simply stated

that the cross-over design was not appropriate for the trial. In other words, no conclusion of the treatment effect could be drawn and the trial would have been conducted in vain.

The difference in the versions of the two-stage procedure could be traced to the two important papers covering such analysis. The original paper which proposed testing the carry-over effect was by Grizzle [4] which suggested that the design was not suitable if the carry-over effect was to be found statistically significant ("*If residual effects are thought to be unequal, the change-over design should be avoided unless residual effects are of interest in their own right and are not regarded as a nuisance.*"). The best-known (results from the citation analysis in Chapter 4 - ) publication of the two-stage procedure was published by Hills and Armitage [7]. They completed the two-stage procedure by suggesting that the treatment effect be estimated using the first period data. Both Grizzle [4] and Hills and Armitage [7] versions of the two-stage procedure were reflected in general medical statistics books. It is conceivable that with the possibility of not gaining any information about the treatment effect from the cross-over trial, Grizzle's version of the two-stage procedure was more likely to discourage some clinical researchers from using the cross-over design. Although 11 of the books cited Grizzle [4] or Hills and Armitage [7], the name originally used by Grizzle [4] was not referred to by any of those books.

**Significance level of the preliminary test**

Due to its low power, Grizzle [4] set the significance level of the preliminary test of the two-stage procedure at 10%. It was one of the most important features of the procedure. However, amongst the 9 books which recommended the two-stage procedure, only 3 books (Rosner 1991, Chow & Liu 1992, Buncher & Tsay (eds) 1994) specified the 10% nominal level for the carry-over effect.

It is most likely that the 5% significance level would otherwise have been used if the 10% level had not been indicated. There is one book (Everitt, 1994) which, despite citing Armitge [7], gave 5% as the significance level regardless of what was recommended in the original papers. This shows that general medical statistics books cannot always be relied on for a correct or complete analysis of the two-stage procedure.

**Not using the two-stage procedure N=7**

Seven of the 25 books reviewed did not recommend the two-stage procedure for analysing cross-over trials. These included books which did not mention testing for the carry-over effect at all and those who explained why it should no longer be used. There were 5 books (Jennison, Turbull. Group Sequential Methods with Applications to Clinical Trials, Smart. Elements of Medical Statistics 1993, Chow,editor. Encyclopedia of Biopharmaceutical Statistics, Zolman. Biostatistics – Experimental Design and Statistical Inference, Goldstone, 1983) which did not describe how carry-over effects could be tested statistically. Their methods for

analysing cross-over trials were usually the simple cross-over difference for the treatment effect for continuous endpoints. For example:

> *"As with the cross-over trial, the analysis for matched pairs is fairly simple and lends itself to a technique known as sequential analysis."(Goldstone. 1983)*

The two books which gave the most complete information regarding analysis of cross-over trials were Matthews (2000) and Chow (editor, 1992). They both all mentioned the two-stage procedure and explained why it should not be used.

> *"The approach is no longer recommended. There are several problems .".(Matthews. An Introduction to Randomized Controlled Clinical Trials)*

Although not falling in the category of the general medical statistics, it is worth mentioning that there is a book: Guiloff, Clinical trials in neurology (2001) went a step further to warn readers what they might come across in other medical statistics books. This is considered highly informative. Because of the popularity of the two-stage procedure, the awareness of its deficiency can only be raised by statements like this:

> *"As a result of a paper by Freeman, medical statisticians working on the methodology of cross-over trials have now abandoned the so-called two-stage procedure. The reader should be warned, however, that general introductory text-books on medical statistics are still being written which recommend this biased procedure. In fact, most introductory textbooks cannot be trusted as regards the advice given on this issue". (Guiloff. editor- Clinical Trials in Neurology 2001)*

**Not clear N=2 (Books which did not specify the second stage of the analysis)**

Two books which could not be categorised into either of the groups mentioned above were by Mainland & Sanders and by Ingram and Bloch.

> *"In order to find out whether the sequence per se had any influence on the outcome, we use from each patient the sum of his changes, period 1+ period2." (Mainland & Sanders. Elementary Medical Statistics. 1963)*

> *"If such time period difference is found, then the interpretation of the trial becomes rather difficult; interpretation depends more on the medical than the statistical pre-supposition. As a technical point, the statistical test to tell whether the results differ in the two time periods is not very powerful, so that merely deciding the difference between results is 'not statistical significant' may be misleading." (Ingram, Bloch. Mathematical Methods in Medicine. 1983)*

However, Ingram and Bloch continued to recommend the references by Hills and Armitage [7] and Armitage and Hills [49]. Readers of this book would be inclined to use the two-stage procedure.

### 6.4.4. Other: Two atypical books

There were books that were not typical general medical statistics books: One was the proceedings for a symposium and the other was a systematic review of cross-over trials. However, their descriptions gave some indications of the type of analysis they would consider suitable for cross-over trials.

In the book edited by Bithell and Coppi (1981), cross-over designs were discussed in the chapter on repeated measurements. And it was suggested that

> *"One possible approach is the introduction of so-called "carry-over" effects, e.g. the effect of treatment A given in the first period interacting with the effects of other treatments in the second period, third period etc. It should be noticed that these carry-over effects are confounded with other interactive effects between treatments and period and must be interpreted very carefully." (pp169, Bithell, Coppi 1981)*

The chapter on cross-over trials in the book edited by Bailer et al was a report of crossover studies that appeared in the New England Journal of Medicine between 1978 and 1979. Both Grizzle [4] and Hills & Armitage [7]) were cited in the chapter and the authors commented that

> *"Unless order effects are known to be negligible, the crossover design loses its advantage. The assessment of order effects must be based on a statistical model." (Bailer, Mosteller. 1992).*

### 6.4.5. Deficiencies of the two-stage procedure in general medical statistics books

Ideally, the most up to date and fully informative book on the analysis of cross-over trials would describe the two-stage procedure and explain why it should no longer be used so that readers did not only learn the correct method for analysing cross-over trials but were aware of what might be seen in the literature. Nevertheless, such suggestions could not be expected from books published before 1989. There were only three books which met all these criteria: two of them were published in the year 2000 and one in the year 2001.

It is essential for a book to provide information on the deficiencies of the two-stage procedure if the procedure is to be discussed.

However, with the exception of three books, all the authors of the books reviewed failed to do so and the two-stage procedure remained as recommended without giving sufficient warning to readers. Although readers of the medical statistics books which did not mention the two-stage procedure may also use the simple cross-over difference to evaluate the treatment effect, the book was also likely to be perceived as not giving enough information, and readers who were aware of the existence of the two-stage procedure may resort to other references. Therefore, the most informative books regarding the analysis of cross-over trials were those which give information about the two-stage procedure and why it should not be applied any more.

### 6.4.6. The coverage of the low power of the two-stage procedure

The low statistical power of the two-stage procedure was mentioned in Hills and Armitage [7], yet it had never been considered a serious enough issue for the two-stage procedure to be discouraged. Seven of the 25 books reviewed gave some information about the low power of the two-stage procedure and five were from books which recommended the two-stage procedure.

> *"A problem with the analysis of crossover trials is that the important test for a possible treatment-period interaction is noted for its lack of statistical power". (Altman, Practical statistics for medical research. 1990)*

> *"Unfortunately, this comparison is subject to between-subjects variation, and is therefore less precise than the full crossover approach. .. A similar drawback applies to the test for interaction. This is based on between-subjects variation,*

*and if the latter is very large the t test will be relatively insensitive." (Armitage, Berry. Statistical methods in medical research.1994)*

*"The preliminary test has comparatively low power." Buncher, Tsay. Statistics in pharmaceutical industry.1994)*

*"This test usually has less power than the cross-over efficacy test, or alternatively requires a greater sample size to achieve a given level of power." (Rosner. Fundamentals of biostatistics. 1991)*

*"Unfortunately, this test for interaction is not very sensitive since it is based on between-patient comparison, so that particularly in small crossover trials one may fail to detect an interaction when it is present." (Pocock. Clinical trials – a practical approach. 1987)*

*"Unfortunately, to detect the influence of carry-over effects on the results, we must use a method based on the variability between individuals "(Spriet, Dupin-Spriet, Simon. Methodology of clinical drug trials. 1994)*

*"This approach is no longer recommended." (Matthews. An introduction to randomized controlled clinical trials 2000)*

### 6.4.7. The coverage of the type I error rate of the two-stage procedure in medical statistics books

The issue of type I error rate is rarely mentioned in medical statistics. Amongst all 16 books published after 1989, only 1 book (J.N.S. Matthews 2000, An Introduction to Randomized Controlled Clinical Trials) informed the readers about the problem with the type I error rate.

*"This approach is no longer recommended. There are several problems. The most transparent one is that the $\xi_i$ (patient effect) have not been eliminated from the test of $\gamma=0$. Therefore, this test is affected by between-patient variation. This is likely to be large and as the size of the trial would be determined by a sample size calculation based on the smaller variance $\sigma$, it is likely that the test of $\gamma=0$ has poor power. Consequently, the decision to follow the procedure in section*

*10.3 (crossover difference) may well be taken even in the presence of substantial non-zero value of γ. A more subtle problem with the approach is that Step 1 is not what it seems. The first period of the trial appears to be a parallel group trial, so the difference in treatment means from the first period only should be an unbiased estimator of τ (treatment effect). However, if the analysis is only performed when the null hypothesis γ=0 has been rejected, the analysis will be biased. This is because γ=0 is based on $S_i=X_{i1}+X_{i2}$, which is highly correlated with $X_{i1}$. Consequently, if the difference in the treatments means based on $X_{i1}$ is computed only when there is a noticeable difference between the groups of $S_i$, the estimate cannot be expected to be unbiased. The recommended approach is not to use this particular crossover design when there is a possibility of a carryover effect. You should try to use non-statistical arguments, perhaps based on the half-lives of drugs etc., to decide how long treatment effects are likely to persist. The AB/BA design can then be used if the treatment periods are separated by 'washout periods' whose duration is sufficient to ensure that carryover cannot occur." (Matthews. An introduction to randomized controlled clinical trials 2000)*

### 6.4.8. Sample Size

Although sample size calculation is a common requirement for clinical trial reports, it is not discussed here. The reason is the same as in Section 5.9.9.

### 6.4.9. Patient withdrawal

Patient withdrawal was considered to be very problematic for the cross-over design. A high drop-out rate may cause a deficiency in the same size that is required to achieve the power it was originally designed for. However, the drop out rate should also be considered when calculating sample size. In the analysis of cross-over trials, it is possible to fit a term for patient as a random effect. By doing this information on patients who drop out is not wasted. This approach

to the analysis of cross-over trials was rarely mentioned in general medical statistics books.

> *"Crossover trials are particularly vulnerable to the effects of patient withdrawal. If a patient withdraws after the first period they cannot be included in the analysis because they never received the other treatment." (Practical statistics for medical research, Altman, 1990)*

### 6.4.10. Misconceptions

There were also some misconceptions of the cross-over design, for example *"Finally, all these problems can be made much worse when testing three or more treatments"* (Ingram, Bloch. Mathematical methods in medicine, 1983)

The more complicated design (more periods and treatments) of cross-over trials is the alternative to the AB/BA design, as the carry-over effect can usually be eliminated by design. The problem with three or more periods is that the treatment period will be longer and hence uneconomical and cause a higher drop out rate.

## 6.5. Conclusion

The results of the comprehensive review on general medical statistics showed that in all general medical statistics books which included analyses of cross-over trials, the two-stage procedure was recommended by more than a third of the books. Information given on its deficiency was, however, very limited. None of the 9 books that recommended the two-stage procedure mentioned its high type I error rate. This problem was only explained in three of the books which did not recommend the two-stage procedure.

In addition to Grizzle's (1965) original paper on the two-stage procedure and Hills and Armitage's (1979), medical statistics books are most likely to be the sources of references for authors of cross-over trial reports published in medical journals. With more than a third of general medical statistics recommending the two-stage procedure, it explained the use of the two-stage procedure in medical journals while Grizzle (1965) or Hills and Armitage (1979) was not cited.

However, very little consensus existed among authors of general medical statistics books even when core approaches were similar. For example, details (e.g. the nominal significance level of the preliminary test) and the 2nd stage of the two-stage procedure differed from book to book. Applying the methods to the cross-over trials may result in different methods of analysis when different books are referenced. The diversity in methods of analyses of cross-over trials that are published in medical journals is therefore not surprising.

