Cross-over Trials in Clinical Research

Second Edition

Stephen Senn

*Department of Statistical Science
and Department of Epidemiology and Public Health
University College London, UK*

Cross-over Trials
in Clinical Research

STATISTICS IN PRACTICE

Advisory Editor

Stephen Senn
University College London, UK

Founding Editor

Vic Barnett
Nottingham Trent University, UK

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Cross-over Trials in Clinical Research

Second Edition

Stephen Senn

*Department of Statistical Science
and Department of Epidemiology and Public Health
University College London, UK*

JOHN WILEY & SONS, LTD

Contents

4   Other outcomes and the *AB/BA* design   89

5   Normal data from designs with three or more treatments   157

6   Other outcomes from designs with three or more treatments   187

Preface to the Second Edition

The reception of the first edition of this work was much better than I dared hope. I took two uncompromising positions on the subject of carry-over and these went against much conventional wisdom. Despite this, many seemed to find the book helpful, and it is as a result of this positive response that a second edition is possible.

First, I condemned all strategies that relied on pre-testing for carry-over as a means of determining the final form of the analysis of the treatment approach. Such an approach had been extremely common when dealing with the *AB/BA* design, but I considered that the implications of Freeman's (1989) devastating examination of the two-stage procedure made it untenable. One reviewer misread this as meaning that I also disapproved of the *AB/BA* design itself, but this is incorrect. It is an opinion I never expressed. In fact, I consider that, where circumstances permit, an *AB/BA* cross-over is an extremely attractive design to use.

Second, I expressed extreme scepticism concerning common approaches to adjusting for carry-over that relied on simplistic models for it, in particular assuming that the carry-over from an active treatment into an active treatment would be the same as into placebo. Although I worked primarily in phases II to IV whilst employed by CIBA-Geigy, I came into contact with statisticians who worked in phase I on pharmacokinetic-pharmacodynamic (PK/PD) modelling, and it puzzled me that an approach that would have been considered naïve and wrong in an earlier stage of development could be accepted as reasonable later on. In this connection it seemed to me that the criticisms of the standard carry-over model which had been made by Fleiss (1986b, 1989) were unanswerable except by abandoning it.

I consider that with the nearly ten years that have passed since the first edition, these positions look more and more reasonable. In particular, it now seems to be more or less universally the case amongst those who research into the methodology of planning and analysing cross-over trials that the two-stage procedure has been abandoned as illogical. General medical statistics textbooks have lagged behind in this respect—but, within the pharmaceutical industry at

least, it seems to be well understood. For example, the ICH E9 (International Conference on Harmonisation, 1999) statistical guidelines, whilst discussing cross-over trials, do not require a two-stage analysis, despite the fact that as recently as the late 1980s, industry-based guidelines were recommending it. My PhD student, Sally Lee, did a survey of members of Statisticians in the Pharmaceutical Industry (PSI) in 2001 and found that in many firms this procedure was no longer being used. The position on adjusting for carry-over in higher-order designs has moved more slowly. Here my own view would still be characterized as extreme by some medical statisticians, although many pharmacokineticists and others involved in PK/PD modelling would regard it as reasonable and, indeed, natural. Nevertheless, the position looks less extreme than it did at the time of the first edition. Thus, in revising the book I have seen no need to revise these positions. Indeed, part of the revision consists of new material supporting them.

A feature of the first edition was that, although a great deal of space was devoted to explaining (mainly for didactic rather than practical reasons) how analysis could be carried out with a pocket calculator, the only statistical package whose use was described was SAS®. The second edition now includes, in appendices to several chapters, descriptions and code for performing analyses with GenStat® and S-Plus®. I am greatly indebted to Peter Lane and Roger Payne for help with the former and to my former colleagues at CIBA-Geigy, Andreas Krause and Skip Olsen, for help with the latter. I am also very grateful to Kelvin Kilminster and Julie Jones for help with SAS®. Any infelicities of coding that remain are my fault. Please note also that where I say that a particular analysis cannot be carried out with a particular package, I mean that my search of the help file and manual has failed to find a way to do it. No doubt all these packages have resources I did not discover.

Also included now are descriptions of analysis with Excel®, in particular with the help of the excellent add-in StatPlus® (Berk and Carey, 2000), and, for non-parametrics, with StatXact®. As regards the latter, I am particularly grateful to CYTEL for having provided me with a copy of this superb software.

In his generally positive review of the first edition (Gough, 1993) the late Kevin Gough remarked that it was a shame that recovering inter-block information in incomplete blocks designs had not been included. This has now been rectified. I have also added descriptions of the use of SAS® and GenStat® for finding sequences for designs, analysis of frequency data using Poisson regression, an explanation of how to remove the bias inherent in the two-stage procedure (together with a recommendation to avoid it altogether!), Bayesian analysis of the *AB/BA* design, permutation tests, more material on binary data, including random effect modelling, and survival analysis, as well as reviews of more recent literature on most topics.

As was the case with the first edition, I have marked some passages with an asterisk (*). This is either because the material is rather more difficult than the

general level of the book or because, whatever its theoretical interest, its utility to the analyst is low.

In the time since the first edition I have acquired as co-authors in various papers on cross-over trials Pina D'Angelo, Farkad Ezzet, Dimitris Lambrou, Sally Lee, Jürgen Lilienthal, Francesco Patalano, Diane Potvin, Bill Richardson, Denise Till and, on several occasions, Andy Grieve. I am most grateful to all of these for their collaboration. I am also particularly grateful to Andy Grieve for many helpful discussions over the years on this topic, to John Nelder for many helpful discussions on modelling and to John Matthews, Gary Koch, Dimitris Lambrou, Kasra Afsaranijad and the late Kevin Gough for helpful comments on the first edition. My thanks are due also to Schein Pharmaceuticals for having sponsored some research into carry-over in three-period designs.

Since the first edition was published, I have left the pharmaceutical industry. This has made it much more difficult for me to obtain data to illustrate the use of cross-over trials. I am therefore particularly grateful to Sara Hughes and Michael Williams of GlaxoSmithKline for providing me with data to illustrate the use of Poisson regression, to John Guillebaud and Walli Bounds of University College London for providing me with data to illustrate generalized linear mixed models and to Bob Shumaker of Alpharma for giving permission to cite data from Shumaker and Metzler (1998) to illustrate individual bioequivalence.

