University of Southampton

Faculty of Mathematical Studies

Mathematics

The Design of Cross-over Studies Subject to Dropout

by

Janice Lorraine Low

Thesis submitted for the degree of Doctor of Philosophy

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MATHEMATICAL STUDIES
MATHEMATICS

Doctor of Philosophy

THE DESIGN OF CROSS-OVER STUDIES SUBJECT TO DROPOUT

by Janice Lorraine Low

A cross-over study is a comparitive experiment in which subjects receive a sequence of two or more treatments, one in each of a series of successive time periods, and the response of each subject is measured at the end of every period. A common problem, particularly in medicine, is that subjects fail to complete a study through dropping out during the later stages of the trial for reasons unrelated to the treatments received. Current practice is to select a design for a study on the basis of its performance under the assumption that no subjects drop out, using a criterion such as A-optimality. This is an unrealistic assumption for many medical applications. This thesis investigates how studies should be designed when it is unrealistic to assume that subjects will not drop out.

A method of assessing cross-over designs is presented which judges how accurately all the pairwise treatment comparisons are estimated under the assumption that each subject has a fixed probability of dropping out during the final period, independent of treatment received and the other subjects. The method of design assessment is computationally intensive even for studies involving a relatively small number of subjects. Ways of reducing the amount of computation required are presented through establishing the link between implemented designs and a colouring problem in combinatorial theory. The reductions achieved make feasible investigations of currently used designs for cross-over studies.

The results of investigations are presented for designs for the cases of particular practical importance, namely four treatment, four period and three treatment, three period studies, in which a simple carry-over model is assumed for the observations. Designs which are more robust to final period dropout than the currently favoured designs are identified.

Contents

List of Tables

List of Figures

Acknowledgements

This research was carried out under a research studentship from the Scientific and Engineering Research Council.

I am grateful to my supervisors Dr. S. M. Lewis and Dr. P. Prescott for their advice, guidance and support through the many highs and lows of this research.

My thanks also go to Dr. K. G. Russell for his advice and encouragement during his visits to Southampton, Dr. B. D. Mckay for his assistance in developing a computer program and Denise Williams for typing part of the manuscript.

I wish to thank my parents for their constant and unconditional love which has been a continual source of comfort and inspiration.

I am indebted to my husband Stephen for his endless patience and devotion and for being the best husband a research student could ever wish for.

Finally, I thank the Lord for his continuing grace.

Chapter 1

Introduction

1.1 Cross-over Trials

Two popular study designs employed in clinical and medical research are the parallel group study and the cross-over trial. In a parallel group study each subject is assigned at random to a group which receives one treatment for the duration of the study. Treatment differences are then estimated by making comparisons between the subject groups. In contrast, in a cross-over trial each subject is randomly assigned to a sequence of treatments. One treatment is given in each of a series of successive time periods and the response of each subject is measured at the end of each period. Cross-over trials are widely used in many areas including clinical and medical research, agriculture and human factors engineering.

In this thesis we consider the design of cross-over trials for clinical and medical research involving human subjects, either patients or healthy volunteers, to investigate the effects of different drug or therapeutic treatments. Since each subject provides a direct comparison of the treatments he/she has received, it is possible for comparisons to be made within subjects rather than between subjects as in a parallel group study. In clinical research involving human subjects, the variation between subjects usually greatly exceeds the within-subject variation. Therefore, with a cross-over design it is possible to estimate the important contrasts of interest

with greater efficiency since, to obtain estimates of equal precision to those obtained from a parallel group study, fewer subjects and observations are required. This leads to a considerable saving of resources.

Cross-over trials should only be used in situations where the treatments being applied are not expected to have a permanent effect upon the subjects. They are unsuitable for studying a condition in which subjects may experience a considerable improvement or deterioration in their condition, regardless of treatment received, during the course of the trial. Thus, in a medical context, cross-over trials are best suited to investigating treatments for chronic conditions in which the aim is to alleviate symptoms, rather than permanently improve the condition. For example, trials concerned with comparing the relative efficacy of different treatments for conditions such as asthma, diabetes, angina, epilepsy, hypertension or arthritis are often cross-over experiments.