## 6.6. Books reviewed

**Table 6.6-1: General medical statistics books which give recommendataions to analyses of cross-over trials**

| | | |
|---|---|---|
| Altman, DG. | Practical statistics for medical research | 1994 |
| Armitage, P, Berry, G | Statistical methods in medical research | 1994 |
| Armitage, P. Colton, T. | Encyclopedia of biostatistics | 1998 |
| Bailar III, JC. Mosteller, F. | Medical uses of statistics | 1992 |
| Bolton, S. | Pharmaceutical statistics - practical and clinical applications | 1997 |
| Buncher, JC, Tsay, JY. | Statistics in the pharmaceutical industry | 1994 |
| Chow, SC. Liu, JP. | Design and analysis of bioavailability and bioequivalence studies | 1998 |
| Colquhoun, D. | Lectures on Biostatistics: An introduction to statistics with applications in biology and medicine | 1971 |
| Coppi, B | Perspectives in medical statistics proceedings of the European Symposium of Medical Statistics, Rome, 1980 | 1981 |
| Everitt, BS. | Statistical methods for medical investigations | 1994 |
| Goldstone, LA. | Understanding medical statistics | 1983 |
| Ingelfinger, Mosteller, Thibodeau, Ware | Biostatistics in clinical medicine | 1983 |
| Ingram, D, Bloch, RF | Mathematical methods in medicine | 1984 |
| Jennison, C. Turbull, BW | Group sequential methods with applications to clinical trials | 2000 |
| Mainland, D. Sanders, WB. | Elementary medical statistics | 1963 |
| Matthews-JNS | An introduction to randomized controlled clinical trials | 2000 |
| Petrie, A | Lecture notes on medical statistics | 1987 |
| Piantadosi | Clinical Trials - A methodologic perspective | 1997 |
| Pocock, S | Clinical trials - a practical approach | 1983 |
| Rosner, B. | Fundamentals of biostatistics | 1995 |
| Smart, JV | Elements of medical statistics | 1963 |
| Spriet, Dupin-Spriet, Simon | Methodology of clinical drug trials | 1992 |
| Zolman, JF. | Biostatistics - experimental design and statistical inference | 1993 |

# Chapter 7 - Analysis of cross-over trials in the pharmaceutical industry - A questionnaire survey

## 7.1. Introduction

Clinical trials carried out by the pharmaceutical industry play an important role in determining whether a new treatment will be brought to the general public. Despite their impact, the results of cross-over trials carried out in the pharmaceutical industry are rarely published and the data is only available internally. In order to fill this gap and investigate methods of analyses of cross-over trials in the pharmaceutical industry, a survey with very specific questions concerning the conduct and analysis of cross-over trials was conducted amongst statisticians working in the pharmaceutical industry.

## 7.2. Objective

The objective of the survey is to study current methods used to analyse cross-over trials and general view of such designs amongst statisticians working in the pharmaceutical industry.

## 7.3. Sample population

The sampling frame of the survey consists of every pharmaceutical company in the UK with a listing amongst PSI (Statisticians in the Pharmaceutical Industry) members. The survey was assisted by the PSI, which is a non-profit making, UK-based organisation and aims to promote professional standards of statisticians in matters

pertinent to the pharmaceutical industry. There were 89 statistical units listed on PSI. The number of statisticians working in these unit varies from 1 to 58. Each unit received one questionnaire which was sent to the person nominated as the contact for that unit.

The PSI members' list is published every year in alphabetical order of the companies/institutes and the memberships are categorised into 7 types: ordinary, associate, academic, retired, teacher, affiliate and student. The classification of the membership is by their affiliation with the pharmaceutical industry and the type of organisation they work for. For example, members who work in academic institutes would be qualified as academic members while ordinary members are those who work in the pharmaceutical industry. As the objective of the survey was to investigate how cross-over trials were analysed in the pharmaceutical industry, the target respondents were therefore ordinary members.

Since one of the objectives of the questionnaire was to find out how many trials have been analysed with the two-stage procedure and the total number of clinical trials and cross-over trials analysed by each unit every year, the questionnaire was only sent to the contact person of the pharmaceutical unit. The questionnaire was also designed to be anonymous, it would not have been possible to estimate the percentage of units using the two-stage procedure if each unit was sent more than one questionnaire.

## 7.4. Method

The method of the survey was a questionnaire designed in three formats: an electronic format on the web-site, a Microsoft Word® document sent as an email attachment, and a printed copy. The content of all three formats was identical and each unit listed on the PSI contact list was sent a questionnaire. The web address of the questionnaire was sent to target statisticians by email, in addition to the word document as an attachment for respondents who were required to answer on either the web-site or word document. Statisticians who chose to answer by word document were invited to return the questionnaire by e-mail or by post. The questionnaire was sent by post to statisticians who did not provide their email addresses for PSI contact.

### 7.4.1. The permission of using the PSI member list

The first approach to the PSI was through the PhD supervisor who was an associate member, with the chairman, John Chapman in January 2001, to request the permission of using the PSI members list. The PSI chairman suggested that it would be most appropriate and efficient to send the questionnaire by email from the PSI executive office to the contact person at each company or organisation.

Further contact with the PSI regarding the details of the survey was through the executive secretary, Joanie Lee-Irving. Details discussed before the questionnaire was distributed included

whether it would be possible to send the questionnaire to the ordinary members only, methods of distribution and the appropriate formats of the questionnaire. Other members of the PSI who have assisted with the questionnaire distribution were Lesley Smith and Robin Bickersteth.

### 7.4.2. The questionnaire distribution and the reminder

The final decision on how questionnaires should be distributed was emailed to Lesley Smith of the PSI in May 2001 followed by the questionnaire distribution on the 31st of July 2001. In order to increase the response rate, two months after the questionnaire was first distributed (October 2001), an advertisement of reminder was considered. To explore the possibility and feasibility of such a reminder, an email was sent out to the PSI committee. The permission was again requested from the chairwoman of the PSI at that time, Sue McKeown.

It was also discovered that some units/companies were not sent any questionnaire as they had not nominated a contact person for the PSI. Therefore one statistician was selected from each of the 9 units where there was no PSI contact person and sent the questionnaire. The reminder and the 9 sets of questionnaires were sent out in January 2002.

## 7.5. Objective of the questionnaire

The questionnaire was designed to investigate pharmaceutical statisticians' methods of analysing cross-over trials and their views on such a design and its problems. In particular, the questionnaire asked about how statistical models were fitted, the use of the two-stage procedure, and respondents' awareness of its deficiencies.

## 7.6. Contents of the questionnaire

Copy of the questionnaire can be found in Appendix E

The questionnaire was divided into three sections: the quantity of trials analysed by the unit, various aspects of analyses of cross-over trials, and open questions for personal opinions.

Questions asked regarding the quantity of cross-over trials included the number of clinical trials, cross-over trials and AB/BA designs analysed in the year 2000. This provided the information required to estimate the proportion of cross-over trials among all clinical trials carried out in the pharmaceutical industry, as well as the proportion of AB/BA design used among cross-over trials.

The second section was composed of questions on analyses and the designs of cross-over trials. The questions included respondents' approaches to the carry-over effect and the difficulty posed by various effects in analysing cross-over trials. For statistical models, respondents were asked about the effects they included to fit the statistical model and their views on the two deficiencies of the two-stage procedure.

The last section of the questionnaire contained two open questions where statisticians were asked to give their opinions and references in relation to their analyses of cross-over trials.

The questionnaire was designed in a way that answers to different questions could be cross-checked to confirm the credibility of the results. For example, statisticians who chose to approach the possibility of the carry-over effect by design only (in Q.4) would not be expected to incorporate the carry-over effect in the model or to use the Hills and Armitage method to analyse cross-over trials (Q.7 and Q.9). On the other hand, statisticians who chose to use the statistics to eliminate the possibility of the carry-over effect (in Q.4) would be expected to include the carry-over effect in the model (Q.7) and to answer yes to the use of the two-stage procedure (in Q.9).

In addition, using the two-stage procedure is equivalent to fitting the carry-over effect in the statistical model. Therefore those who answered yes to the use the two-stage procedure (in Q.9) would be expected to incorporate the carry-over effect in the statistical model (in Q.7).

## 7.7. The pilot survey

A pilot survey of 6 statisticians was conducted in order to ensure that there were no snags in the questionnaire and the questions served the purposes as they were intended to. The 6 statisticians (Fiona Campbell, Eddie Channon, Steve Julious, Stephen Jones, Susan Talbot and Angela Macleod) were asked to answer the

questions and to give any suggestions they had regarding the design or wording of the questionnaire. Their answers to the questions were not used to modify the context of the questions but to check whether the objective of the survey was met. As a result of the pilot survey, a few minor changes, mainly to the format of the questionnaire were made while the main questions remained the same.

## 7.8. Results

### 7.8.1. Response from the original questionnaire distribution

The first reply from the statisticians was received on the 31st of July. Forty-two sets of questionnaires were returned between 31st of July and 24th of September. Fifteen of the 42 responses replied to explain why they did not respond to the questionnaire. The most common reasons were that the cross-over design was not or rarely used in their unit or/and they were either working on pre-clinical phases or therapeutic areas where cross-over design was not applicable (e.g. oncology, fertility, stroke and long-term prevention trials). The returned questionnaires were then entered into a Microsoft® Excel file for the analysis.

### 7.8.2. Response from the reminder

Ten further questionnaires were returned after the reminder was sent out. These included 3 statisticians who replied that the reasons they did not respond to the earlier questionnaire was that

they were not working in this area or had not been working on cross-over trials for a long time. Answers from the other 7 statisticians were entered to the Excel® file and analysed together with the questionnaires received during the first round.

### 7.8.3. Response rate

The response rate was around 50%. This included statisticians who sent back the questionnaire and replied that cross-over design was not used in their unit. A poster of the preliminary results was present in the PSI 25th Annual Conference and the PSI chairwomen considered the response rate to be 'huge'.

There were a few reasons for the non-response. First, not all listed units were working on the design and analysis of clinical trials. Twenty units replied that they were working at pre-clinical research discovery or non-clinical areas. For those who were working in clinical trials, the cross-over design may not be appropriate for the therapeutic areas they worked in. One statistician replied that he was working in fertility and cross-over trial was not applicable in his units. Other areas that cross-over trial are rarely used include vaccine, cancer and AIDS trials.

Other reasons that could have caused the non-response was that the questionnaire was sent out in summer when statisticians might be on holiday and might have too much (e)mail to deal with when they came back, therefore the questionnaire might have had a low priority.

There were 52 returned questionnaires with 34 effective samples. Fourteen replied to notify that they did not analyse any cross-over trials in the year 2000, or that they did not work in the field of cross-over trials. The analysis of results was based on the 34 effective questionnaires returned.

### 7.8.4. Quantity and designs of cross-over trials

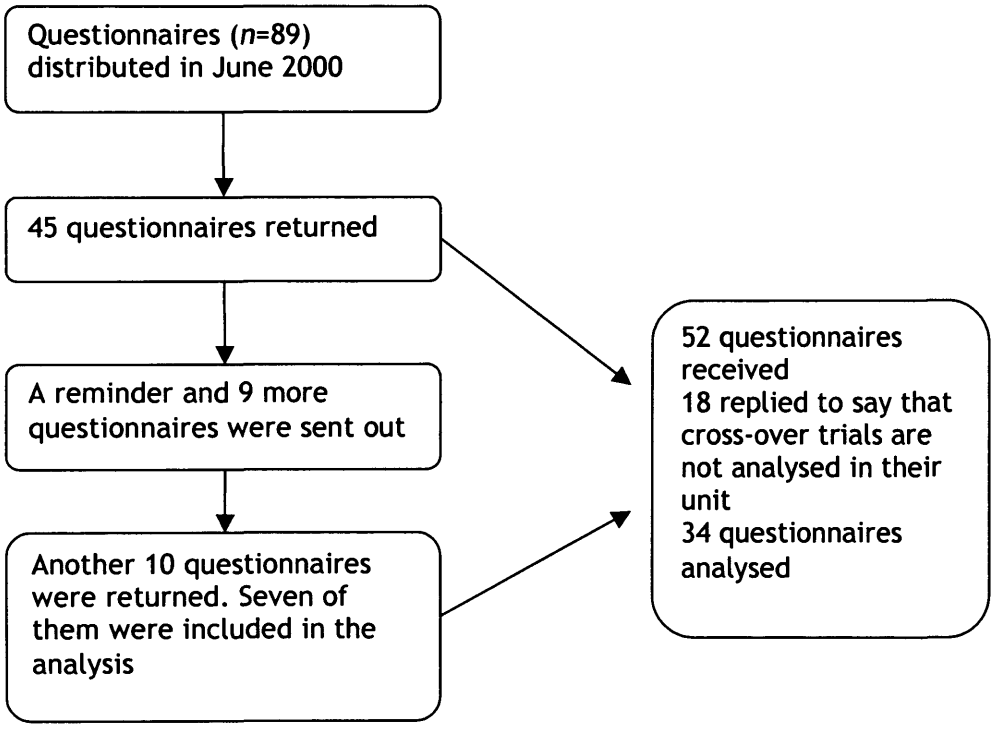
Amongst the 34 effective questionnaires returned, thirty-one units provided information regarding the number of clinical trials or percentage of cross-over or AB/BA trials amongst all clinical trials they analysed in the year 2000. The 31 responses included 4 units that did not analyse any trials. Numbers of clinical trials analysed by the remaining 27 units varied from 2 to 50. They analysed 436 clinical trials in total with a mean of 14.5 trials per unit.

Fourteen of the 34 statistical units analysed cross-over trials in year 2000. These included twelve units which specified the number of clinical trials analysed and two units which only indicated the percentage of cross-over trials amongst all trials they analysed.

Among the 436 clinical trials analysed by all the 27 statistical units which gave the number of clinical trials analysed in their units, 166 were by cross-over trial design. The average percentage of cross-over trials used among all clinical trials was 38%. Since the figure was derived from the units where the cross-over design was used (a few statisticians replied that the cross-over design was not used in their units and therefore did not answer the questions), it cannot be

used to approximate the percentage of cross-over trials of all clinical trials.

The most commonly used design of cross-over trials was the two-period two-treatment design (which was also the targeted design of the two-stage procedure). This design accounted for 121 of 166 cross-over trials analysed in the year 2000. The more complicated designs of cross-over trials used less frequently in the pharmaceutical industry: less than a third (45/166) of cross-over trials analysed by the pharmaceutical industry involved more than two periods or two treatments.



### 7.8.5. Approaches to the carry-over effect

The objective of the two-stage procedure was to test the carry-over effect using statistics. As a consequence, the fact that the carry-over effect is fundamentally a design problem rather than a statistical problem is often overlooked. There is very little that can be done about the carry-over effect once the trial has been completed. The only solution for its elimination is by design, to allow an adequate wash-out period between any two treatment periods.