At Wiley I thank Helen Ramsey, Siân Phillips, Rob Calver, Richard Leigh and Sarah Corney for their support, and in particular Sharon Clutton for having encouraged me to undertake the revision. It was with my wife Victoria's support and encouragement that I came to write the first edition and, now as then, I should like to thank her for setting me on the path to do so in the first place.

Finally, I should like to leave this word of encouragement to the reader. Cross-over trials are not without their problems and there are indications and occasions where their use is not suitable. Nevertheless, cross-over trials often reach the parts of drug development and medical research other designs cannot reach. They are extremely useful on occasion, and this is always worth bearing in mind.

Stephen Senn
Harpenden
January 2002

Preface to the First Edition

This book has been written for two types of reader. The first is the physician or biologist who carries out or wishes to carry out cross-over trials and either has to analyse these trials himself or needs to interpret the analyses which others perform for him. I have kept in my mind whilst writing the book, as a concrete example of such a person, a hospital pulmonologist carrying out research into the treatment of asthma but with a lively amateur interest in statistics. The second type of reader is the applied statistician with no particular experience as of yet in cross-over trials but who wishes to learn something about this particular field of application. The concrete example I have kept in mind here is that of a statistician working in the pharmaceutical industry whose experience to date has been largely with parallel group trials but now has to work with cross-overs.

Obviously, the needs and capabilities of these two sorts of readers are not identical. The former will occasionally find the book hard going and the latter will sometimes find it elementary and wonder why I have been at such pains to describe with words what I might have said with ease with symbols. I beg each type of reader's pardon but ask them to use the book according to their taste. The former will find that no harm comes of skipping what little algebra there is in the book. The latter may easily cut the discussion of what is grasped at once. In particular I advise the former not to bother reading Chapter 2 unless inspired by the rest of the book to do so. The other passages I have marked with an asterisk (*) are those whose practical importance is not sufficient to justify study by those who are anxious to master relevant techniques in the minimum of time.

In Chapter 10, I have covered various matters which others consider important but I do not. In my opinion this chapter may also be omitted by the reader who is only concerned with practical matters.

I have also deliberately used different styles in writing the book. Few topics in clinical trials are as controversial as the cross-over trial. The history of the two-stage analysis of the *AB/BA* design and the recent revolution in attitudes to it brought about largely by Freeman's (1989) paper are a good example. It would

be dishonest and misleading for any author in this field always to maintain the impersonal voice of neutral infallibility. Accordingly, although I have either adopted the passive form or used 'we' where I feel confident that most or at least many statisticians would agree with what I have to say, where I am aware that some authorities on cross-over trials would disagree, I have deliberately used 'I'. In such cases, the intention is not to irritate the reader but to warn him that others hold different opinions.

I also apologize to all female readers for only using masculine personal and possessive pronouns and adjectives. I am, of course, aware that many physicians, biologists and statisticians, not to mention patients, are female. In using 'he' and 'his' I plead guilty to the sin of linguistic laziness but I have not meant to offend.

There are many practical worked examples in the book. Most come from one very large project comprising over 100 clinical trials, most of them cross-overs, with which I have been involved. This is deliberate. I do not wish to help perpetuate the notion that definitive statements regarding treatments are usually produced using single trials. In drug development in particular, and in science more generally, it is the cumulative effect of analysing studies with different designs and related objectives which constitutes the true advance in our knowledge. This is particularly true in cross-over trials where the problem of carry-over may make the interpretation of individual studies problematic but where the diseases studied are of necessity not life threatening and therefore the ethical problems associated with repeating studies are not severe.

I must also point out that, although in all cases I have presented what I believe to be reasonable analyses of the trials I have used as examples, the analysis presented is in most cases neither the only reasonable approach nor necessarily the one used for reporting the trial. My task in using these examples for writing this book has been made much easier than it would otherwise have been by the fact that the trials have already been reported on. I should like to thank my colleagues Nathalie Ezzet, Nicole Febvre, Roman Golubowski, Roger Heath, Hartwig Hildebrand, Denise Till and Elizabeth Wehrle, as well as Rolf Hambuch of the Institute for Biometrics in Freiburg in Breisgau and Anne Whitehead of Reading University, for work on these trials.

I also have to thank my colleagues Petra Auclair and Hartwig Hildebrand for collaborating on papers whose results I have partly incorporated in this book, Harald Pohlmann for contributing SAS® code and Bernhard Bablok, Farkad Ezzet, Nathalie Ezzet, Albert Kandra and Gunther Mehring for helpful comments. I am also very grateful to Vic Barnett, Peter Freeman and Robin Prescott for comments on individual chapters, to Tony Johnson for extremely helpful and encouraging comments on every aspect of the book and to Gilbert Rutherford for careful reading of the first draft. The basic idea for Chapter 6 grew out of a hint of Gary Koch's, and Peter Freeman proposed an analysis for the *AB/BA* design with single baselines which had not occurred to me. I am also grateful to

1
Introduction

1.1 THE PURPOSE OF THIS CHAPTER

In clinical medicine, cross-over trials are experiments in which subjects, whether patients or healthy volunteers, are each given a number of treatments with the object of studying differences between these treatments. The commonest of all such designs is one in which approximately half of the patients are first given an active treatment or *verum* and on a subsequent occasion a dummy treatment or *placebo* whereas the rest of the patients are first given placebo and then on a subsequent occasion verum. This is a simple example of a type of design which we shall consider in detail in Chapter 3.

The purpose of this chapter, however, is simply to provide some gentle exposition, in very general terms, of some features of cross-over trials. In particular we shall:

- define cross-over trials;
- explain why they are performed;
- mention clinical specialties for which they are useful;
- point to some dangers and difficulties in performing them;
- as well as explain some general attitudes which will be adopted throughout the book.

Methods for analysing cross-over trials will not be dealt with in this chapter but form the subject matter of Chapters 3 to 7 inclusive. The fact that we defer the issue of analysis until later enables us to begin the discussion of cross-over trials with the help of a very famous (but relatively complex) example, of considerable historical interest, which we now consider below.