The order in which each subject receives the treatments under investigation is determined by the particular design chosen for the study. Designs used in cross-over trials are referred to as **cross-over, change-over** or **repeated measurement** designs. The latter term, however, is used more widely to include longitudinal studies in which sequences of observations are made on subjects who receive repeated applications of the same treatment.

An example of a cross-over design to compare the effects of four treatments over four treatment periods using four different treatment sequences is given in Example 1.1. In the example, and throughout this thesis, designs will be shown as two-way layouts with rows corresponding to treatment sequences and columns corresponding to periods.

Example 1.1 A trial to compare the effects of four treatments, labelled A, B, C and D, using four periods $(1, \ldots, 4)$ and 16 subjects might employ the four treatment sequences $(1, \ldots, 4)$ given in Table 1.1, and allocate four subjects to each sequence. For example, a subject allocated to sequence 1 receives treatment A in period 1,

treatment B in period 2, treatment D in period 3 and treatment C in period 4.

Table 1.1: Cross-over design for Example 1.1

		Period			
		1	2	3	4
	1	A	B	D	C
Treatment	2	B	C	A	D
sequence	3	C	D	B	A
	4	D	A	C	B

The layout of treatment symbols in the example forms a Latin square in that every symbol appears once in each row and once in each column. It has the further property that every symbol precedes every other symbol the same number of times. Latin squares having this property were used by Williams (1949) to give treatment sequences for designs for agricultural experiments and hence are known as Williams squares. Williams developed the designs to address the particular problem of treatments persisting or carrying over beyond the period in which they are applied and the designs have since been widely used in cross-over studies. The early work in this area assumed a simple model for the observations from the studies. More recently, more complicated models have been proposed and these, together with the simple model, are reviewed in the following section. In Section 1.3 the controversies surrounding cross-over studies are discussed. The problem of how to design studies when subjects may drop out is described in Section 1.4 where the aims of the thesis are stated. In Section 1.5 the estimation under a simple model of direct and carry-over treatments effects is summarised for reference later in the thesis. Finally, Section 1.6 gives definitions of design selection criteria used in later chapters.

1.2 Models used in Cross-over Studies

In planning a cross-over experiment, it is often important to allow for the possibility that the contribution to a subject's response from a particular treatment may persist beyond the period in which the treatment is administered. This is because, in many situations, it is unrealistic to assume that a treatment ceases to have an effect immediately it is stopped. Several different models have been proposed in the literature for the observations from cross-over studies. In recent years controversies have arisen regarding the most appropriate models to adopt at the planning stage of an experiment. In this section, a variety of models is reviewed for cross-over trials to compare the effects of $t \geq 2$ treatments over $p \geq 2$ periods using a total of mn subjects, where n subjects are allocated to each of m distinct treatment sequences.

1.2.1 The Simple Carry-over Model

The model most frequently assumed for the observation obtained from the jth subject allocated to the ith sequence in the kth period is the additive model:

$$\begin{aligned} y_{ijk} &= \mu + s_{ij} + \pi_k + \alpha_{d(i,j,k)} + \lambda_{d(i,j,k-1)} + \varepsilon_{ijk} \\ &\quad (i = 1, \ldots, m;\ j = 1, \ldots, n;\ k = 1, \ldots, p), \end{aligned} \tag{1.1}$$

where μ is the overall mean, s_{ij} is the effect of the jth subject receiving the ith sequence, π_k is the kth period effect, $\alpha_{d(i,j,k)}$ is the direct effect of the treatment $d(i,j,k)$ administered to the jth subject receiving the ith sequence in the kth period, $\lambda_{d(i,j,k-1)}$ $(k = 2, \ldots, p)$ is the first-order carry-over effect of the treatment administered in the $(k-1)$th period to the jth subject receiving the ith sequence, $\lambda_{d(i,j,0)} = 0$ and the ε_{ijk}'s denote random errors which are assumed to be independently and identically distributed with zero mean and variance σ^2. All effects in the model are assumed to be constants, that is the effects are fixed.