In order to understand the rationale of their choice of method for analysing cross-over trials, statisticians were first asked about their approaches to the carry-over effect in cross-over trials (Table 7.8-1). The table shows that the most popular approach to the carry-over effect was by design alone. Forty-four per cent of statisticians considered that the possibility of carry-over effect should have been eliminated by design alone. There were nearly as many statisticians who would use both statistics and design to eliminate the carry-over effect as statisticians who considered it to be solely a design problem. Combining these two categories, approximately 85% of the statisticians recognised that design was a solution for the carry-over effect if not the only one. On the other hand, there were three statisticians who considered that the carry-over effect should be dealt with by statistics. The problem of the carry-over effect was considered to be a very serious problem by one of the statisticians, and he was therefore reluctant to use cross-over designs. An

alternative method suggested by a statistician was that the carry-over effect should be eliminated by design and tested in secondary analysis.

The majority (85%) of the statisticians recognised that design is a solution for the carry-over effect, while 50% of them also considered statistics to be a solution. With more than half of the respondents considering statistics to be one of the solutions for the carry-over effect, it is conceivable that the carry-over effect was included in statistical analyses of cross-over trials.

**Table 7.8-1: Approaches to the carry-over effect**

| Approaches to Carry—over Effect | Frequency | Percentage |
|---|---|---|
| Should have been eliminated by design | 15 | 44.1% |
| Is dealt with by statistical tests | 3 | 8.8% |
| Should be eliminated by trial design and tested by statistics | 14 | 41.2% |
| Carry-over Effect cannot be dealt with | 1 | 2.9% |
| Other | 1 | 2.9% |
| **TOTAL** | **34** | **100%** |

### 7.8.6. Difficulty of each effect when analysing cross-over data

In addition to the carry-over effect, there were other factors which were prone to occur in cross-over trials and which therefore made evaluating treatment effects difficult. Such factors common to cross-over designs were the period effect and the patient effect.

When statisticians were asked to rate the difficulty caused by each effect in analysing cross-over trials, the carry-over effect was found to be more problematic than the period or patient effect (Table 7.8-2). More than two-thirds of the statisticians found that the carry-over effect presented some difficulty or a serious problem.

There were more statisticians who found the carry-over effect posed some difficulty or was a serious problem than the other two effects.

In comparison, period and patient effects were of much less concern. Both effects were only considered to be a serious problem by one statistician. The patient effect was the least problematic amongst the three. Nearly a third of statisticians (29%) found the period effect provided some difficulties while only 4 (12%) found the patient effect in this category.

**Table 7.8-2: Rating of difficulty for carry-over, period and patient effects**

| | Carry-over Effect | | Period Effect | | Patient Effect | |
|---|---|---|---|---|---|---|
| **Is not a problem** | 10 | 29% | 23 | 68% | 29 | 85% |
| **Provides some difficulty** | 15 | 44% | 10 | 29% | 4 | 12% |
| **Is a serious problem** | 9 | 26% | 1 | 3% | 1 | 3% |
| **TOTAL** | **34** | **100%** | **34** | **100%** | **34** | **100%** |

### 7.8.7. Approaches to the carry-over effect versus difficulty of the carry-over effect

It was shown in the previous section that the carry-over effect had posed some degree of difficulty for 70 per cent of the statisticians. While the main objective of the two-stage procedure was to test for the significance level of the carry-over effect, this section uses the cross-tabulation to examine whether the approaches towards the possibility of the carry-over effects were related to the perceived difficulty that this effect might cause (Table 7.8-3). Although results of the survey cannot establish the cause and effect relationship between the approach to and the perceived difficulty of the carry-over effect, statisticians' rating of the difficulty of the carry-over effect gave an indication as to whether they were confident with their approaches.

It was noted that 7 of the 10 statisticians who did not consider 'carry-over' to be a problem eliminated it by design alone. Of the twenty-four statisticians who found the carry-over effect to pose some difficulty or to be a serious problem, only 8 chose 'design' as the only solution for the carry-over effect.

The carry-over effect was least likely to be a problem when the statisticians eliminated it by design alone. It was most likely to be a serious problem when statisticians used both design and statistics to eliminate the carry-over effect. It can also be argued that considering the carry-over effect a serious problem was the cause of the statisticians using both design and statistical methods to eliminate the problem, and that eliminating the carry-over effect by design only was the result of carry-over effect not being a problem.

**Table 7.8-3: Approaches vs. perceived difficulty of the carry-over effect**

| | | Perceived difficulty of the carry-over effect in the analysis of cross-over trials | | | |
|---|---|---|---|---|---|
| | | Is not a problem | Provides some difficulty | Is a serious problem | TOTAL |
| **Approaches to Carry-over Effect** | Should have been eliminated by design | 7 | 5 | 3 | **15** |
| | Is dealt with by statistical tests | 1 | 2 | 0 | **3** |
| | Should be eliminated by trial design and tested by statistics | 1 | 7 | 6 | **14** |
| | Cannot be dealt with, therefore I am reluctant to use cross-over design | 1 | 0 | 0 | **1** |
| | Other | 0 | 1 | 0 | **1** |
| | **TOTAL** | **10** | **15** | **9** | **34** |

## 7.8.8. Guidelines and standard operations

Guidelines for analysing cross-over trials were used in nearly one fifth of the pharmaceutical companies. The two-stage procedure

was once treated as a standard operation of cross-over trials in the pharmaceutical industry. If a guideline is applied in that unit, answers given by the statistician who responded to the questionnaire represented the methods used to analyse cross-over trials in the entire unit. Otherwise, methods used by other statisticians in the unit were likely to be different.

**Table 7.8-4: Number of trials analysed VS. use of guideline**

| | With Guidelines (N=6) | Without Guidelines (N=28) |
|---|---|---|
| Total Number of Clinical Trials | 93 | 343 |
| Mean | 18.6 | 15.6 |
| Total Number of Cross-over Trials | 42 | 124 |
| Mean | 8.4 | 5.6 |

Total number of clinical trials and the mean were calculated based on the units that specified the number trials analysed. Five of the 6 units that had guidelines and 22 of the 28 units that did not have guidelines were included in this calculation.

### 7.8.9. Model fitting

The statistical model suggested by Grizzle [4] included treatment effect, carry-over effect, period effect and patient effect. Statisticians were asked which effects they had included in model fitting (Table 7.8-5). Although it was shown in 6.7.2 that 17 statisticians would use the statistics to test for the carry-over effect, only 9 statisticians would always fit the carry-over effect in the model.

**Table 7.8-5: Statistical model fitting of carry-over, patient and period effects for AB/BA design**

| | Model Fitting for AB/BA Design | | |
|---|---|---|---|
| | Carry-over Effect | Patient Effect | Period Effect |
| Always in the model | 9 | 29 | 26 |
| Never in the model | 12 | 2 | 2 |
| Included if significant | 4 | 3 | 3 |
| Included if it makes the treatment effect significant | 0 | 0 | 1 |
| Other | 9 | 0 | 2 |
| **TOTAL** | **34** | **34** | **34** |

**Table 7.8-6: Statistical model fitting of carry-over, patient and period effect for cross-over trial designs with more than 2 periods**
Figures of this table are based on the results

| | Model Fitting | | | |
|---|---|---|---|---|
| | 1st Order Carry-over Effect | 2nd Order Carry-over Effect | Patient Effect | Period Effect |
| Always in the model | 9 | 2 | 28 | 25 |
| Never in the model | 9 | 20 | 1 | 2 |
| Included if significant | 7 | 4 | 3 | 4 |
| Included if makes the treatment effect significant | 0 | 0 | 0 | 0 |
| Other | 6 | 2 | 0 | 1 |
| Missing | 3 | 6 | 2 | 2 |
| **TOTAL** | 34 | 34 | 34 | 34 |

## 7.8.10. Use of the two-stage procedure

It was shown that the two-stage procedure has been the most cited method of cross-over trials (Chapter 4 - ) and has been recommended by more than a third of the general medical statistics books (Chapter 6 - ). When asked whether the two-stage procedure (Table 7.8-7) was used in their units, 7 out of the 34 statisticians considered the method as being used in their units. This figure was not consistent with results of statistical model fitting in Table 7.8-5.

The two-stage procedure tended to be used by bigger units (more clinical trials analysed in year 2000) with a smaller number of cross-over trials. Around 17% of all cross-over trials were analysed by the two-stage procedure in the pharmaceutical industry in the year 2000.

**Table 7.8-7: Number of trials analysed VS. use of the two-stage procedure**

| | The two-stage procedure **USED** (N=7) | The two-stage procedure **NOT USED** (N=27) |
|---|---|---|
| **Clinical trials** | | |
| Total number of clinical trials analysed in year 2000 among all companies | 107 | 329 |
| Mean number of clinical trials analysed | 17.8 | 15.7 |
| **Cross-over trials** | | |
| Total number of cross-over trials analysed by all companies | 24 | 142 |
| Mean number of cross-over trials | 4 | 6.8 |

Total number of clinical trials and the mean were calculated based on the units that specified the number trials analysed. Six of the 7 units that used the two-stage procedure and 21 of the 27 units that did not have guidelines were included in this calculation

## 7.8.11. The two-stage procedure used as a standard operation

The use of the two-stage procedure was compared with the use of guidelines in pharmaceutical companies in order to study whether the two-stage procedure was adopted as a standard operation. In pharmaceutical companies where there were guidelines for analysing cross-over trials, only one of them used the two-stage procedure.

For the 27 units not using the two-stage procedure, 22 of them did not have guidelines for analysing cross-over trials. This indicated that the two-stage procedure might still be used by different statisticians in their units.

**Table 7.8-8: Use of guidelines VS. the two-stage procedure**

| | **The two-stage procedure USED** | **The two-stage procedure NOT USED** | **TOTAL** |
|---|---|---|---|
| Guidelines | 1 | 5 | **6** |
| No Guidelines | 6 | 22 | **28** |
| **TOTAL** | **7** | **27** | **34** |

### 7.8.12. Approaches to the carry-over effect versus use of the two-stage procedure

Theoretically speaking, decisions to use the two-stage procedure can be predicted from answers which described respondents' approaches to the carry-over effect, Table 7.8-9 examines whether that is the case and cross the approaches by the use of the two-stage procedure. There were some inconsistencies in the answers to the two questions. Two statisticians who chose to approach the carry-over effect by design alone worked in pharmaceutical companies where the two-stage procedure was used. There were also two units where statisticians would eliminate the carry-over effect by statistics but answered that the two-stage procedure was not used in their unit.

Therefore, discussion involving the use of the two-stage procedure refers to whether the two-stage procedure was 'considered' for use rather than whether it was actually applied.

**Table 7.8-9: Approaches to the carry-over effect VS. use of the two-stage procedure – survey results**

| | Use of the Two-stage Procedure | | |
|---|---|---|---|
| | Yes | NO | TOTAL |
| Eliminated by design | 2 | 13 | 15 |
| By statistics | 1 | 2 | 3 |
| By both design and statistics | 4 | 10 | 14 |
| Cannot be dealt with | 0 | 1 | 1 |
| Other | 0 | 1 | 1 |
| **TOTAL** | **7** | **27** | **34** |
| | | | |

### 7.8.13. Use of the two-stage procedure versus difficulty of the carry-over effect

The cross-tabulation of the two-stage procedure and the difficulty of the carry-over effect (Table 7.8-10) examined whether testing for the significance of the carry-over effect using the two-stage procedure eased the problem. It does not give information to establish whether the conceived difficulty of the carry-over effect explained why the two-stage procedure was used or, whether knowing the problems of the two-stage procedure makes the carry-over effect difficult to deal with. The table shows however, that analysing cross-over trials with the two-stage procedure does not make the carry-over effect less problematic for the statisticians. Five of the 7 statisticians from the units where the two-stage procedure was used considered the carry-over effect a very serious problem. In contrast, only 4 of the 27 statisticians whose units did not use the two-stage procedure found the carry-over effect a serious problem.

Not using the two-stage procedure also meant that the statisticians were more likely to find the carry-over effect did not to pose a problem. Among the 10 statisticians who considered the carry-over effect not a problem, 8 worked in units where the two-stage procedure was not considered for use. In comparison, 9 of the statisticians who considered the carry-over effect a very serious problem, 5 worked in statistical units where the two-stage procedure was considered for use. Therefore, it can be concluded

that the two-stage procedure did not help to ease the difficulty of the carry-over effect.

**Table 7.8-10: Use of the two-stage procedure VS. difficulty of the carry-over effect**

| | | Perceived difficulty of the carry-over effect | | | |
|---|---|---|---|---|---|
| | | Is not a problem | Provides some difficulties | Is a serious problem | TOTAL |
| **Use of the two-stage procedure** | YES | 2 | 0 | 5 | 7 |
| | NO | 8 | 15 | 4 | 27 |
| | **TOTAL** | **10** | **15** | **9** | **34** |

Results of the perceived difficulty of the carry-over effect and the method of statistical analyses were somewhat different when different variables were compared. Due to the inconsistency between use of the two-stage procedure and statistical model fitting, a cross-tabulation between the difficulty of the carry-over effect and statistical models for the carry-over effect is shown below.

In the category in which that the carry-over effect was always in the model, there was only one statistician who found such an effect a serious problem. This contradicts the previous result that 5 of the 7 units which considered the two-stage procedure for use thought the carry-over effect to be a serious problem. In the category of never fitting the carry-over effect in the model, half of them considered the carry-over effect not to be a problem.

These results implied that never fitting the carry-over effect in the model contributed the highest to the category of the carry-over effect not being a problem, whereas in the previous table not using the two-stage procedure was the main contributor to the category that the carry-over effect provided some difficulty.