1.2 AN EXAMPLE

Example 1.1 Cushny and Peebles (1905) reported the results of a clinical trial conducted on their behalf by Richards and Light of the effect of various optical isomers on duration of sleep for a number of inmates of the

Michigan Asylum for Insane at Kalamazoo. Three treatments in tablet form were examined:

- laevorotatory hyoscine hydrobromate, 0.6 mg (which we shall refer to either as *L-hyoscine HBr* or *laevo-hyoscine*);
- racemic hyoscine hydrobromate, 0.6 mg (*R-hyoscine HBr* or *racemic hyoscine*);
- laevorotatory hyoscyamine hydrobromate, 0.6 mg (*L-hyoscyamine HBr* or simply *hyoscyamine*).

Patients were given each of these treatments on a number of evenings and also studied on a number of control nights for which no treatment had been administered. According to the authors,

> As a general rule a tablet was given on each alternate evening ... (on) the inter-vening control night ... no hypnotic was given. Hyoscyamine was thus used on three occasions and then racemic hyoscine, and then laevo-hyoscine. Then a tablet was given each evening for a week or more, the different alkaloids following each other in succession (Cushny and Peebles, 1905, p. 509.)

Table 1.1 summarizes the results in terms of hours of sleep for the patients studied.

Remark   These data are of particular historical interest not only because Cushny was a pioneer of modern pharmacology and did important work on optical isomers (Parascondola, 1975) but because they were quoted (incor-rectly) by Student (1908) in his famous paper, 'The probable error of a mean'. The data were in turn copied from Student by Fisher (1990a), writing in 1925, and used as illustrative material in *Statistical Methods for Research Workers*.

Table 1.1   (Example 1.1) Number of observations and mean hours of sleep by treatment and patient in a trial of hypnotics.

	Treatment							
	Control		0.6 mg L-Hyo-scyamine HBr		0.6 mg L-Hyo-scine HBr		0.6 mg R-Hyo-scine HBr	
Patient	Number	Mean	Number	Mean	Number	Mean	Number	Mean
1	9	0.6	6	1.3	6	2.5	6	2.1
2	9	3.0	6	1.4	6	3.8	6	4.4
3	8	4.7	6	4.5	6	5.8	6	4.7
4	9	5.5	3	4.3	3	5.6	3	4.8
5	9	6.2	3	6.1	3	6.1	3	6.7
6	8	3.2	4	6.6	3	7.6	3	8.3
7	8	2.5	3	6.2	3	8.0	3	8.2
8	7	2.8	6	3.6	6	4.4	5	4.3
9	8	1.1	5	1.1	6	5.7	5	5.8
10	9	2.9	5	4.9	5	6.3	6	6.4
11	—	—	2	6.3	2	6.8	2	7.3

These data thus have the distinction of having been used in the paper which inaugurated the modern statistical era (since it was the first to deal explicitly with small sample problems) and also in what is arguably the single most influential textbook written on the subject. (See Plackett and Barnard, 1990, and Senn and Richardson, 1994, for historical accounts.)

The particular feature of these data which is of interest here, however, is that they were obtained by giving each of a number of subjects a number of treatments to discover something about the effects of individual treatments. They thus come from what we would now call a *cross-over trial* (sometimes also called a *change-over trial*) which we may now define as follows.

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).

Remark   It is probable that the word cross-over has come to be used for trials in which patients are given a number of treatments, because in the commonest type of trial of this sort (see Section 1.1 above) two treatments *A* and *B* (say) are compared. Patients are given either *A* or *B* in the first period and then 'crossed over' to the other treatment. In more complicated designs, however, such simple exchanges do not occur but the word cross-over is nevertheless employed. The essential feature of the cross-over is not crossing over *per se* but is as captured in our definition above.

Further remark   Note that the fact that patients in a given clinical trial are assigned to sequences of treatment does not alone cause the trial to be a cross-over. For example in clinical trials in cancer it is usual for patients to be given many treatments: some simultaneously, and some sequentially. In a trial investigating a new therapy, patients might well be assigned in the first instance either to a standard first-line treatment or to the new therapy with the purpose of studying the difference of the effects of treatment on remission. Patients who failed to respond or suffered a relapse would then be given alternative therapies and so on. At the end of the trial the difference between the effects of the new and alternative therapy on time to remission (or relapse) might form the object of an analysis. We could then regard the patients as having been allocated different *treatments* with the purpose of studying differences between them. Alternatively, we might study the effect on survival as a whole of allocating the patients to the different sequences (starting with new or alternative therapy). We would then be examining the difference between *sequences*. For neither of these two ends could we regard ourselves as having conducted a cross-over trial.

Before going on to consider cross-over trials in more detail some general points may usefully be made using the Cushny and Peebles data quoted in Example 1.1.

The first point to note is the *ethical* one. It may reasonably be questioned as to whether the initial investigation of substances of this sort ought to be carried out in the mentally ill who may be not in a position to give their free consent on the basis of an understanding of the potential risks and benefits of the trial. Such agreement on the part of the patient is referred to as *informed consent* in the literature on clinical trials. (It should be noted, however, that Cushny and Peebles tried the drugs out on themselves first. A similar step is undertaken in modern pharmaceutical development where so-called Phase I trials are undertaken with healthy volunteers in order to establish tolerability of substances.) Ethical considerations provide an important constraint on the design of all clinical trials and cross-over trials are no exception. This is a point which should constantly be borne in mind when designing them. In particular the fundamental right, which should not only be granted to all patients but also made clear to them, to be free to withdraw from a trial at any time, is one which can, if exercised, cause more problems of interpretation for cross-over trials than for alternative designs.

The second point to note concerns *purpose*. The trial had a specific scientific purpose. Cushny and Peebles wished to discover if there were any differences between two optical isomers: laevorotatory (L) and dextrorotatory (D) hyoscine HBr. For practical reasons the differences had to be inferred by comparing the effect of L-hyoscine HBr to the racemic form (the mixture of L and D), R-hyoscine HBr. This pharmacological purpose of the trial was of more interest to Cushny and Peebles than any details of treatment sequences, patient allocation or analysis. I mention this point because in my opinion some of the methodological research in cross-over trials over the past few decades can be justified more easily in terms of mathematical interest *per se* rather than in terms of its utility to the practising scientist.