Model (1.1) expresses the carry-over from a treatment in its simplest form and

for this reason is often referred to as the **simple carry-over model**. The model is widely used in the literature and has been adopted by many authors in their search for optimal or efficient cross-over designs for many experimental situations, as discussed in Chapters 4 and 5. In recent years the simple carry-over model has received some criticism in the medical literature mainly because some of the assumptions required by the model are not always satisfied in the context of pharmacological measurements. In Section 1.3 we review the criticisms of the model together with additional controversies concerning cross-over experiments.

1.2.2 Inclusion of Interactions

It is possible to allow for interactions between the factors in the experiment by extending model (1.1) to include interaction terms. In this section we describe some of the more important interactions which may be included and outline situations in which they may be needed in the model.

(i) Direct treatment $\times$ period interaction. Such an interaction occurs when the effect of a treatment is modified according to the period in which it is administered. For example, consider a two treatment, two period cross-over experiment to compare the efficacy of two different asthma drugs in patients who also suffer from hay-fever. If period one occurs during the hay-fever season, but period two does not take place until after the season is over, then a significant period effect may be observed. Suppose that one of the drugs is an effective treatment for asthma in the absence of hay-fever, but is ineffective when a subject is suffering from hay-fever, but the action of the other drug is unaffected by hay-fever, then a treatment $\times$ period interaction may arise.

If evidence of a treatment $\times$ period interaction is present, it can be very difficult to interpret the results and draw conclusions about the efficacy of the treatments in the absence of a period effect.

Several authors have included a treatment $\times$ period interaction term in their

models when seeking efficient cross-over designs. For example, Baalam (1968) uses a model with a treatment × period interaction, but no carry-over effects, to find efficient designs for t treatments, t^2 experimental units and two treatment periods. Lasarre (1991) finds efficient two treatment designs using a similar model but incorporating random, rather than fixed, subject effects.

(ii) Direct treatment × subject interaction. Senn (1993) argues that, in certain medical situations, a general treatment effect may not exist, but the effect of a treatment may vary from patient to patient. It is then necessary to include a direct treatment × subject interaction term in the model. In these circumstances, it will be very difficult to draw conclusions about the general efficacy of the treatment. In particular, it will be very difficult to predict in advance how effective the treatment will be for an individual patient.

The problem of interpretation may be one reason why very little discussion of the direct treatment × subject interaction appears in the literature. Alternatively, the absence of models including the interaction could be because the existence of a significant interaction will only increase the amount of variation in the trial, rather than render any analysis invalid, as discussed by Cox (1984).

(iii) Subject × period interaction. Another interaction term which may be included in the model is a subject × period interaction. Such an interaction may be present if time trend effects occur which are not the same for all subjects. For instance, in some studies patients are not always recruited together but over a considerable length of time. Consequently, environmental conditions may vary for patients recruited at different times. In the asthma study described previously, suppose that patients recruited to the study at an early stage have their first period before the hay-fever season and their second period during it, whilst patients recruited at a later stage have their first period during the hay-fever season and their second after it has finished. Then a significant subject × period interaction may be observed.

(iv) Direct treatment × first-order carry-over interaction. A potentially important term which could be included in the model is a direct treatment × first-order carry-over interaction which is needed if the carry-over effect of the treatment administered during period $(k-1)$ to the jth subject on the ith sequence varies according to the treatment administered to the subject during period k. Fleiss (1986) and Matthews (1993) draw attention to the possibility that the amount of carry-over a particular treatment exerts onto itself may differ from the amount of carry-over it exerts onto a different treatment. More discussion of this particular issue appears in Section 1.3.

A consequence of adding any of the interaction terms discussed in this section to model (1.1) is that the number of parameters to be estimated will be increased. The resultant increase in the number of subjects required, together with the difficulties of interpreting the results and of knowing in advance which particular interactions need to be included, may be why few authors include interaction terms in models for designing cross-over experiments. An exception is the algorithmic approach of Jones and Donev (1994).