**Table 7.8-11: Fitting of the carry-over effect in the model VS. difficulty of the carry-over effect**

| | | Perceived difficulty of the carry-over effect | | | TOTAL |
|---|---|---|---|---|---|
| | | Is not a problem | Provides some difficulty | Is a serious problem | |
| Fitting the model | Always in the model | 4 | 4 | 1 | 9 |
| | Never in the model | 6 | 3 | 3 | 12 |
| | Included if significant | 0 | 2 | 2 | 4 |
| | Other | 0 | 6 | 3 | 9 |
| | TOTAL | 10 | 15 | 9 | 34 |

## 7.8.14. Use of the two-stage procedure versus approaches to the carry-over effect

The inconsistency of the above results can be explained by the difference in the perception of model fitting (Table 7.8-12). Seven of the 27 statisticians whose units did not use the two-stage procedure stated that the carry-over effect was always fitted in the model. Another inconsistency is that 2 of the 7 statisticians from the units where the two-stage procedure was used considered the carry-over effect never fitted in the model.

It was shown that whether the carry-over effect had been fitted in the statistical model was inconsistent with the statisticians' approach to the carry-over effect. Due to the inconsistency between use of the two-stage procedure and model fitting of the carry-over effect, statisticians' approaches to the carry-over effect were checked against the use of the two-stage procedure and further inconsistencies were found. The two-stage procedure was used in two units where the statisticians' approach to the possibility of a carry-over effect was that it should have been eliminated by design.

Such inconsistency may indicate that there were other statistical methods being used, and that they simply did not consider fitting

the carry-over effect in the model to be part of the two-stage procedure.

**Table 7.8-12: Fitting of the carry-over effect VS. use of the two-stage procedure**

| | | The two-stage procedure | | |
|---|---|---|---|---|
| | | Used | Not Used | TOTAL |
| Fitting the model | Always in the model | 2 | 7 | 9 |
| | Never in the model | 2 | 10 | 12 |
| | Included if significant | 1 | 3 | 4 |
| | Other | 2 | 7 | 9 |
| | TOTAL | 7 | 27 | 34 |

## 7.8.15. Lack of power

The low power of the two-stage procedure was first mentioned in Hills and Armitage [7]. Its disadvantage has also been flagged in several general medical statistics books where the two-stage procedure is recommended (Chapter 6 - ). However, lack of power has not been used as a reason to avoid the practice of the two-stage procedure.

The statisticians were asked whether they agreed, disagreed or did not know that the two-stage procedure lacks power. The deficiency was acknowledged by 70 per cent (24/34) of the statisticians. Although such a deficiency was mentioned in the most cited paper (Hills and Armitage [7]) for the analysis of cross-over trials (Chapter 4 - ), 2 statisticians disagreed and 8 did not know one way or the other.

**Table 7.8-13: Recognition of low power of the two-stage procedure**

| | Count | Percentage |
|---|---|---|
| Agree | 24 | 70.6% |
| Disagree | 2 | 5.9% |
| Don't know | 8 | 23.5% |
| Total | 34 | 100% |

### 7.8.16. High type I error rate

The high type I error rate of the two-stage procedure was a more recent finding compared to the problem of low power. The original reference for this deficiency is by Freeman in Statistics in Medicine [8]. Further references following this issue include Senn (1993, 1994, 1995).

Similar to the low power of the two-stage procedure, statisticians were asked about their views on the high type I error rate of the two-stage procedure (Table 7.8-14). Among statisticians who completed this survey, more of them were unaware of this problem than those who were. There were also three statisticians who disagreed with the statement.

**Table 7.8-14: Recognition of low type I error rate of the two-stage procedure**

| | Count | Percentage |
|---|---|---|
| **Agree** | 15 | 44.1% |
| **Disagree** | 3 | 8.8% |
| **Don't know** | 16 | 47.1% |
| **Total** | **34** | **100%** |

### 7.8.17. Deficiencies of the two-stage procedure

When recognitions of the two deficiencies were compared, lack of power was the better-known deficiency of the two-stage procedure. The frequency of recognition of both deficiencies is examined in Table 7.8-15. More than a third of statisticians recognised both deficiencies of the two-stage procedure, followed by nine statisticians who agreed with the power problem but were not aware of the high type I error rate. There were also 7 statisticians who were not aware of either of the problems and one who disagreed with both problems.

**Table 7.8-15: Recognition of both deficiencies**

| | | Low Power | | | |
|---|---|---|---|---|---|
| | | Agree | Disagree | Don't know | TOTAL |
| Type I error Rate | Agree | 13 | 1 | 1 | 15 |
| | Disagree | 2 | 1 | 0 | 3 |
| | Don't Know | 9 | 0 | 7 | 16 |
| | TOTAL | 24 | 2 | 8 | 34 |

## 7.8.18. Deficiencies VS. number of cross-over trials analysed

**Table 7.8-16: Number of trials analysed VS. recognition of the deficiencies**

| | Type I Error Rate | | | Power | | |
|---|---|---|---|---|---|---|
| | Agree (n=15) | Disagree (n=3) | Don't Know (n=16) | Agree (n=24) | Disagree (n=2) | Don't Know (n=8) |
| **Clinical Trials** | | | | | | |
| **Total number of clinical trials analysed** | 125 | 41 | 270 | 319 | 33 | 84 |
| **Mean number of clinical trial analysed** | 11.4 | 13.7 | 20.8 | 16 | 16.5 | 16.8 |
| **Cross-over Trials** | | | | | | |
| **Total number of cross-over trials analysed** | 16 | 7 | 81 | 66 | 2 | 33 |
| **Mean number of cross-over trials analysed** | 1.5 | 2.3 | 6.2 | 3.3 | 1.0 | 6.6 |

Total number of clinical trials and the mean were calculated based on the units that specified the number trials analysed.
Numbers of units included in the calculations were 11, 3 and 13 for the categories of agree, disagree and don't know about the type I error rate respectively.
Numbers of units included in the calculations were 20, 2 and 5 for the categories of agree, disagree and don't know about the power of the two-stage procedure.

## 7.8.19. Deficiencies versus use of the two-stage procedure

In order to investigate whether such awareness of deficiencies discouraged the statisticians from applying the method, statisticians' recognitions of the two deficiencies of the two-stage procedure were tabulated against the use of the two-stage procedure (Table 7.8-17). There were four broad categories in terms of recognition of the deficiency and use of the two-stage procedure:

**1. *Acknowledging* the deficiency/deficiencies and *not using* the two-stage procedure**

The majority of the statisticians who recognised both deficiencies of the two-stage procedure did not use the method to analyse cross-over trials. However, the two-stage procedure was used in 3 of the 13 units where the statisticians agreed with both deficiencies. In other words, the percentage acknowledging both deficiencies was higher among statistical units where the two-stage procedure was used than where it was not. Three statisticians of the 7 units which used the two-stage procedure were aware of both deficiencies compared to 10 of the 27 units which did not use the two-stage procedure. Such a result could indicate that not using the two-stage procedure was not necessarily the result of understanding its deficiency, and knowing the deficiencies did not always lead to avoiding such a procedure (unless the two-stage procedure is the standard operation of the unit. There was 1 unit which fell in this category).

**2. *Acknowledging* the deficiency/deficiencies and still *using* the two-stage procedure**

Among the 24 statisticians who agreed that the power of the two-stage procedure was low, 5 of them worked in units where the two-stage procedure was considered for use. A similar result applied to the high type I error rate: despite 15 statisticians agreeing with the high type I error rate, the two-stage procedure remained in use in three of these units.

There was no significant difference between the two deficiencies in influencing the decision to use the two-stage procedure. Agreeing with either of the deficiencies was almost equally likely (low power 12/15, type I error rate19/24) to lead to not using the two-stage procedure.

**3. *Not acknowledging* the deficiencies and *not using* the two-stage procedure**

Six statisticians of the 27 units that did not use the two-stage procedure were not aware of the deficiencies. Thirteen of them did not know about its high type I error problem, 7 did not know about the low power and 6 did not know either.

**4. *Not* acknowledging the deficiencies and *using* the two-stage procedure**

Not being aware of the deficiencies was considered to be the chief reason that the two-stage procedure continued to be used. However, less than half of the units where the two-stage procedure was used fell into this category.

Amongst the 7 statisticians who worked in units where the two-stage procedure was used, 3 of them did not know the problem of high type I error rate and 1 did not know about the low power.

In addition, amongst the three statisticians who disagreed with the high type I error rate of the two-stage procedure, 2 of them worked in the units where the method was not considered for applications.

One of the two statisticians who disagreed that the two-stage lacks power worked in the units where the two-stage procedure was used

while the other worked in the units where the method was not used.

**Rationale:**

The statistician who disagreed with both deficiencies worked in a unit where the two-stage procedure was considered for application.

The implication was that whether the two-stage procedure was used was not directly influenced by the awareness of the deficiencies of the two-stage procedure. Agreeing with both deficiencies was not followed by giving up the method. On the other hand, not knowing any of the deficiencies of the two-stage procedure did not necessarily indicate keeping the method.

**Table 7.8-17: Use of the two-stage procedure VS. recognition of the two deficiencies**

| | | | High Type I Error Rate | | | |
|---|---|---|---|---|---|---|
| | | | Agree | Disagree | Don't Know | TOTAL |
| **The two-stage procedure USED** | Low Power | Agree | 3 | 0 | 2 | **5** |
| | | Disagree | 0 | 1 | 0 | **1** |
| | | Don't know | 0 | 0 | 1 | **1** |
| | | **TOTAL** | **3** | **1** | **3** | **7** |
| **The two-stage procedure NOT used** | Low Power | Agree | 10 | 2 | 7 | **19** |
| | | Disagree | 1 | 0 | 0 | **1** |
| | | Don't Know | 1 | 0 | 6 | **7** |
| | | **TOTAL** | **12** | **2** | **13** | **27** |

## 7.8.20. Recognition of the deficiencies and difficulties of the carry-over effect

The purpose of this section is to investigate whether the awareness of the deficiencies of the two-stage procedure contributed to the perceived difficulty of the carry-over effect. All statisticians who agreed on the high type I error rate of the two-stage procedure, there was an equal number of statisticians who found the carry-

over effect to be a serious problem and those who thought carry-over effect was not a problem at all. The distribution was similar amongst the statisticians who agreed on the low power of the two-stage procedure. Statisticians who 'did not know' about the high type I error rates were also equally likely to find the carry-over effect not a problem or a serious problem.

Statisticians who agreed on either of the deficiencies were more likely to think that the carry-over effect provided some difficulties than other statisticians who held other views on the deficiencies of the two-stage procedure. The carry-over effect providing some difficulties was also the category that was more likely to occur than the other two categories where statisticians agreed on either of the deficiencies.

Half of the 8 statisticians who did know about the low power of the two-stage procedure considered the carry-over effect not a problem. Almost a third of the statisticians who did not know about the type I error rate problem considered the carry-over effect not a problem.

Difficulty of the carry-over effect was evenly distributed among statisticians who 'did not know' about the high type I error rate of the two-stage procedure.

**Table 7.8-18: Recognition of the two deficiencies VS. difficulty of the carry-over effect**

| | | Perceived difficulty of the carry-over effect | | | |
|---|---|---|---|---|---|
| | | Is not a problem | Provides some difficulties | Is a serious problem | TOTAL |
| **Low Power** | Agree | 5 | 13 | 6 | **24** |
| | Disagree | 1 | 0 | 1 | **2** |
| | I don't know | 4 | 2 | 2 | **8** |
| | **TOTAL** | **10** | **15** | **9** | **34** |
| | | | | | |
| **High Type I Error Rate** | Agree | 4 | 7 | 4 | **15** |
| | Disagree | 1 | 2 | | **3** |
| | Don't know | 5 | 6 | 5 | **16** |
| | **TOTAL** | **10** | **15** | **9** | **34** |

## 7.9. Conclusion

The two-stage procedure was considered for use by nearly a fifth of the statistical units and the carry-over effect in the cross-over design remained problematic. For individual questions in the survey, there was no consensus on approaches to the carry-over effect or how it should be analysed among the pharmaceutical statisticians. In addition, there were some inconsistencies between answers given to questions regarding analyses and approaches to the carry-over effect.

The reason why the two-stage procedure was used has not been established. The recommendation by Freeman (1989) who found the major deficiency of the two-stage procedure has not been generally adopted. Agreeing with the deficiency of the two-stage procedure did not always discourage researchers from using the method and not knowing either the lower power or high type I error rate of the two-stage procedure did not indicate that the two-stage procedure would still be used.

The design of the questionnaire was based on the concept that the two-stage procedure was a common problem and its deficiency of high type I error rate, which was only published 25 years after the method was introduced, might have had some impact on its use. However, the results of the survey showed that the choice of analysis of cross-over trials involved more than whether its deficiencies were recognised and there was a question in consensus of fundamental perception of the two-stage procedure and model fitting of the carry-over effect.

# Chapter 8 - Summary of Results

This chapter presents a comprehensive description of the use of the two-stage procedure in different aspects of medical statistics by comparing the results drawn from the four independent studies (Chapter 4 - Chapter 5 - Chapter 6 - and Chapter 7 - ).

The summary of results includes direct comparisons between any two aspects of medical statistics as well as overviews by combining results from two or more aspects of general medical statistics books to give a complete picture of analyses of cross-over trials and to compare the methods used to analyse cross-over trials amongst different professions.

### 8.1.1. Medical journals - citation analysis and systematic review

The citation analysis and the systematic review were the two research methods used to investigate the two-stage procedure in periodicals. The citation analysis included all scientific periodicals while systematic review focused on cross-over trial reports published in medical journals.