Nevertheless, the third point to note concerns *sequences* of treatments. These were not necessarily wisely chosen and in any case are not clearly described. If we label a control night, *C*, L-hyoscyamine HBr, *X*, R-hyoscine HBr, *Y*, and L-hyoscine HBr, *Z*, it seems that the general rule was to use a sequence:

$$X \ C \ X \ C \ X \ C \ Y \ C \ Y \ C \ Z \ C \ Z \ C \ Z \ C \ X \ Y \ Z \ X \ Y \ Z \ X \ Y \ Z$$

(Preece, 1982). This would certainly produce the number of observations recorded for patients 1 and 2, although not for any other. If there were any general tendency for patients to improve or deteriorate such a scheme would bias comparisons of treatments since, for example, *Z* is on average given later than *X*.

Despite this criticism, the fourth point, which relates to the proper *interpretation* of this trial, is to note that the conclusions of Cushny and Peebles (1905), which are that 'hyoscyamine is of no value in the dose given as a hypnotic, while the laevorotatory and racemic forms of hyoscine have about the same influence in inducing sleep' (p. 509), being based on the right data, are probably not unreasonable. This may be seen by studying Figure 1.1. The figure gives the
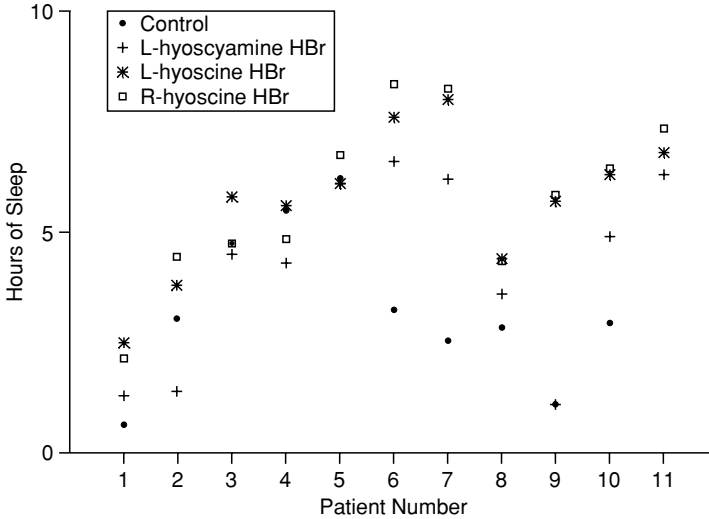
Figure 1.1 (Example 1.1) The Cushny and Peebles data. Mean hours of sleep for 11 patients for three active treatments and a control.

mean hours of sleep for the three treatments as well as the controls for patients number 1 to 10. If we compare the four results for each patient with each other, we shall see that the on the whole the values for the two forms of hyoscine are the highest of the four obtained but similar to each other. On the other hand, Student and Fisher, using incorrectly labelled data, concluded that there was a difference between optical isomers of hyoscine HBr. The moral is that the contribution to correct conclusions made by good data is greater than that made by sophisticated analysis. (In making this point I mean no disrespect to either Student or Fisher.)

The final point concerns *conduct* of the experiment. It may be noted that the patients did not each receive an equal number of treatments. Whether this was through careless planning or accident in execution one cannot say but the result is that the data bear the hallmarks of a real experiment: the data are imperfect. Missing observations continue to be one of the major problems in interpreting clinical trials, and cross-overs are no exception.

1.3 WHY ARE CROSS-OVER TRIALS PERFORMED?

We mentioned in Section 1.2 that not all trials in which patients are assigned to sequences of treatments are cross-over trials. For the trial to be a cross-over the sequences have to be of incidental interest and the object of the trial must be to study differences between the individual treatments which make up the sequences.

This was, in fact, the purpose of the trial in hyoscines reported as Example 1.1 above. Here the sequence in which the patients were given the drugs was not of interest. In fact, as we may deduce from their conduct and reporting of the trial, Cushny and Peebles (1905) probably considered the sequence in which the individual treatments were allocated as being of no consequence whatsoever. (As we pointed out above this is not always a wise point of view to take but, on the other hand, not always as disastrous as some modern commentators imply.) The purpose of the trial was to investigate the difference between the individual treatments. It is this which makes it a cross-over trial.

It is instructive to consider an alternative procedure that might have been used above. Each patient could have been assigned one treatment only. We should then have a *parallel group* trial. If we ignore the observations for patient 11 above it would thus have been necessary to study 40 (i.e., 4×10) patients to have obtained as many mean results per treatment as were obtained above. Even so the information would not have been as useful. In looking at Figure 1.1 it is noticeable that on the whole (there were some exceptions) patients who had high control values had high values for the three treatments. This point can be brought out by recasting the data (as Peebles and Cushny did, in fact, themselves) in the form of differences to control as has been done in Table 1.2 below. The data are also shown in this form in Figure 1.2. (No control values for patient 11 having been recorded, he is omitted from this table and figure.)

Just presenting the data in this form is revealing. Immediately it highlights the relatively poor performance of L-hyoscyamine HBr compared to the two forms of hyoscine HBr. (Even more revealing for this purpose, of course, would be calculating the difference between these treatments for each patient.) This feature of the data has been brought out by using every patient as his own control, a device which permits a particular source of variation, *between-patient*

Table 1.2 Mean hours of sleep per patient expressed as a difference from the mean obtained for the control.

Patient	Treatment		
	0.6 mg L-Hyo-scyamine HBr	0.6 mg L-Hyo-scine HBr	0.6 mg R-Hyo-scine HBr
1	0.7	1.9	1.5
2	−1.6	0.8	1.4
3	−0.2	1.1	0.0
4	−1.2	0.1	−0.7
5	−0.1	−0.1	0.5
6	3.4	4.4	5.1
7	3.7	5.5	5.7
8	0.8	1.6	1.5
9	0.0	4.6	4.7
10	2.0	3.4	3.5
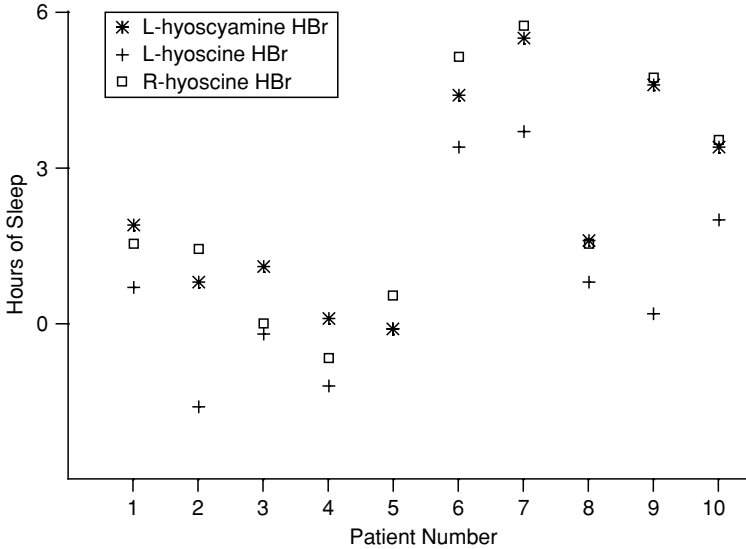
Figure 1.2   (Example 1.1) Mean hours of sleep for three active treatments expressed as a difference from control.