1.2.3 Higher Order Carry-over Effects

In the simple carry-over model it is assumed that the carry-over effect of a treatment will persist for no more than one period. However, in cross-over studies involving $p > 2$ treatment periods there is the possibility that carry-over effects may persist for longer. A carry-over treatment effect which lasts up to and including the rth period after the treatment has ceased is known as an *rth-order* carry-over effect. Model (1.1) can be extended to include all possible higher order carry-over effects as follows:

$$\begin{aligned} y_{ijk} &= \mu + s_{ij} + \pi_k + \alpha_{d(i,j,k)} + \lambda^{(1)}_{d(i,j,k-1)} + \lambda^{(2)}_{d(i,j,k-2)} + \ldots + \lambda^{(r)}_{d(i,j,k-r)} + \varepsilon_{ijk} \\ &\quad (i = 1, \ldots, m;\ j = 1, \ldots, n;\ k = 1, \ldots, p), \end{aligned}$$

where $\lambda^{(r)}_{d(i,j,k-r)}$ is the carry-over effect, observed in period k of the treatment administered in the $(k-r)$th period to the jth subject on the ith sequence, such that $\lambda^{(h)}_{d(i,j,k-h)} = 0$ for $k \leq h$ where $h = 1, \ldots, r$. All other terms are as for model (1.1).

Senn (1992) questions whether it is appropriate to exclude higher order carry-over effects from a model when first-order effects are included. Decisions on the anticipated duration of carry-over effects should be made on the basis of knowledge of the nature of the treatments to be administered.

Williams (1949) considered the design and analysis of experiments when, not only first and second-order carry-over effects may exist, but also their interactions cannot be assumed negligible. The design and analysis of experiments involving higher order carry-over effects is also considered by Patterson and Lucas (1962).

1.2.4 Models Including Correlated Errors

In the simple carry-over model the random errors are assumed to be independently and identically distributed. This may be an unrealistic assumption since several measurements are made on the same subject and hence may be correlated. One way of extending the model to allow for this possibility is to assume some form of serially correlated error structure. If we assume that the measurements on different subjects are independent and that measurements taken on the same subject are correlated, then we can replace the independence assumption of model (1.1) by

$$Var(\varepsilon) = (I_{mn} \otimes V_p)\sigma^2,$$

where $\varepsilon = (\varepsilon_{111}, \ldots, \varepsilon_{11p}, \ldots, \varepsilon_{1n1}, \ldots, \varepsilon_{1np}, \ldots, \varepsilon_{mn1}, \ldots, \varepsilon_{mnp})$ is an $mnp \times 1$ vector containing the random errors for each observation, I_{mn} is the $mn \times mn$ identity matrix, V_p is the $p \times p$ variance-covariance matrix for the observations on each subject and $\otimes$ denotes the Kronecker product.

The majority of work in the literature on finding designs under the assumption of correlated errors has assumed that V_p has the form for a first-order autoregressive process, see for example Kunert (1985, 1991) and Matthews (1987), defined as

follows.

Stationary first-order autoregressive process. If we assume the errors for the jth subject ($j = 1, \ldots, n$) receiving the ith sequence ($i = 1, \ldots, m$) follow a stationary first-order autoregressive process we can define

$$\varepsilon_{ijk} = \begin{cases} \eta_{ijk} & \text{for } k = 1 \\ \rho\varepsilon_{ij(k-1)} + \eta_{ijk} & \text{for } k > 1 \end{cases}$$

where η_{ijk} are independent, identically distributed random variables with zero mean and variance σ_e^2 and ρ is the autoregression parameter.
This results in a variance-covariance matrix for the within-subject errors, $V_p = (v_{ab})$, of order p, whose elements are given by

$$v_{ab} = \frac{\rho^{|a-b|}}{1-\rho^2}\sigma_e^2 \quad (a, b = 1, \ldots, p),$$

where $-1 < \rho < +1$ and $|a - b|$ denotes the absolute difference of a and b.

An error structure of this form, in which the correlation gradually decreases over time, may provide a reasonable approximation to reality for some experiments. However, the popularity of such an error structure in the literature may be due to its mathematical tractability rather than evidence of its plausibility. A disadvantage of models including autoregressive errors is that the autoregression parameter ρ is assumed to be known. Kunert (1985) considers such models, assuming there are no carry-over effects, and defends their use by arguing that it is often possible to obtain information about ρ from previous similar studies.