The main objective of the citation analysis was to investigate how often the two-stage procedure was used. Citation analysis is based on the assumption that a citation given to a paper describing the two-stage procedure indicated the usage of the method. It has not been established that not citing the method implied that the two-stage procedure was not used. The systematic review of medical

journals was used to give a more detailed examination on analyses of cross-over trials and assisted in verifying the relationship between a citation and the actual application.

The key finding of the citation analysis was that the two papers describing the two-stage procedure have been more cited than the paper about its deficiencies. Such a difference is most substantial in the field of medical research. Results of the systematic review on cross-over trial reports showed that the carry-over effect was often tested without citing the two-stage procedure or specifying the method used. However, as the method of testing was often not specified in the report, it could not be established whether the two-stage procedure was applied. As a result, the relationship between citation/non-citation and application/non-application was not established. Therefore, results drawn from the citation analysis cannot be used as an indicator of the use of the two-stage procedure for analyses of cross-over trials published in medical journals. However, it can be confirmed that incidences of testing for the carry-over effect were more common than citations given to the two-stage procedure.

**The relationship between citation and use of the method**

As a result of the insufficient information given in cross-over trial reports, a citation on the two-stage procedure indicates the application of the method was assumed but was not verified. On the other hand, neither did the reverse apply. The use of the two-stage procedure did not always imply that a citation was given to a

reference of the method. Such examples were found in systematic review in Chapter 5 - . Similarly, not citing the two-stage procedure did not imply that the method was not used. The only association that can be assumed was that not using the method usually indicates that the method was not cited.

### 8.1.2. The pharmaceutical industry and medical journals - a comparison

Both pharmaceutical industry and medical journals involve actual analyses of cross-over trials. Results drawn from the survey and systematic review showed how cross-over trials had been analysed and the difference in methods of analyses by different research populations.

It was shown in Chapter 5 - and Chapter 7 - that the use of the two-stage procedure could not be estimated in either medical journals or the pharmaceutical industry. For cross-over trial reports published in medical journals, information given in cross-over trial reports was not sufficient to determine whether the two-stage procedure was applied. For analyses of cross-over trials in the pharmaceutical industry, the inconsistency in fitting the statistical models and the use of the two-stage procedure was found in the results of the survey for pharmaceutical statisticians. Seven of the 34 pharmaceutical companies who responded to the questionnaire answered yes to whether the two-stage procedure had been used in their units. However, results drawn from questions regarding their approaches to the carry-over effect and the variables they fitted in

the statistical model showed that using statistics to test for the carry-over effect was estimated to be higher than the rate of the use of the two-stage procedure. It highlighted the issue that the two-stage procedure had not been well-recognised and it might have been practised without knowing the method. It therefore makes raising the awareness of the deficiencies even more difficult.

Another similarity in the finding of the two studies was the percentages of AB/BA design used amongst cross-over trials in medical journals and the pharmaceutical industry. The AB/BA design was used by 45 (70%) of the 64 cross-over trial reports reviewed and 121 (73%) of the 166 cross-over trials analysed in the pharmaceutical industry.

### 8.1.3. Medical statistics books and pharmaceutical industry - A comparison

Medical statistics books and the pharmaceutical industry involve different aspects and stages of analyses of cross-over trials. However, research methods for both studies served the purpose of investigating the awareness of the deficiencies of the two-stage procedure, views on the carry-over effect and methods used/recommended for analyses of cross-over trials.

The role of medical statistics books in terms of the deficiencies of the two-stage procedure is to give readers the information regarding such problems while pharmaceutical statisticians were directly asked about their awareness of the two deficiencies of the two-stage procedure.

The results of the comprehensive review and the survey showed that the two-stage procedure was more commonly recommended and the deficiencies were less acknowledged by general medical statistics books than the pharmaceutical industry. However, it cannot be concluded that the pharmaceutical industry was more up to date than general medical statistics books in terms of the analyses of cross-over trials as the comprehensive review of general medical statistics covered publication years from 1963 to 2000 while the survey was conducted in the year 2001. What can be concluded was that the two-stage procedure continued to be recommended and applied by these two research populations while its deficiencies were not fully acknowledged and understood in both realms of medical statistics.

**Table 7.9-1: Comparison of main results from general medical statistics books and the pharmaceutical industry**

| | General Medical Statistics Books | Pharmaceutical Industry |
|---|---|---|
| Use of the two-stage procedure | 9/26 | 7/34 |
| Awareness of low power | 7/26 | 24/34 |
| Awareness of type I error rate | 3/26 | 15/34 |

### 8.1.4. Citation analysis and medical statistics books

The citation analysis on the papers about the two-stage procedure and the comprehensive analysis on medical statistics books were both used to investigate references for analysing cross-over trials. High citation rates on Grizzle [4] and Hills and Armitage [7], together with more than a third of general medical statistics books recommending the two-stage procedure, showed that the method has been around and will remain so for some time in the future.

### 8.1.5. Citation analysis and the pharmaceutical industry

The citation analysis was used as an indicator of the use of the two-stage procedure in published cross-over trial reports. Whether the two-stage procedure was used in their unit was one of the survey questions for the pharmaceutical statisticians. The results compared the use of the two-stage procedure in different realms of medical statistics. As in the results in Section 8.1.4. the two-stage procedure is likely to be applied by some cross-over trial reports published in medical journals and some pharmaceutical statisticians for some time.

### 8.1.6. Medical statistics books and medical journals

The purpose of comparing results drawn from the comprehensive review of medical statistics books and the systematic review on medical journals is to investigate whether medical statistics books had any influence on analyses of cross-over trials in medical journals.

The results in Chapter 4 - showed that most citations given by cross-over trial reports were papers from periodicals rather than medical statistics books. However, there was a considerable percentage of cross-over trial reports that did not give citations to any statistical references. These included reports which mentioned that the carry-over effect had been tested. It indicated that there were likely to be other sources of references for the analysis. Medical statistics books are considered to be the most commonly

used sources that provide methods of analyses. Furthermore, it was found that many books have recommended testing for the carry-over effect without referring it as the two-stage procedure. Therefore, cross-over trial reports that indicated testing for the carry-over effect without giving information on methods used could well indicate the influences of general medical statistics books on analyses of cross-over trials published in medical journals.

## 8.2. Summary of Results

The original objective of the thesis was to investigate the use of the two-stage procedure for analysing cross-over trials in four different aspects of medical statistics. However, the results from the four studies showed that it is difficult to estimate the prevalence of the use of the two-stage procedure in cross-over trials. The overall conclusion of this thesis is that the two-stage procedure is not the only area in the analysis of cross-over trials that has deficiencies. None of the four methods was able to give an estimation of the percentage of cross-over trials being analysed with the two-stage procedure because of the deficiencies in each realm of medical statistics.

The reasons for not meeting the objective of the studies were that those objectives were based on assumptions which were only found to be unreliable when results of the studies were analysed. Violations of the assumptions were the major finding of the thesis. Below is the discussion regarding the assumptions and the results.

1. As an assumption of bibliometric study, it was assumed that citation indicated use. Although the evidence found in the systematic review of cross-over trial reports was not strong enough to reject such an assumption, neither was it supported. It was also assumed that the analysis section of cross-over trial reports would indicate whether the two stage procedure was used or not. However, results of the systematic review on cross-over trials published in medical journals showed that the two-stage procedure was cited while it was not mentioned anywhere in the trial report. On the other hand, the carry-over effect was described as 'non-existent' while the method of testing was not indicated. Therefore citations cannot be treated as a reliable indicator of the use of the two-stage procedure. The major finding of the systematic review was that the design, methods and discussion that had been published was generally not sufficient to judge whether the design and analyses of cross-over trials were appropriate.

2. Information provided by cross-over trial reports published in medical journals was inadequate and did not meet the requirement of the 'CONSORT statement' or journal checklists. As a consequence, the relationship between citations and usages was not established. Insufficient information from cross-over trial reports also applies to the section of statistical analyses.

3. There were unforeseen inconsistencies between answers to the questions of the survey for the pharmaceutical statisticians. Use of the two-stage procedure was not consistent with the

statistical model fitting. In addition, neither did their views on how carry-over effects should be dealt with reflect on the statistical methods they used.

The conclusion that can be drawn from the results of the four studies was that without being recognised by those who analysed trials, the two-stage procedure had been more commonly used than the evidence suggested. It is impossible to ascertain the aspect of medical statistics in which the two-stage procedure was most or least used. It can only be concluded that the two-stage procedure continued to be highly used in all aspects of medical statistics studied. The reason why it was still used was associated with the fact that its deficiencies have not been as well-recognised as the method itself. Furthermore, the fact that testing for the carry-over effect without realising it was the so-called 'two-stage procedure' or 'the Hills-Armitage method' made it more difficult to raise awareness of its deficiencies.

**Table 8.2-1**: Comparison of main results from the four studies

| Aspect of medical statistics | Periodicals | Medical journals | General medical statistics books | Pharmaceutical industry |
|---|---|---|---|---|
| **Research Method** | **Citation analysis** | **Systematic review** | **Comprehensive review** | **Questionnaire survey** |
| **The two-stage procedure** | 1,328 citations on Grizzle (1965) and Hills and Armitage (1979) | Cannot be estimated | 9 of 26 books recommended the two-stage procedure | 7 of 34 pharmaceutical companies used the two-stage procedure |
| **Acknowledging Low power** | 3 trial papers cited Freeman (1989) | 2 | 7/26 | 24/34 |
| **Acknowledging High type I error rate** | 3 trial papers cited Freeman (1989) | 2 | 3/26 | 15/34 |

Regarding the content of the two-stage procedure, it was found that there have been different versions of the two-stage procedure in general medical statistics books and a lack of understanding in the pharmaceutical industry.

As for the major deficiency (high type I error rate) of the two-stage procedure, this has been acknowledged by newly published general medical statistics books. Yet, although infrequent, the awareness could be found in cross-over trial reports published in medical journals. In general, analyses of cross-over trials were often associated with the two-stage procedure. Although it is hoped that there will be a change in analyses of cross-over trials, the two-stage procedure is unlikely to be eliminated from analyses of cross-over trials soon.

## 8.3. Suggestions for improvements

The results drawn from the four studies suggest that improvements in analyses of cross-over trials are needed in each realm of medical statistics. Suggestions are given below.

### 8.3.1. Authors of cross-over trials reports to be published in medical journals

The two fundamental issues of cross-over trial reports published in medical journals are reporting and analyses. Both areas are equally important as it is often impossible to differentiate whether it is a problem of reporting or analyses.

**For reporting of trials:** Details of design should always be included in the report. Designs of cross-over trials should always be specified. The essential information regarding designs of cross-over trials are numbers of treatments, treatment periods and treatment groups. As the carry-over effect is one of the most problematic areas for concern in using cross-over designs, whether a wash-out period has been incorporated and the choice of its length should always be justified.

**For analyses of cross-over trials:** Report of cross-over trials is similar to an ordinary parallel trial except when carry-over effects are tested. There should always be clear indications of which methods have been used and on which variables they have been tested on. The two-stage procedure is not recommended for analyses of cross-over trials due to its high type I error rate and low power. However, if the two-stage procedure has to be used, how carry-over effects have been tested and at what nominal level should be stated clearly. Results of the test for carry-over effects and how treatment effects are evaluated should always be provided.

Statisticians should be consulted whenever possible, and the focus should be on the validity of the method for their data.

In terms of citations of references, references should be given if statistical analyses involve more than testing for simple cross-over differences. References should always be cited if they have been mentioned in the report. Giving correct citations is not only for the author's credibility of their work but will benefit readers if further details/reading are needed.

### 8.3.2. For the editorial board/peer review

Editorial board/peer review plays an important role in the quality of cross-over trials published in medical journals. In order to improve the reporting and analyses of cross-over trials published in medical journals, there should be guidelines on how cross-over trials should be analysed and a checklist of items included in cross-over trial reports. Such a guideline/checklist would not only improve the quality of cross-over trials published in medical journals but also direct authors to better reports. Such guidelines should include all items mentioned in Section 8.3.1.

### 8.3.3. General medical statistics books

Despite the major deficiencies of the two-stage procedure, it is not the purpose of this thesis to condemn all general medical statistics books which recommended the two-stage procedure. A good general medical statistics book as a reference for analyses of cross-over trials would give information on deficiencies of the two-stage

procedure even if the method is recommended. However, it was found that many general medical statistics books did not provide the correct version of the two-stage procedure, such as the recommended nominal level of testing for the carry-over effect and how treatment effects should be analysed when carry-over effects are found to be significant or non-significant. If they know the deficiencies of the two-stage procedure and still consider the two-stage procedure the optimal method for analyses of cross-over, authors should give a correct and complete version of the two-stage procedure.

Suggestions for authors are to emphasise the deficiencies of the two-stage procedure and make it clear that despite its attractive feature of testing for the carry-over effect, it actually has a high price to pay in terms of the false positive (type I error) rate. If the authors were not sure about their views or the latest development of the method, they would skip the topic about analyses of cross-over trials and/or refer to other books which specialise in cross-over trials or experiment designs.

As it was found that there were no two books which gave identical suggestions on analyses of cross-over trials, it must have caused researchers who analyse cross-over trials data trouble choosing which book to follow. The suggestion for them is that newly published medical statistics books are more reliable than the old ones regarding analyses of cross-over trials. If more specific

analysis is needed, readers should consult specialty books such as Senn's.

## 8.4. Pharmaceutical industry

No conclusion can be drawn from the survey amongst the pharmaceutical statisticians in regard to the use of the two-stage procedure as the results showed some inconsistency in methods of the two-stage procedure.

The deficiencies of the two-stage procedure have not always been acknowledged. The high type I error rate is a more serious problem than lower power. Therefore it is important for statisticians to be informed about the problem before they decide to analyse cross-over trials with the two-stage procedure.