variation, to be eliminated. Thus, we can see that, although when the L-hyoscine HBr values and the control values from Table 1.1 are mixed together there is considerable overlap, seven of the values under treatment being lower than the highest control value, yet only one of the differences, that for patient 5, is negative.

These then are the main reasons why a cross-over trial may be preferred to a parallel group trial. First, to obtain the same number of observations fewer patients have to be recruited. Second, to obtain the same precision in estimation fewer observations have to be obtained. A cross-over trial can thus lead to a considerable saving in resources.

1.4   WHAT ARE THE DISADVANTAGES OF CROSS-OVER TRIALS?

There are disadvantages as well as advantages to the cross-over trial when compared to the parallel group trial. It is worth considering what these are.

First, there is the problem of *drop-outs*. These are patients who discontinue their programme of treatment before the trial is complete. Drop-outs cause difficulties for analysis and interpretation in parallel group trials as well but here at least the time until discontinuation for a patient may yield information which can be recovered. In cross-over trials this is extremely difficult to do and of course the patient can provide no direct information on the treatments

he did not even start if, for example, he drops out during the first treatment period.

Second, there are many conditions, or *indications*, for which cross-over trials would be a quite unsuitable approach. Obviously any disease in which there is a non-negligible probability that the patient will die during the period of observation is totally unsuited for study through a cross-over trial but so, more generally, is any condition in which the patient may be expected to suffer considerable deterioration or improvement during the course of treatment. This, for example, usually makes infectious diseases (which are diseases which may on the one hand prove fatal and on the other for which the patient may be cured), an unsuitable field for cross-over trials.

Third, there is a problem which is related to that above, namely that of *period by treatment interaction*, a phenomenon which occurs if the effect of treatment is not constant over time. If it is likely that the period in which a treatment is given will modify to any important degree the effect of that treatment then not only may a given cross-over trial become difficult to interpret but the very problem itself may be difficult to detect. There is thus the danger that the investigator or 'trialist' may confidently make incorrect assertions. One such cause of period by treatment interaction is that of *carry-over*, which may be defined as follows.

Definition   Carry-over is the persistence (whether physically or in terms of effect) of a treatment applied in one period in a subsequent period of treatment.

Remark   If carry-over applies in a cross-over trial we shall, at some stage, observe the simultaneous effects of two or more treatments on given patients. We may, however, not be aware that this is what we are observing and this ignorance may lead us to make errors in interpretation. This topic will be covered in more detail below and will not be discussed further at this point.

Fourth, there is the problem of *inconvenience to patients*. Cross-over trials may place the patients at particular inconvenience in that they are required to submit to a number of treatments and the total time they spend under observation will be longer. It should be noted, however, that this particular feature can sometimes be turned to advantage in that it may be of interest for a patient to have the opportunity to try out a number of treatments for himself in order to gain personal experience of their effects.

Finally, there is a difficulty of *analysis*. Although there is a considerable and growing literature on cross-over trials, it is true to say that there are a number of problems still lacking totally satisfactory algorithms for their solution. For example, a type of measurement commonly encountered in clinical trials is the so-called 'ordered categorical outcome'. Such outcomes are obtained when measurements are made using a rating scale such as: poor, moderate, good. There are no easy ways of analysing such outcomes for cross-over trials with three or more treatments.

1.5   WHERE ARE CROSS-OVER TRIALS USEFUL?

Cross-over trials are most suited to investigating treatments for ongoing or chronic diseases: for such conditions where there is no question of curing the underlying problem which has caused the illness but a hope of moderating its effects through treatment. A particularly suitable indication is asthma, a disease which may last for a lifetime and remain relatively stable for years. Rheumatism is another suitable condition, as is migraine. Mild to moderate hypertension and epilepsy (chronic seizures) are also conditions in which cross-over trials are frequently employed.

Even within these areas, however, the use of a cross-over design may be more or less appropriate according to the question being investigated. Thus *single-dose trials*, in which the patient is given a single administration of the treatment under study at any particular time, even if he may be subsequently crossed over to other treatments, are usually more suitable than long-term trials in which the patient is given regular repeated administrations of the same treatment. For the latter, the danger of patients dropping out, the possibility that repeated dosing may cause difficulty with carry-over and the sheer total amount of time that may be necessary to give one patient a number of therapies may make the cross-over trial an unwise choice.

Again certain types of therapy may lend themselves more easily to cross-over trials. Thus in asthma, bronchodilators (a class of drugs which has a rapid, dramatic and reversible effect on airways) are more suitable candidates than are steroids, which have a less marked but also more persistent effect.

Cross-over trials are also very popular in single-dose *pharmacokinetic* and *pharmacodynamic* studies in healthy volunteers as well as in Phase I tolerability studies and trials for *bioequivalence*. We shall not define or discuss such studies further here. Pharmacokinetics, pharmacodynamics and bioequivalence are topics which are discussed in Chapters 7 and 10.

1.6   WHAT ATTITUDE TO CROSS-OVER TRIALS WILL BE ADOPTED IN THIS BOOK?

The basic attitude towards cross-over trials which will be adopted in this book is one of cautious optimism. There are certain problems which may occur with cross-over trials which are less likely to cause problems with parallel trials but it is quite wrong to regard cross-over trials as uniquely problematical as has been suggested in some commentaries. There are many areas in which cross-over trials are particularly useful and in such cases what the practitioner needs are simple, useful, techniques for analysing a variety of outcomes as well as practical advice regarding planning of trials. *I shall assume*, in fact, that in designing a trial the practitioner has a particular background scientific question in which

he is interested (as had Cushny and Peebles) and that his interest in a given analytical technique is purely in terms of its utility to him in answering this question. A consequence of this is that a particular attitude will be adopted towards the problem of carry-over which I shall now explain.