In practice, however, previous studies can only provide an approximation to the true value of ρ. This is an important consideration which may effect design selection since, as Kunert demonstrates, an optimal design for a particular size of study, in the sense of minimising the maximum variance of the treatment effects, will usually depend upon the value of the autoregression parameter ρ. Methods for constructing the optimal designs proposed by Kunert have been given by Street (1989).

Several authors have sought optimal or efficient two treatment designs under model (1.1) with the assumption of independent errors replaced by an autoregressive error structure. Matthews (1987) gives optimal designs, in the sense of minimising the variances of $\hat{\alpha}$ and $\hat{\lambda}$, for experiments involving three and four periods which require the allocation of unequal subject numbers to the treatment sequences. He also gives sub-optimal designs which have equal numbers of subjects per sequence as he considers that these are preferred by experimenters. Kunert (1991) gives efficient designs for $p \geq 3$.

An alternative approach to modelling the correlation between the within-subject errors is to use the following first-order moving average process.

Stationary first-order moving average process. The errors for the jth subject $(j = 1, \ldots, n)$ receiving the ith sequence $(i = 1, \ldots, m)$ have the form

$$\varepsilon_{ijk} = \begin{cases} \eta_{ijk} & \text{for } k = 1 \\ \eta_{ijk} - \rho\eta_{ij(k-1)} & \text{for } k > 1 \end{cases}$$

where η_{ijk} are independent, identically distributed random variables with zero mean and variance σ_e^2 and ρ is the moving average parameter. The variance-covariance matrix, V_p, for the within-subject errors then has elements of the form

$$v_{ab} = \begin{cases} \rho^{|a-b|}\sigma_e^2 & \text{if } |a-b| \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $-1/2 \leq \rho \leq 1/2$.

Designs which are known to be optimal or efficient using models containing certain correlated error structures have practical application only to experiments in which the error structure is known at the planning stage. In many fields of investigation, including medical trials, it is not usually possible to predict in advance the nature of the correlation. In these situations, a design which can be shown to provide efficient estimates of the contrasts of interest under models using several different, plausible error structures is desirable, that is, a design which is **robust**

to mis-specification of the error structure. If the assumed error structure is subsequently found to have been inappropriate, then the design will have been selected on the basis of incorrect values for the variances of the estimated treatment effects. This may have serious consequences, unless the design is known to be robust to assumptions on the errors. Matthews (1990) examined this issue, but only for the designs in Matthews (1987) and the two correlated error structures described in this section.

In practice, most experimenters adopt a design which is efficient when the errors are assumed to be independent because they cannot predict the error structure in advance.

1.3 Controversies Concerning Cross-over Trials

In this section some of the issues and controversies concerning cross-over trials are briefly reviewed. The issues arise from consideration of the existence, nature and duration of carry-over effects.

1.3.1 The Need for Carry-over Effects

The size of physical carry-over effects can sometimes be reduced by the use of wash-out periods, that is time intervals inserted between the treatment periods during which the subjects receive no active treatment. The aim of the wash-out period is to allow the contribution of the treatment administered in the previous period to lessen to such an extent that its effect may be assumed to be negligible. Unfortunately, in many practical applications, there is no guarantee that wash-out periods will achieve this aim and, as pointed out by Matthews (1993), for certain therapeutic studies there may be ethical objections to their use. For pharmacological applications, Senn (1993) argues that when the time for which the drugs under investigation remain in the body is known, it will be possible to make the wash-out period long enough for any carry-over effects to be eliminated. However, there is no guarantee that

carry-over effects of a non-physical nature will not exist.

In medicine it is widely believed that the act of being treated can have a profound effect upon some patients which cannot always be accounted for by the physical action of the treatment(s) being administered. Such psychological effects can be observed in clinical trials in which placebo treatments or sham procedures are used. A placebo is an inactive substance which is manufactured so that it is identical in appearance and taste to that of the active treatment under investigation. A sham procedure is a dummy procedure in which the patient undergoes the same regime as those patients receiving an experimental procedure.