## 8.5. Future work

In order to improve the statistics in medical journals, it will be useful to investigate how clinicians acquire their knowledge of statistics. In addition, more effort should be made in demonstrating and recommending the correct method for analyses of cross-over trials. Another area that requires attention is potential researchers/clinicians' understanding of the design of cross-over trials.

# Appendices

## Appendix A - Number of cross-over trials listed on the MEDLINE

| Year | Total Number of Cross-over Trials |
|---|---|
| **1981** | 95 |
| **1982** | 126 |
| **1983** | 121 |
| **1984** | 129 |
| **1985** | 134 |
| **1986** | 168 |
| **1987** | 160 |
| **1988** | 156 |
| **1989** | 165 |
| **1990** | 148 |
| **1991** | 182 |
| **1992** | 181 |
| **1993** | 177 |
| **1994** | 166 |
| **1995** | 166 |
| **1996** | 167 |
| **1997** | 161 |
| **1998** | 126 |
| **1999** | 122 |
| **2000** | 124 |
| **2001** | 188 |
| Total | 3162 |

## Appendix B - 1988 List of cross-over trial reports sampled for review

1988-1 Schaer DH, Buff LA, Katz RJ. Sustained antianginal efficacy of transdermal nitroglycerin patches using an overnight 10-hour nitrate-free interval. Am.J Cardiol. 1988;61(1):46-50.
Ref ID: 3

1988-2 McMenemin IM, Parbrook GD. Comparison of the effects of subanaesthetic concentrations of isoflurane or nitrous oxide in volunteers. Br.J Anaesth. 1988;60(1):56-63.
Ref ID: 5

1988-3 Campbell N, Paddock V, Sundaram R. Alteration of methyldopa absorption, metabolism, and blood pressure control caused by ferrous sulfate and ferrous gluconate. Clin Pharmacol.Ther. 1988;43(4):381-6.
Ref ID: 24

1988-4 Sasso E, Perucca E, Calzetti S. Double-blind comparison of primidone and phenobarbital in essential tremor. Neurology 1988;38(5):808-10.
Ref ID: 30

1988-5 Stromberg C, Seppala T, Mattila MJ. Acute effects of maprotiline, doxepin and zimeldine with alcohol in healthy volunteers. Arch.Int.Pharmacodyn.Ther. 1988;291:217-28.
Ref ID: 32

1988-6 Niemi MK, Keinanen-Kiukaanniemi SM, Salmela PI. Long-term effects of guar gum and microcrystalline cellulose on glycaemic control and serum lipids in type 2 diabetes. Eur.J Clin Pharmacol. 1988;34(4):427-9.
Ref ID: 57

1988-7 Gleeson JG, Price JF. Controlled trial of budesonide given by the nebuhaler in preschool children with asthma. BMJ 1988;297(6642):163-6.
Ref ID: 63

1988-8 Wheatley D. New hypnotic agents: clinical studies in general practice. Pharmacol.Biochem.Behav. 1988;29(4):811-3.
Ref ID: 69

1988 -9 Gsinger C, Jeremy J, Speiser P, Mikhailidis D, Dandona P, Schernthaner G. Effect of vitamin E supplementation on platelet thromboxane A2 production in type I diabetic patients. Double-blind crossover trial. Diabetes 1988;37(9):1260-4.
Ref ID: 75

1988-10 Theron AJ, Anderson R. Investigation of the effects of oral administration of ascorbate on the functional activity of serum alpha-1-protease inhibitor and oxidant release by blood phagocytes from cigarette smokers in a placebo-controlled, doubleblind, crossover trial. Int.J Vitam.Nutr.Res. 1988;58(2):218-24.
Ref ID: 84

1988-11 Catterall JR, Rhind GB, Whyte KF, Shapiro CM, Douglas NJ. Is nocturnal asthma caused by changes in airway cholinergic activity? Thorax 1988;43(9):720-4.
Ref ID: 99

1988-12 Ney P, Neal T, Manku MS. Double blind cross-over trial with fenfluramine. Can.J Psychiatry 1988;33(6):574.
Ref ID: 102

1988-13 Engelhart M. Ketanserin in the treatment of Raynaud's phenomenon associated with generalized scleroderma. Br.J Dermatol. 1988;119(6):751-4.
Ref ID: 108

1988-14 Weisweiler P. Effects of bezafibrate and gemfibrozil on serum lipoproteins in primary hypercholesterolemia. Arzneimittelforschung. 1988;38(7):925-7.
Ref ID: 111

1988-15 Stoller JK, Wiedemann HP, Loke J, Snyder P, Virgulto J, Matthay RA. Terbutaline and diaphragm function in chronic obstructive pulmonary disease: a double-blind randomized clinical trial. Br.J Dis.Chest 1988;82(3):242-50.
Ref ID: 137

1988-16 Armbrecht U, Lundell L, Stockbrugger RW. The benefit of pancreatic enzyme substitution after total gastrectomy. Aliment.Pharmacol.Ther. 1988;2(6):493-500.
Ref ID: 140

## Appendix B - 1992 List of cross-over trial reports sampled for review

1992-1 Atala A, Amin M, Harty JI. Diethylstilbestrol in treatment of postorchiectomy vasomotor symptoms and its relationship with serum follicle-stimulating hormone, luteinizing hormone, and testosterone. Urology 1992;39(2):108-10.
Ref ID: 8

1992-2 Brodows RG. Benefits and risks with glyburide and glipizide in elderly NIDDM patients. Diabetes Care 1992;15(1):75-80.
Ref ID: 11

1992-3 Hood MA, Smith WM. Adenosine versus verapamil in the treatment of supraventricular tachycardia: a randomized double-crossover trial. Am.Heart J 1992;123(6):1543-9.
Ref ID: 48

1992-4 Clifton PM, Wight MB, Nestel PJ. Is fat restriction needed with HMGCoA reductase inhibitor treatment? Atherosclerosis 1992;93(1-2):59-70.
Ref ID: 49

1992-5 Uden S, Schofield D, Miller PF, Day JP, Bottiglier T, Braganza JM. Antioxidant therapy for recurrent pancreatitis: biochemical profiles in a placebo-controlled trial. Aliment.Pharmacol.Ther. 1992;6(2):229-40.
Ref ID: 53

1992-6 Lieberman D, Heimer D. Effect of dietary sodium on the severity of bronchial asthma. Thorax 1992;47(5):360-2.
Ref ID: 62

1992-7 Hay JG, Stone P, Carter J, Church S, Eyre-Brook A, Pearson MG et al. Bronchodilator reversibility, exercise performance and breathlessness in stable chronic obstructive pulmonary disease. Eur.Respir.J 1992;5(6):659-64.
Ref ID: 74

1992-8 Laffi G, Marra F, Carloni V, Azzena G, De Feo ML, Pinzani M et al. Thromboxane-receptor blockade increases water diuresis in cirrhotic patients with ascites. Gastroenterology 1992;103(3):1017-21.
Ref ID: 88

1992-9 Ramaekers JG, Uiterwijk MM, O'Hanlon JF. Effects of loratadine and cetirizine on actual driving and psychometric test

performance, and EEG during driving. Eur.J Clin Pharmacol. 1992;42(4):363-9.
Ref ID: 95

1992-10 Hauser SL, Doolittle TH, Lopez-Bresnahan M, Shahani B, Schoenfeld D, Shih VE et al. An antispasticity effect of threonine in multiple sclerosis. Arch.Neurol. 1992;49(9):923-6.
Ref ID: 100

1992-11 Goel A, Suri JC, Aggarwal K. Role of corticosteroids in the management of chronic obstructive lung disease: factors predicting response. Indian J Chest Dis.Allied Sci. 1992;34(1):11-7.
Ref ID: 101

1992-12 Riddle MA, Scahill L, King RA, Hardin MT, Anderson GM, Ort SI et al. Double-blind, crossover trial of fluoxetine and placebo in children and adolescents with obsessive-compulsive disorder. J Am.Acad.Child Adolesc Psychiatry 1992;31(6):1062-9.
Ref ID: 116

1992-13 Buchsbaum MS, Potkin SG, Siegel B-VJ, Lohr J, Katz M, Gottschalk LA et al. Striatal metabolic rate and clinical response to neuroleptics in schizophrenia. Arch.Gen.Psychiatry 1992;49(12):966-74.
Ref ID: 120

1992-14 Kishida H, Hata N, Kunimi T, Miyagawa H, Nishiyama H, Katoh K. Antianginal effects of amlodipine at a single dose on exertional angina patients using treadmill exercise testing—a randomized crossover study in comparison with placebo. Cardiovasc.Drugs Ther. 1992;6(5):481-7.
Ref ID: 122

1992-15 Palmer AJ, Fletcher AE, Rudge PJ, Andrews CD, Callaghan TS, Bulpitt CJ. Quality of life in hypertensives treated with atenolol or captopril: a double-blind crossover trial. J Hypertens 1992;10(11):1409-16.
Ref ID: 139

1992-16 Sanderson A, Carpenter R. Eye movement desensitization versus image confrontation: a single-session crossover study of 58 phobic subjects. J Behav.Ther.Exp.Psychiatry 1992;23(4):269-75.

## Appendix B - 1996 List of cross-over trial reports sampled for review

1996-1 Hansen O, Pfeiffer P, Madsen B, Andersen I, Hansen B, Mathiesen B. Sustained-release metoclopramide plus methylprednisolone versus placebo plus methylprednisolone as antiemetic prophylaxis during non-cisplatin chemotherapy. A randomized double-blind cross-over trial. Acta Oncol. 1996;35(1):57-61.
Ref ID: 17

1996-2 Hong D, Regehr G, Reznick RK. The efficacy of a computer-assisted preoperative tutorial for clinical clerks. Can.J Surg. 1996;39(3):221-4.
Ref ID: 19

1996-3 Sekine I, Nishiwaki Y, Kakinuma R, Kubota K, Hojo F, Matsumoto T et al. A randomized cross-over trial of granisetron and dexamethasone versus granisetron alone: the role of dexamethasone on day 1 in the control of cisplatin-induced delayed emesis. Jpn.J Clin Oncol. 1996;26(3):164-8.
Ref ID: 28

1996-4 El Chaar GM, Mardy G, Wehlou K, Rubin LG. Randomized, double blind comparison of brand and generic antibiotic suspensions: II. A study of taste and compliance in children. Pediatr Infect.Dis.J 1996;15(1):18-22.
Ref ID: 31

1996-5 Gatto LM, Hallen GK, Brown AJ, Samman S. Ascorbic acid induces a favorable lipoprotein profile in women. J Am.Coll.Nutr. 1996;15(2):154-8.
Ref ID: 35

1996-6 Levin A, Goldstein MB. The benefits and side effects of ramped hypertonic sodium dialysis. J Am.Soc.Nephrol. 1996;7(2):242-6.
Ref ID: 36

1996-7 Sakai T, Antoku Y, Matsuishi T, Iwashita H. Tetrahydrobiopterin double-blind, crossover trial in Machado-Joseph disease. J Neurol.Sci. 1996;136(1-2):71-2.
Ref ID: 38

1996-8 Winer KK, Yanovski JA, Cutler G-BJ. Synthetic human parathyroid hormone 1-34 vs calcitriol and calcium in the

treatment of hypoparathyroidism. JAMA 1996;276(8):631-6.
Ref ID: 53

1996-9 Parry BL, Hauger R, LeVeau B, Mostofi N, Cover H, Clopton P et al. Circadian rhythms of prolactin and thyroid-stimulating hormone during the menstrual cycle and early versus late sleep deprivation in premenstrual dysphoric disorder. Psychiatry Res. 1996;62(2):147-60.
Ref ID: 58

1996-10 Fux M, Levine J, Aviv A, Belmaker RH. Inositol treatment of obsessive-compulsive disorder. Am.J Psychiatry 1996;153(9):1219-21.
Ref ID: 60

1996-11 Striebel HW, Romer M, Kopf A, Schwagmeier R. Patient controlled oral analgesia with morphine. Can.J Anaesth. 1996;43(7):749-53.
Ref ID: 64

1996-12 Dowson A. Can oral 311C90, a novel 5-HT1D agonist, prevent migraine headache when taken during an aura? Eur.Neurol. 1996;36 Suppl 2:28-31.
Ref ID: 69

1996-13 MacLeod KM, Gold AE, Frier BM. A comparative study of responses to acute hypoglycaemia induced by human and porcine insulins in patients with Type 1 diabetes. Diabet.Med 1996;13(4):346-57.
Ref ID: 90

1996-14 Clark HE, Matthews DR. The effect of glimepiride on pancreatic beta-cell function under hyperglycaemic clamp and hyperinsulinaemic, euglycaemic clamp conditions in non-insulin-dependent diabetes mellitus. Horm.Metab Res. 1996;28(9):445-50.
Ref ID: 106

1996-15 Kluft J, Beker L, Castagnino M, Gaiser J, Chaney H, Fink RJ. A comparison of bronchial drainage treatments in cystic fibrosis. Pediatr Pulmonol. 1996;22(4):271-4.
Ref ID: 107

1996-16 Evans DJ, Barnes PJ, Spaethe SM, van Alstyne EL, Mitchell MI, O'Connor BJ. Effect of a leukotriene B4 receptor antagonist, LY293111, on allergen induced responses in asthma. Thorax 1996;51(12):1178-84.
Ref ID: 128

## Appendix B - 2000 List of cross-over trial reports sampled for review

2000-1 Pernerstorfer T, Stohlawetz P, Kapiotis S, Eichler HG, Jilma B. Partial inhibition of nitric oxide synthase primes the stimulated pathway of vWF-secretion in man. Atherosclerosis 2000;148(1):43-7.
Ref ID: 1

2000-2 Balasekaran R, Porter JL, Santa-Ana CA, Fordtran JS. Positive results on tests for steatorrhea in persons consuming olestra potato chips. Ann.Intern.Med 2000;132(4):279-82.
Ref ID: 10

2000-3 Pope H-GJ, Kouri EM, Hudson JI. Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: a randomized controlled trial. Arch.Gen.Psychiatry 2000;57(2):133-40.