1.7 CARRY-OVER

The problem of *carry-over* or more generally *period by treatment interaction* is taken by many commentators to be the outstanding problem of cross-over trials. It is worth considering first of all, therefore, whether this problem can cause difficulties elsewhere as well. This point can be examined with the help of an example.

Salbutamol is a standard treatment (at the time of writing undoubtedly the most popular of its class, that of beta-agonists) for patients suffering from asthma. If a new beta-agonist is to be introduced on to the market it will certainly be tested at some stage or other against salbutamol. Consider the case of a trial of a more newly developed beta-agonist, formoterol, against salbutamol.

Suppose this were done using a very simple type of cross-over in which half the patients were given salbutamol for a fixed period followed (possibly) by a period in which no treatment was given (a so-called wash-out, to be defined more precisely in Section 1.8) and then given formoterol, the other half being given formoterol first, followed by the wash-out and then the treatment with salbutamol. This sort of design is sometimes referred to as the *two-treatment, two-period cross-over*, or alternatively (and more precisely) as the *AB/BA cross-over* (where in this case *A* could be salbutamol and *B* formoterol). A particular problem which could occur with this trial is that the effect of one or other of the treatments in the first period might be such that by the end of the wash-out the patients would not be in the state they would have been in had they not been given treatment.

This might occur in a number of ways. For example there might be a physical persistence of the drug or the drug might have shown a curative effect. These are both examples of types of carry-over. In the former case there is the danger of a drug–drug interaction occurring and in the latter the second treatment may appear to benefit the patient when in fact the previous treatment is responsible. The consequence of these types of carry-over would be to bias the estimates of the effect of treatment.

Alternatively, during the time in which the cross-over trial is run the condition of the patients might suffer a *secular change* (some factor other than treatment might slowly be affecting the condition of most patients) and the benefit (or otherwise) of one drug compared to the other might be dependent on the current state of the patient. This would provide a case of period by treatment interaction. Again interpretation might be problematical.

It has been regularly overlooked, however, that for the sort of conditions in which cross-over trials are commonly used similar problems cannot be ruled out for parallel group trials. For example, it is fairly common in *'long-term' parallel-group trials* of asthma to treat patients for a year only (and frequently no longer than three months) with a view to being able to make recommendations regarding much longer periods of therapy. If the results of the trial are to be used with confidence, therefore, it must be believed that the effect of treatment beyond one year is the same as it is during the year. Obviously if the effect of treatment wears off over time—a phenomenon known as 'tachyphilaxis' (Holford and Sheiner, 1981)—then this form of period by treatment interaction may cause the results from such a trial to be quite misleading.

Again, the patients entering a parallel group trial may well have been under previous treatment. In asthma most of them will have been receiving salbutamol for many years. If there is salbutamol carry-over, then only under the very special assumption that this will be purely additive (i.e. that it will be the same regardless of which therapy follows) will this be unproblematical.

In fact, tachyphilaxis is suspected of occurring for beta-agonists, although it has usually been considered (fortunately) to be more important in terms of cardiac side-effects than in terms of efficacy. If, however, for the example above both salbutamol and formoterol showed tachyphilaxis for efficacy, then the results might be misleadingly disadvantageous for salbutamol since patients at the end of one year's apparent treatment with salbutamol would, in fact, if one takes account of their pre-trial experience, have been treated for many more. A possible consequence of this would be that the apparent advantage to formoterol could be reversed at some future date by studying a population which had previously been treated with formoterol.

Thus carry-over and period by treatment interaction are not uniquely a problem for cross-over trials as is sometimes claimed. They may affect parallel group trials as well. Nevertheless there are probably more occasions when carry-over in particular might more plausibly affect cross-over trials. It is worth considering, therefore, what may be done about it.

1.8   WHAT MAY BE DONE ABOUT CARRY-OVER?

It is easier to look first of all at what may not be done.

For many years the standard recommended analysis of the *AB/BA* cross-over was the so-called *two-stage procedure* (Grizzle, 1965; Hills and Armitage, 1979). This will be considered in more detail in Chapter 3 below, not because it may be recommended as a form of analysis, but because it is worth studying to bring home the dangers of *pre-testing* (i.e. carrying out preliminary tests of assumptions before choosing a substantive model). For the moment it is sufficient to describe it as follows. The two-stage procedure consists first of all of performing a statistical test on the data to examine the possibility of carry-over having

occurred. If it is not judged to have occurred then a within-patient test, whereby each patient's result under treatment A is referred to his result under B, is performed. If it is judged to have occurred then, on the basis that carry-over could not possibly have affected the values in the first period, a between patient test is carried out on the first period values, comparing the results under treatment *A* for one group of patients to the results under *B* of the other group.

The overall performance of this procedure has now been studied in depth by Freeman (1989) and has been shown to be vastly inferior in almost any conceivable circumstance to the simple alternative of always doing the within-patient test. An explanation as to why this is so must wait until Chapter 3 but a simple analogy with medicine may be helpful at this point. The initial test for carry-over in the two-stage procedure is similar to a screening test for a medical condition. It has a false positive and false negative rate associated with it. Furthermore it turns out that the 'cure' one would envisage for a case known to require treatment has a high probability of being fatal when the disease is absent. Because of this the conservative approach of not screening at all turns out to be best.

There are other similar two-stage, or even *multi-stage*, testing procedures which have been proposed for more complicated cross-over designs than the *AB/BA* design. It is not known that these approaches definitely perform as badly as the two-stage approach for the *AB/BA* cross-over. It is also not known, however, that these approaches are safe and it is known that the problem which arises with the two-stage procedure is potentially a problem for all pre-testing procedures. Accordingly in this book I shall not describe any multi-stage testing procedures.

We shall at some points indicate how tests for carry-over may be performed. Regarding this, however, it is appropriate to issue the following warnings. First, that the reader should on no account consider modifying a proposed analysis for the purpose of estimating or testing a treatment effect on the basis of the result of a test for carry-over performed on the same data. Second, that the reader should be extremely cautious about interpreting the results of such tests. They are virtually impossible to interpret reasonably independently of the treatment effect and this is true even for designs and models where the carry-over and treatment estimates may be assumed to be independent. For these reasons I never test for carry-over myself.

Another approach which is popular for more complex designs than the *AB/BA* design has been to include parameters for carry-over and estimate treatment and carry-over effects simultaneously: that is to say, estimate treatment in the presence of carry-over and vice versa. This approach suffers, however, from the fundamental flaw that it is necessary to make restrictive assumptions about the nature of the possible carry-over in order to model the phenomenon successfully. These assumptions are not at all reasonable (Fleiss, 1986b) and involve, for example, in a dose-finding trial assuming that the carry-over of effect from the highest to the lowest dose is the same as that from the highest to

the next-highest. Furthermore, it has been shown (Senn, 1992; Senn and Lambrou, 1998) that if slightly more realistic forms of carry-over apply, then using these models and their associated designs can actually be worse than doing nothing at all about carry-over.

Again, although these models are considered briefly in Chapter 10, I must issue the following warnings. First, it must be clearly understood that these models cannot guarantee protection against realistic forms of carry-over adversely affecting treatment estimates. Second, I can think of no cases where the assumptions made under these models would even approximately apply unless the carry-over were so small as to be ignorable anyway. And third, that such models often impose a penalty in efficiency of estimates. For these reasons I never use them myself.

The third approach to dealing with carry-over is that of using a *wash-out period*. This may be defined as follows.

Definition   A wash-out period is a period in a trial during which the effect of a treatment given previously is believed to disappear. If no treatment is given during the wash-out period then the wash-out is *passive*. If a treatment is given during the wash-out period then the wash-out is *active*.

When a wash-out period is employed it is assumed that all measurements taken after the wash-out are no longer affected by the previous treatment. If a passive wash-out is employed the patient is assumed to have returned to some natural background state before the next treatment is started. For example in the case of single-dose trial of beta-agonists in asthma it is generally believed that a wash-out period of a few days is more than long enough to eliminate all effects of previous treatment. In a multi-dose trial we might use a different approach. Patients might be given repeated doses of one therapy during a month after which they might be switched over to an alternative therapy for another month. As a precaution against carry-over we might limit the observation period to the second two weeks of each treatment period. Obviously this only makes sense if we wish to observe the steady-state behaviour of each treatment and believe that this will be reached after two weeks at the latest under each treatment regardless of what has happened before. Note, however, that a similar assumption would have to be made in a parallel group trial with the same objective.

The main difficulty with the wash-out approach to dealing with carry-over is that *we can never be certain that it has worked*. This would be a serious objection under one or both of two conditions. First, suppose it were the case in general (except for cross-over trials) that we could say that using the results from clinical trials did not require us to make assumptions we could not 'verify'. This is not, however, the case. All analyses of clinical trials depend on dozens of assumptions we choose to ignore because we are unable or unwilling to investigate them. For example we assume that trials carried out in patients who give

consent yield results which are applicable to patients in general, including those who would refuse to enter a trial. In a modern democracy there is no way that this assumption could ever be examined using clinical trials and it might even plausibly be maintained that for certain mental diseases it is unlikely to be true. Second, it would be a serious objection if there were a realistic alternative. However, there is not. Even the most enthusiastic proponents of the modelling approach to carry-over concede that one has to assume that if carry-over occurs it has taken a particular form. Such an assumption is not only not verifiable but *a priori* unlikely to be true.

A further approach to the problem of carry-over is to recognize in general that the adequacy of assumptions made in clinical trials is tested by carrying out many studies with different designs. The isolated study in which all assumptions must be 'verified' (whatever that might mean) in order that the conclusion, which is then to stand for all scientific posterity, can be known to be true (or have reasonably been declared to be true using some probabilistic rule) is an unrealistic paradigm of research. Cross-over trials are carried out where diseases are not life-threatening. There is no reason why trials should not be repeated; there is every advantage in so doing in trying radically different designs. It is in this way that scientific knowledge is increased. As different trials with different designs come up with similar results it becomes increasingly difficult to maintain that some peculiar form of carry-over could be the common explanation.

Thus, in this book I shall be making the assumption that the practitioner will deal with carry-over as follows. First he will design his trials cautiously using what he believes to the best of his knowledge to be adequate wash-out periods (whether passive or active). Second he will accept that his findings will always be conditional on an assumption (amongst many!) that carry-over has not seriously distorted his results and that there is always the possibility that different trials with different designs may not repeat them.

1.9   OTHER ATTITUDES TO BE ADOPTED

The treatment of statistics in this book will be eclectic but largely restricted to frequentist methods. (A possible Bayesian analysis is included in Chapter 3 but this is the only such example.) That is to say, a variety of frequentist approaches which I personally consider to be useful will be employed. This does not imply any hostility on my part to the *Bayesian* programme (in fact I am only too happy to acknowledge the important contribution which Bayesians like Peter Freeman and Andy Grieve have made in correcting certain frequentist errors) but merely reflects a recognition that for the time being practicalities dictate that the majority of analyses of clinical trials in general and cross-over trials in particular will be frequentist. (Although it has been predicted that by the year 2020 things will be different!)

Heuristic arguments will be employed in developing analyses. There will be no formalism and little algebra.

Global tests of significance will not be covered. In my experience it is rarely of interest for the trialist to test the null hypothesis that all treatments are equal (where there are more than two treatments). Instead the testing and estimation of specific treatment contrasts with calculation of associated significance levels and confidence intervals will be covered.

A rough agreement between *modelling* and *randomization* will be maintained. (Although this is not always a simple or obvious matter in cross-over trials, it is very easy to put a foot wrong and I can give no guarantees that I will not do so.) For example for the *AB/BA* cross-over if patients are allocated completely at random to the two sequences I should regard it as being permissible to ignore any period effect in the model used for analysis: not so much because of any randomization argument *per se* but because this form of allocation is consistent with a belief that the period effect is negligible. In this case, however, I should also permit the fitting of a period effect because this form of allocation is also consistent with a belief that the period effect is important if it is known that it will be dealt with by analysis. On the other hand if the investigator had blocked the trial so as to balance the sequences and ensure that an equal number of patients were assigned to each, I would regard him as being bound to fit a period effect because his behaviour shows that he considers such effects to be important.