2000-4 Benton RE, Sale M, Flockhart DA, Woosley RL. Greater quinidine-induced QTc interval prolongation in women. Clin Pharmacol.Ther. 2000;67(4):413-8.
Ref ID: 36

2000-5 Burstein AH, Fisher KM, McPherson ML, Roby CA. Absorption of phenytoin from rectal suppositories formulated with a polyethylene glycol base. Pharmacotherapy 2000;20(5):562-7.
Ref ID: 47

2000-6 Bermingham EC, Papich MG, Vivrette SL. Pharmacokinetics of enrofloxacin administered intravenously and orally to foals. Am.J Vet.Res. 2000;61(6):706-9.
Ref ID: 58

2000-7 Marzo A, Monti NC, Tettamanti RA, Crivelli F, Dal Bo L, Mazzucchelli P et al. Bioequivalence of inhaled formoterol fumarate assessed from pharmacodynamic, safety and urinary pharmacokinetic data. Arzneimittelforschung. 2000;50(6):559-63.
Ref ID: 70

2000-8 Frier BM, Ewing FM, Lindholm A, Hylleberg B, Kanc K. Symptomatic and counterregulatory hormonal responses to acute hypoglycaemia induced by insulin aspart and soluble human insulin in Type 1 diabetes. Diabetes Metab Res.Rev. 2000;16(4):262-8.

Ref ID: 81

2000-9 Shamir E, Barak Y, Plopsky I, Zisapel N, Elizur A, Weizman A. Is melatonin treatment effective for tardive dyskinesia? J Clin Psychiatry 2000;61(8):556-8.
Ref ID: 89

2000-10 Chao CK, Yu LL, Su LL, Liu CM, Yang TH, Chen CM. Bioequivalence study of tramadol by intramuscular administration in healthy volunteers. Arzneimittelforschung. 2000;50(7):636-40.
Ref ID: 92

2000-11 Williamson L, Illingworth H, Smith D, Mowat A. Oral quinine in ankylosing spondylitis: a randomized placebo controlled double blind crossover trial. J Rheumatol. 2000;27(8):2054-5.
Ref ID: 96

2000-12 Larson VD, Williams DW, Henderson WG, Luethke LE, Beck LB, Noffsinger D et al. Efficacy of 3 commonly used hearing aid circuits: A crossover trial. NIDCD/VA Hearing Aid Clinical Trial Group. JAMA 2000;284(14):1806-13.
Ref ID: 111

2000-13 Parry BL, Javeed S, Laughlin GA, Hauger R, Clopton P. Cortisol circadian rhythms during the menstrual cycle and with sleep deprivation in premenstrual dysphoric disorder and normal control subjects. Biol.Psychiatry 2000;48(9):920-31.
Ref ID: 128

2000-14 Darbinyan V, Kteyan A, Panossian A, Gabrielian E, Wikman G, Wagner H. Rhodiola rosea in stress induced fatigue—a double blind cross-over study of a standardized extract SHR-5 with a repeated low-dose regimen on the mental performance of healthy physicians during night duty. Phytomedicine. 2000;7(5):365-71.
Ref ID: 130

2000-15 Stangier J, Su CA. Pharmacokinetics of repeated oral doses of amlodipine and amlodipine plus telmisartan in healthy volunteers. J Clin Pharmacol. 2000;40(12 Pt 1):1347-54.
Ref ID: 143

2000-16 Lacaille B, Julien P, Deshaies Y, Lavigne C, Brun LD, Jacques H. Responses of plasma lipoproteins and sex hormones to the consumption of lean fish incorporated in a prudent-type diet in normolipidemic men. J Am.Coll.Nutr. 2000;19(6):745-53.
Ref ID: 151

# Appendix C List of medical statistics books reviewed

## MEDICAL STATISTICS BOOKS WITH RECOMMENDATIONS TO ANALYSIS OF CROSS-OVER TRIALS (N= 26 )

| | |
|---|---|
| Altman-DG | Practical Statistics for Medical Research |
| Armitage-P and Colton-T | Encyclopedia of Biostatistics |
| Armitage-P, Berry-G | Statistical Methods in Medical Research |
| Bailar III-JC, Mosteller-F | Medical Uses of Statistics |
| Bithell-JF, Coppi-R | Perspectives in Medical Statistics-Proceedings of the European Symposium on Medical Statistics, Rome, 1980 |
| Bolton-S | Pharmaceutical Statistics: Practical & Clinical Applications |
| Buncher-CR, Tsay-JY | Statistics in the Pharmaceutical Industry |
| Chow-SC | Encyclopedia of biopharmaceutical statistics |
| Chow-SC, Liu-JP | Design and Analysis of Bioavailability and Bioequivalence Studies |
| Colquhoun-D | Lectures on Biostatistics: An Introduction to Statistics with Applications in Biology and Medicine |
| Everitt-BS | Statistical Methods for Medical Investigations |
| Friedman-LM, Furberg-CD, DeMets-DL | Fundamentals of Clinical Trials |
| Goldstone-LA | Understanding Medical Statistics |
| Ingelfinger-JA, Mosteller-F, Ware-JH, Thibodeau-LA | Biostatistics in Clinical Medicine |
| Jennison, Turnbull | Group Sequential Methods with Applications to Clinical Trials |
| Mainland-D, Sanders-WB | Elementary Medical Statistics |
| Matthews-JNS | An Introduction to Randomised Controlled Clinical Trials |
| Petrie-A | Lecture Notes on Medical Statistics |
| Piantadosi-S | Clinical Trials - A Methodologic Perspective |
| Pocock | Clinical Trials: A Practical Approach |
| Rosner-B | Fundamentals of Biostatistics |
| Smart-JV | Elements of Medical Statistics |
| Spriet-A, Dupin-Spriet-T, Simon-P | Methodology of Clinical Drug Trials |
| Whitehead-J | The Design and Analysis of Sequential Clinical Trials |
| Zolman-JF | Biostatistics: Experimental Design and Statistical Inference |

## MEDICAL STATISTICS BOOK WITH STATISTICAL APPLICATONS OF CROSS-OVER TRIALS (N=21)

| Medical Statistics books | |
|---|---|
| Bertelsmann-A | Drug Epidemiology (English-German Dictionary) |
| Biller-J, Bogousslavsky-J | Clinical Trials in Neurologic Practice 2001 |
| Bland-M | An Introduction to Medical Statistics |
| Bowers-D | Statistics from Scratch |
| Campbell-MJ, Machin-D | Medical Statistics A Commonsense Approach |
| Colton-T | Statistics in Medicine |
| Daly-LE, Bourke-GJ | Interpretation and Uses of Medical Statistics |
| Dawson-Saunders-B, Trapp-RG | Basic & Clinical Biostatistics |
| De Muth-JE | Basic Biostatistics and Pharmaceutical Statistical Applications |
| Elston-RC, Johnson-WD | Essentials of Biostatistics |
| Faragher-B, Marguerie-C | Essential Statistics for Medical Examinations |
| Fisher-LD, Van Belle-G | Biostatistics: A Methodology for the Health Sciences |
| Hill-AB | Principles of Medical Statistics |
| Holland-BK | Probability Without Equations: Concepts for Clinicians |
| Kirkwood-BR | Essentials of Medical Statistics |
| Lindsey-J | Revealing Statistical Principles |
| Motulsky-H | Intuitive Biostatistics |
| Pereira-Maxawell-F | A-Z of Medical Statistics: A Companion for Critical Appraisal |
| Riffenburgh-RH | Statistics in Medicine |
| Rimm-AA | Basic Biostatistics in Medicine and Epidemiology |
| Wooding-WM | Planning Pharmaceutical Clinical Trials - Basic Statistical Principles |

## MEDICAL STATISTICS BOOK WITHOUT MENTIONING CROSS-OVER TRIALS (N=83)

| | |
|---|---|
| Andersen-B | Methodological Errors in Medical Research |
| Armitage-P | Sequential Medical Trials |
| Bahn-AK | Basic Medical Statistics |
| Bailey-NTJ | Mathematics, Statistics and Systems for Health |
| Bailey-NTJ | The Mathematical Approach to Biology and Medicine |
| Bellman-R | Mathematical Methods in Medicine |
| Bowers-D | Statistics Further from Scratch: For Health Care Professionals |
| Brown, Swanson | Medical Statistics on Personal Computers |
| Brown-BW, Hollander-M | Statistics: A Biomedical Introduction |
| Castle-WM, Morth-PM | Statistics in Small Doses |
| Clark-GM | Statistics and Experimental Design-An Introduction for Biologists and Biochemists |
| Cleophas-TJ, et al | Statistics Applied to Clinical Trials 2000 |
| Coggon-D | Statistics in Clinical Practice |
| Daniel-WW | Biostatistics - A Foundation for Analysis In the Health Sciences |
| Duncan-RC | Introductory Biostatistics for the Health Sciences |
| Dunn-G | Statistics in Psychiatry |
| Dunn-G, Everitt-B | Clinical Biostatistics - An Introduction to Evidence Based Medicine |
| Dunn-OJ, Clark-V | Basic Statistics: A Primer for the Biomedical Sciences |
| Dwyer-JH, et al | Statistical Methods for Longitudinal Studies of Health |
| Eldridge-S, Ashby-D | Statistical Concepts (Master Classes in Primary Care Research No.3) |
| Elwood-M | Critical Appraisal of Epidemiological Studies and Clinical Trials |
| Emery-AEH | Methodology in Medical Genetics: An Introduction to Statistical Methods |
| Feinstein-AR | Clinimetrics |
| Forthoher-RN, Lee-ES | Introduction to Biostatistics: A Guide to Design, Analysis and Discovery |
| Gardner-MJ, Altman-DG | Statistics with Confidence: Confidence Intervals and Statistical Guidelines |
| Gehan-EA, Lemak-NA | Statistics in Medical Research: Developments in Clinical Trials |
| Glantz-SA (editor) | Primer of Biostatistics |
| Glaser-AN | High-Yield Biostatistics |
| Greenbery-NT | Medical Statistical Epidemiology |
| Harris-EK, Boyd-JC | Statistical Bases of Reference Values in Laboratory Medicine |
| Hawkins-B | Elements of Medical Statistics (yr1829) |
| Hirsch-RD, Riegelman-RK | Statistical Operations: Analysis of Health Research Data |
| Holm-S | Ethical Problems in Clinical Practice: The Ethical Reasoning of Health Care Professionals |
| Hoppensteadt-E, Peskin-CS | Mathematics in Medicine and the Life Sciences |
| Hulley-SB | Designing Clinical Research: An Epidemiology Approach |
| Jekel-JF, Elmore-FG, Katz-DL | Epidemiology, Biostatistics and Preventive Medicine |
| Jordan-K | Mastering Statistics: A Guide for Health Service Professionals & Researchers |
| Kachigan-SK | Multivariate Statistical Analysis – A Conceptual Introduction |
| Katz-MH | Multivariable Analysis: A Practical Guide for Clinicians |
| Kuzma-JW | Basic Statistics for the Health Sciences |
| Lange, et al | Case Studies in Biometry |
| Leaverton-PE | A Review of Biostatistics |
| Le-CT, Boen-JR | Health and Numbers: Basic Biostatistical Methods |
| Lewis-AE | Biostatistics (Reinhold Books in the Biological Sciences) |
| Matthews-DE, Farewell-VT | Using and Understanding Medical Statistics |
| McCall-J | Statistics: A Guide for Therapists |
| McFadden-F | Management of Data in Clinical Trials 1998 |
| Michelson-S, Schofield-T | The Biostatistics Cookbook: The Most User-Friendly Guide for the Biomedical Scientists |
| Miller-RG, Efron-B, Brown-BW, Moses-LE | Biostatistics Casebook |
| Morton-RF, Hebel-JR, McCarter-RJ | A Study Guide to Epidemiology and Biostatistics |
| Mould-RF | Introductory Medical Statistics |
| Munro-AJ | Clinical Trial Procedure: Notes for Doctors |
| Munro-BH | Statistical Methods for Health Care Research |
| Murphy-EA | Biostatistics in Medicine |
| Nimmo-W, Tucker-G | Clinical Measurement in Drug Evaluation |
| Norman-GK, Streiner-DL | PDQ Statistics |
| Norman-GR, Streiner-D | Biostatistics: the Bare Essentials |
| Petrie-A, Sabin-C | Medical Statistics at a Glance |

| | |
|---|---|
| Pipkin-FB | Medical Statistics Made Easy |
| Po-LW | Statistics for Pharmacists |
| Puri-BK | Statistics in Practice: An Illustrated guide to SPSS |
| Rao-CR, Chakraborty-R (editors) | Statistical Methods in the Biological and Health Sciences |
| Rees-DG | Essential Statistics for Medical Practice : A Case-Study Approach |
| Reid-NG, Boore-JRP | Research Methods and Statistics in Health Care |
| Roberts-EA | Sequential Data in Biological Experiment |
| Salsburg-D | The Use of Restricted Significance Tests in Clinical Trials |
| Sanders-DH | Statistics: A First Course |
| Selvin-S | Practical Biostatistical Methods |
| Senn-S | Statistical Issues in Drug Development |
| Shott-S | Statistics for Health Professionals |
| Shoukri-NW, Pause-CA | Statistical Methods for Health Sciences |
| Simon-W | Mathematical Techniques for Biology and Medicine |
| Sogliero-Gilbert-G | Drug Safety Assessment in Clinical Trials |
| Strike-PW | Medical Laboratory Statistics |
| Swinscow-TDV, Revised by Campbell-MJ | Statistics at Square One |
| Witten-M (editor) | Mathematical Models in Medicine: Diseases and Epidemics |
| Woolson-RF | Statistical Methods for the Analysis of Biomedical Data |
| Zar-JH | Biostatistical Analysis |