I shall extend the ban on pre-testing for carry-over to apply to all other forms of pre-testing as well. There will be no dropping of terms from models because they are not significant. Similarly, choices will not be made between parametric and non-parametric methods on the basis of tests of normality. Quite apart from any other reason for not performing such tests it is only the within patient errors anyway which need to be normally distributed for normal theory tests to be valid for most cross-over analyses. The 'correct' examination of this point requires the investigator to go so far down the road of fitting the parametric model that it is a charade for him to pretend he has not done so. The trialist should either determine on *a priori* grounds which form of analysis he favours or perform (and of course report) both. On the whole I have not found a great use for non-parametric methods in cross-over trials but I regard them, nevertheless, as useful on occasion and therefore worth covering.

Suitable approaches to analysis will be illustrated using the computer packages SAS® (version 8.02), GenStat® (fifth edition, release 4.2) and S-Plus® (version 6.0) as well as, on occasion, StatXact® (version 4.0.1) and the spreadsheet Excel 97®. However, this book does not provide a course in any of these packages. Books that I myself have found helpful in this respect are Cody and Smith (1997) and Der and Everitt (2002) for SAS®, Harding *et al.* (2000) and McConway *et al.* (1999) for GenStat®, Krause and Olson (2000) and Venables and Ripley (1999) for S-Plus and Berk and Carey (2000) for Excel. StatXact is provided with an extremely scholarly and remarkably readable manual (Mehta

and Patel, 2000). (Other more specialist texts illustrating particular aspects of analysis with these packages are referred to subsequently.) Where possible, the analyses covered will also be illustrated using calculations done on a pocket calculator.

Finally, adjustments for repeated testing will not be covered. It will be assumed that the investigator will make a sensible choice of measures, think hard about what hypotheses are worth investigating, report the results of all analyses he makes (however disappointing) and do his best to make a cautious and sensible overview of his findings as a totality.

1.10   WHERE ELSE CAN ONE FIND OUT ABOUT CROSS-OVER TRIALS?

There is a book by Jones and Kenward (1989), with a rather more theoretical treatment than this one, as well as another by Ratkowsky *et al.* (1993). A web-based tutorial by the author (Senn, 2001a) provides an introduction to the subject. There are also encyclopaedia articles by Kenward and Jones (1998) and Senn (1998a, 2000b) which provide overviews of the field.

2

*Some Basic Considerations Concerning Estimation in Clinical Trials**

2.1   THE PURPOSE OF THIS CHAPTER

The purpose of this chapter is to review various basic statistical concepts regarding clinical trials which will either be referred to subsequently or assumed as general background knowledge. The chapter may be omitted altogether by readers who are already familiar with the general statistical approach to clinical trials or, alternatively, who are happy to learn how to analyse cross-over trials without concerning themselves over much about justifications for methods employed. It may also be passed over and reserved for later perusal (or simply referred to as necessary) by readers who are concerned to proceed directly to the study of cross-over trials. In particular I am anxious to make the point that the reader who dislikes algebra should not allow this chapter to put him off the rest of the book. It most definitely is *not* required reading.

2.2   ASSUMED BACKGROUND KNOWLEDGE

The reader is assumed to be familiar with various basic statistical ideas, to be able to calculate and understand standard descriptive statistics such as means, medians and standard deviations, to have acquired some elementary knowledge of statistical estimation and hypothesis testing, to be at ease with the concepts of random variables, estimators and standard errors, confidence intervals, significance levels, 'P values', etc., and to have encountered and used t tests, F tests and analysis of variance. On the computational side some familiarity with studying computer output from statistical packages, in particular in connection with linear regression, will be assumed.

The book will not make a heavy reliance on such background knowledge. As has already been explained in the Preface and the Introduction, the emphasis is

on heuristic justification and teaching through worked examples. The reader's memory will be jogged regarding background theory where this is appropriate. Equally well, however, this book does *not* include a general course in medical statistics. Further help in that direction will be found by consulting Armitage and Berry (1987), Altman (1991), Campbell and Machin (1990) or Fleiss (1986a). For the background knowledge regarding clinical trials Pocock (1983) is extremely useful.

Frequent reference (often implicitly rather than explicitly) will be made to the following basic statistical ideas.

2.2.1   Linear combinations

We frequently form estimators of treatment effects by weighted addition of individual observations. Such sums are referred to as *linear combinations*. We now state three rules regarding such combinations.

(2.1) If X_i is a random variable with expected value $E[X_i] = \mu_i$ and variance $\text{var}[X_i] = \sigma_i^2$ and a and b are two constants, then the expected value $E[a + bX_i]$ of $a + bX_i$ is $a + b\mu_i$ and its variance, $\text{var}[a + bX_i]$, is $b^2 \sigma_i^2$.

Practical example   Suppose X_i is a random variable standing for temperature measurements in Celsius of a population of patients. Suppose its expected value is 37°C and its variance is $0.25°C^2$. Then, since the corresponding Fahrenheit readings are obtained by multiplying by 9/5 and adding 32, we may substitute 32 for a and 9/5 for b, from which the expected value in Farenheit is 98.6°F and the variance is $0.81°F^2$.

(2.2) If X_i and X_j are two random variables, then $E[aX_i + bX_j] = a\mu_i + b\mu_j$ and $\text{var}[aX_i + bX_j] = a^2 \sigma_i^2 + b^2 \sigma_j^2 + 2ab\sigma_{ij}$, where $\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$ is known as the *covariance* of X_i and X_j. If X_i and X_j are independent then $\sigma_{ij} = 0$.

Practical example   Suppose that in a parallel group trial X_0 is a baseline measurement in forced expiratory volume in one second (FEV_1) and X_1 is the same measurement carried out at the end of the trial. We form a new variable, the difference from baseline, by substracting X_0 from X_1. This corresponds to substituting 0 for i, 1 for j, -1 for a and 1 for b in (2.2). Hence the expected value of this difference is $\mu_1 - \mu_0$ and the variance is $\sigma_0^2 + \sigma_1^2 - 2\sigma_{01}$.

Remark   Difference from baseline is a measure which is commonly encountered in clinical trials. If we assume that baseline and outcome variances are the same, so that $\sigma_0^2 = \sigma_1^2 = \sigma^2$, then it will be seen that providing the covariance σ_{01} is greater than half the variance, σ^2, the difference from baseline measure, $X_1 - X_0$, has lower variance than the raw outcomes measure, X_1.