# Appendix E: The CONSORT Checklist

| PAPER SECTION And topic | Item | Description | Reported on Page # |
|---|---|---|---|
| *TITLE & ABSTRACT* | 1 | How participants were allocated to interventions (e.g., "random allocation", "randomised", or "randomly assigned"). | |
| *INTRODUCTION* Background | 2 | Scientific background and explanation of rationale | |
| *METHODS* Participants | 3 | Eligibility criteria for participants and the settings and locations where the data were collected. | |
| Interventions | 4 | Precise details of the interventions intended for each group and how and when they were actually administered. | |
| Objectives | 5 | Specific objectives and hypotheses. | |
| Outcomes | 6 | Clearly defined primary and secondary outcome measures and, when applicable, any methods to enhance the quality of measurements (e.g., multiple observations, training of assessors). | |
| Sample size | 7 | How sample size was determined and, when applicable explanation of any interim analyses and stopping rules. | |
| Randomization -- Sequence generation | 8 | Method used to generate the random allocation sequence, including details of any restrictions (e.g., blocking, stratification) | |
| Randomization -- Allocation concealment | 9 | Method used to implement the random sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned. | |
| Randomization -- Implementation | 10 | Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups. | |
| Blinding (masking) | 11 | Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated. | |
| Statistical methods | 12 | Statistical methods used to compare groups for primary outcomes(s); Methods for additional analyses, such as subgroup analyses and adjusted analyses. | |
| RESULTS<br><br>Participant flow | 13 | Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome. Describe protocol deviations form study as planned, together with reasons. | |
| Recruitment | 14 | Dates defining the periods of recruitment and follow-up. | |
| Baseline data | 15 | Baseline demographic and clinical characteristics of each group. | |
| Numbers analyzed | 16 | Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat". State the | |

| | | | |
|---|---|---|---|
| | | results in absolute numbers when feasible (e.g., 10/20, not 50%). | |
| Outcomes and estimation | 17 | For each primary and secondary, a summary of results for each group, and the estimated effect size and its prevision (e.g., 95% confidence interval). | |
| Ancillary analyses | 18 | Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indication those pre-specified and those exploratory. | |
| Adverse events | 19 | All important adverse events or side effects in each intervention group. | |
| DISCUSSION Interpretation | 20 | Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes. | |
| Generalizability | 21 | Generalizability (external validity) of the trial findings. | |
| Overall evidence | 22 | General interpretation of the results in the context current evidence. | |

## Appendix E Survey Questionnaire

# QUESTIONNAIRE for ANALYSIS of CROSS-OVER TRIALS

This questionnaire is being sent to one person per statistical unit or grouping as given in the PSI members list of November 2000. If you have received this questionnaire, it is because you are listed as a contact person, or (where no contact person is given) because you have been otherwise chosen.

Questions regarding 'your unit' refer to the group or unit is given in the PSI members list in which your name appears.

**Q1. How many clinical trial reports for which your unit was responsible for analysis have been signed off in the period 1st Jan – 31st Dec 2000?** ☐

**Q2. Among all clinical trials mentioned in Q.1 above, how many were of *cross-over* design?** (please give the answer in approximated percentage if the number of cases cannot be recalled) ☐

**Q3. Among all *cross-over* trials mentioned in Q.2, how many were two-period two-treatment (AB/BA) designs?** (please give the answer in approximated percentage if the number of cases cannot be recalled) ☐

**Q4. What is your approach to the possibility of carry-over effect in cross-over trial?**

☐ Carry-over effect should have been eliminated by design

☐ Carry-over effect is dealt with by statistical tests

☐ Carry-over effect should be eliminated by trial design and tested by statistics

☐ Carry-over effect cannot be dealt with and I am therefore reluctant to use cross-over designs

☐ Other (please state)

**Q5. What do you consider the most serious/difficult problem with statistical analysis of *cross-over* trials? (Please tick the category that applies.)**

| | Is Not a Problem | Provides some Difficulty | Is a Serious Problem |
|---|---|---|---|
| **Carry-over Effect** | | | |
| **Patient Effect** | | | |
| **Period Effect** | | | |
| **Other**<br><br>**Please specify:** | | | |

**Q6. Does your unit have a guideline or standard operating procedure that covers the analysis of *cross-over* trials?**

☐ Yes

☐ No

**Q7. Looking at AB/BA designs, and considering the primary analysis only, (in which either a confidence interval or p-value for the treatment effect is presented), please indicate below your approach to dealing with the factors listed by ticking the appropriate column.** [For example, if you always use the matched-paired approach (either t-test or Wilcoxon signed rank test) you should tick "always in the model" for patient effect and " never in the model" for the other effects.]

| | Always in the Model | Never in the Model | Include if Significant | Include if makes Treatment Significant | Other* |
|---|---|---|---|---|---|
| **Patient Effect** | | | | | |
| **Period Effect** | | | | | |
| **Carry-over Effect** | | | | | |

* Please explain what other strategy is used if any in the space below.

**Q8. Looking at designs with three or more periods and considering the primary analysis only, (in which either a confidence interval or p-value for the treatment effect is presented), please indicate below your approach to dealing with the factors given by ticking the appropriate column.**

| | Always in the Model | Never in the Model | Include if Significant | Include if makes Treatment Significant | Other* |
|---|---|---|---|---|---|
| **Patient Effect** | | | | | |
| **Period Effect** | | | | | |
| **First Order Carry-over Effect** | | | | | |
| **Second Order Carry-over Effect** | | | | | |

* Please explain what other strategy is used if any in the space below.

**Q9. Is the two-stage procedure for analysing AB/BA designs, as proposed by Grizzle and endorsed by Hills and Armitage, used in your unit?**

☐ Yes

☐ No

**Q10. The two-stage procedure for analysing AB/BA designs has raised some concerns about its statistical validity. The following statements are two main concerns regarding this procedure. What is your opinion on the following statements?**

**Q10.1 The Two-Stage Procedure is problematic because the test for carry-over lacks power**

☐ I Agree

☐ I Disagree

☐ I Don't Know

**Q10.2 Analysing cross-over trials by the Two-Stage Procedure generates a higher overall type I error rate than the nominal rate claimed**

☐ I Agree

☐ I Disagree

☐ I Don't Know

**Q11. Could you recommend some references on analysis of cross-over trials?**

**Q12. We would like to hear your opinions on analysis of cross-over trials**

**CONFIDENTIAL Supplementary Questions**

**Name* (optional):**

**Organisation* (optional):**

*Providing us with these details will help us check your answers should we have any queries. However we will not reveal results of individual answers.

## References


[1] Senn SJ. Cross-over Trials in Clinical Research Wiley: Chichester, 2002.
[2] Balaam LN. A two period design with two experimental units. Biometrics 1968; 24: 61-73.
[3] Senn SJ. The AB/BA crossover: past, present and future? Statistical Methods in Medical Research 1994; 3: 303-24.
[4] Grizzle JE. The two-period change over design and its use in clinical trials. Biometrics 1965; 21: 467-480.
[5] Grizzle JE. Correction to Grizzle (1965). Biometrics 1974; 30: 727.
[6] Grieve AP. The two-period changeover design in clinical trials [letter]. Biometrics 1982; 38: 517.
[7] Hills M and Armitage P. The two-period cross-over clinical trial. British Journal of Clinical Pharmacology 1979; 8: 7-20.
[8] Freeman P. The performance of the two-stage analysis of two-treatment, two-period cross-over trials. Statistics in Medicine 1989; 8: 1421-1432.
[9] Larson HJ and Bancroft TA. Biases in prediction by regression for certain incompletely specified models. Biometrika 1963; 50: 391-401.
[10] Senn SJ. Problems with the two stage analysis of crossover trials [letter; comment]. Br J Clin Pharmacol 1991; 32: 133.
[11] Senn SJ. Cross-over trials. Encyclopedia in Biostatistics 1998; 2: 1033-1049.
[12] Senn SJ. Crossover Design. Encyclopedia of Biopharmaceutical Statistics 2000; 142-149.
[13] Garfield E. The Uses and Limitations of Citation Data as Science Indicators - an Overview for Students and Nonspecialists - Introduction to Citation Data - Their Use as Quantitative Indicators for Science and Technology Evaluation and Policy Making by Garfield,Eugene and Welljams-Dorof,Alfred (Reprinted from Science and Public-Policy, Vol 19, Pg 321-327, 1992). Current Contents 1992; 49: 3-13.
[14] Redman J, Willett P, Allen FH and Taylor R. A citation analysis of the Cambridge Crystallographic Data Centre. Journal of Applied Crystallography 2001; 34: 375-380.
[15] Garfield E. Play the new game of twenty citations. Wherein ISI reveals the fifty most frequently cited non-journal items. Current Contents 1971;
[16] De Solla Price DJ. Networks of scientific papers. Science 1965; 149: 510-515.
[17] Nelder JA and Wedderburn RWM. Generalized Linear Models. Journal of the Royal Statistical Society A 1972; 132: 107-120.
[18] McCullagh P and Nelder JA. Generalized Linear Models Chapman and Hall: London, 1989.
[19] Jones B and Kenward MG. Design and Analysis of Cross-Over Trials Chapman and Hall,: London, 1989.
[20] Ratkowsky DA, Evans MA and Alldredge JR. Cross-over experiments : design, analysis, and application Marcel Dekker: New York, 1993.
[21] Cotton JD. Analyzing Within-Subject Experiments Lawrence Erlbaum Associates: Mahwah, New Jersey, 1998
[22] Senn SJ. The AB/BA cross-over: How to perform the two-stage analysis if you can't be persuaded that you shouldn't. Liber Amicorum 1996: 93-100

[23] Redman J et al, A citation analysis of Cambridge Crystallographic Data Centre, Applied Crystallography, 2001. 34, 375-380
[24] Borgmon C L (1980). Scholarly communication and bibliometrics. Sage
[25] Hamers L, Hemeryck Y, Herweyers G, Janssen M, Keters H, Rousseau R, Vanhoutte A. Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula. Information Processing & Management , 25, 1989, 315-318
[26] Garfield E. Essay of an information scientist, vol.2, p.184-187, 1974-76
[27] Bacchetti, Peer review of statistics in medical research: the other problem. *BMJ* 2002; 324: 1271-1273
[28] Garfield. Reprinted in Essays of an Information Scientist, Vol: 6, p580, 1983
[29] Senn SJ. The AB/BA crossover: past, present and future? Statistical Methods in Medical Research 1994; 3: 303-324
[30] Senn SJ. Some controversies in designing and analysing cross-over trials. Biocybernetics and Biomedical Engineering 1995 Vol.15 No.1-2 pp. 27-39
[31] International Conference of Harmonisation ICH-E9 Statistical Principles for Clinical Trials 1998
[32] Begg C et al. Improving the Quality of Reporting of Randomized Controlled Trials The CONSORT Statement. JAMA, 1996 Vol. 276, No.8: 637-639
[33] Bland JM, Altman DG. Misleading statistics: errors in text books, software and manuals. International Journal of Epidemiology 1988 Vol. 17, No.2: 245-247
[34] Felson-DT, et al Misuse of Statistics Methods in Arthritis and Rheumatism 1982 Versus 1967-68 1984. Arthritis and Rheumatism, Vol.27, No.9 September:1018-1022
[35] Cruess-DF, et al Review of Use of Statistics in the American Journal of Tropical Medicine and Hygiene for January-December 1988 Am-J-Trop-Med-Hyg, 41(6), 619-626
[36] The Cochrane collaborations. http://www.cochrane.org/resources/glossary.htm
[37] http://www.consort-statement.org/endorsements/journals/journals.html, last access: Feb 2005
[38] http://www.consort-statement.org/organisations/organisations.htm, last access: Feb 2005
[39] Cohlen BJ, Te Velde ER, Looman CW, et al. Crossover or parallel design in infertility trials? The discussion continues. Fertil Steril 1998;70:40 45.
[40] The Web of Science, http://wos.mimas.ac.uk, last access July 2005
[41] Johns Hopkins Health after 50 http://www.hopkinsafter50.com/html/silos/hypertension/wpARTICLE_sodium.html
[42] Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191-4
[43] http://www.garfield.library.upenn.edu/classics.html, last access August 2005
[44] Pocock SJ, Hughes MD, Lee RJ. Statistical Problems in the Reporting of Clinical Trials. A survey of three medical journals. New England Journal of Medicine. 317 (7): 426-32, 1987 Aug 13

[45] Jones BJ, Lewis J. The case for Cross-Over Trials in Phase III. Statistics in Medicine 1995 14: 1025-38
[46] Grieve AP, Senn S. Estimating Treatment Effects in Clinical Crossover Trials. Journal of Biopharmaceutical Statistics 1998; 8(2): 191-233
[47] Gore SM, et al. Misuse of Statistical Methods: Critical Assessment of Articles in BMJ from January to March 1976. British Medical Journal 1977; 1(8): 85-87
[48] DerSimonian R, et al. Reporting on Methods in Clinical Trials. New England Journal of Medicine 1982; 306 (22): 1332-37
[49] Armitage P, Hills M. The Two-Period Crossover Trial. Statistician. 1982 31(2)L 119-131