Trials of this nature are usually double-blind, that is, neither the patient nor the clinician is aware which treatment is being administered. There are numerous examples in the literature of trials which report a significant placebo effect. Many involve conditions with a psychosomatic cause such as anxiety, but others involve conditions such as epilepsy and post-operative pain. For example, Group (1989) describes a study to investigate the effectiveness of cinromide in reducing the frequency of seizures in epileptics in which 23% of the patients experienced a 50% reduction in seizure frequency whilst receiving a placebo.

There is a great deal of evidence in the medical literature to support the existence and significance of psychological treatment effects. Therefore, the possibility that such effects may persist into subsequent treatment periods cannot be ignored. Baker et al (1982) state that psychological carry-over can occur if the effect of a second treatment is partly dependent upon a subject's attitude following the first treatment. For example, consider a pain relief study to compare an active treatment with a placebo. The subjects who receive the active treatment during the first period may experience a reduction in pain and enter the second period with confidence. Alternatively, the patients who receive the placebo during the first period may receive little or no pain relief and thus enter the second period with apprehension or possibly withdraw from the study completely. Jones and Kenward (1989) and Wilan and Pater (1986) give similar discussions of psychological carry-over effects.

1.3.2 Criticisms of the Simple Carry-over Model

In recent years a number of authors have expressed some concern regarding the use of the simple carry-over model (1.1) for certain clinical cross-over studies, for example Fleiss (1986, 1989), Matthews (1993) and Senn (1992, 1993). The most comprehensive discussions of the disadvantages of the model are given by Senn who argues that the model is not only of little use but also harmful. His main criticism is that the presence of a first-order carry-over effect in the model gives the misleading impression that it is no longer necessary to use effective wash-out periods. He further argues that, in order to justify the use of the simple carry-over model, an experimenter must either know or make assumptions about the persistence of carry-over effects. If the decision to use the model is based on actual knowledge, then it must be possible to design the study in such a way that carry-over effects are eliminated. Thus the need to include a carry-over term in the model is removed. Alternatively, if its use is justified by assumptions, then any conclusions drawn will depend upon the validity of the assumptions. Furthermore, if it is reasonable to assume that the effect of a particular treatment will only persist for one period beyond the period of application, why is it unrealistic to assume that there are no carry-over effects, or that they persist for more than one period? Essentially, Senn questions whether *any* assumptions concerning the duration of carry-over are reasonable. Fleiss (1989) argues that the assumption that carry-over effects only persist for one period is made for mathematical convenience rather than because it is believed that the model is an accurate description of the response.

Nevertheless, much of the literature is in agreement that important carry-over effects may exist and that often carry-over terms higher than first-order are negligible. Clearly, anybody who employs the simple carry-over model does not believe it to be "true". However, they should satisfy themselves that it provides a reasonable approximation to reality through scientific judgement.

A further criticism of the simple carry-over model arises when a design involves

treatment sequences in which a treatment follows itself. Fleiss (1989) and Matthews (1993) question the assumption in the model that the carry-over of a treatment onto itself is identical to its carry-over onto any other treatment. In drug trials this may be an unrealistic assumption to make, since the duration of periods is usually chosen to allow the effect of each treatment to reach its maximum. Consequently, if any treatment is followed by itself, the result may be that there is no carry-over present when the second measurement is made. Matthews (1993) proposes an alternative model for two treatment cross-over studies in which the carry-over effect of a treatment onto itself is set to zero. This can easily be generalised to provide a model for designs involving any number of treatments as follows:

$$\begin{aligned} y_{ijk} &= \mu + s_{ij} + \pi_k + \alpha_{d(i,j,k)} + \lambda_{d(i,j,k-1)}\{1-\phi\} + \varepsilon_{ijk} \\ &\quad (i = 1, \ldots, m;\ j = 1, \ldots, n;\ k = 1, \ldots, p), \end{aligned}$$

where

$$\phi = \begin{cases} 1 & \text{if } d(i,j,k-1) = d(i,j,k) \\ 0 & \text{otherwise} \end{cases}$$

and $d(i,j,k)$ is defined in (1.1).

Another controversy associated with the simple carry-over model is whether or not estimates based on it are efficient. Senn (1993) states that he believes estimates obtained using this model will be inefficient and illustrates his argument via an example. In the example he shows that the estimates of the direct treatment effects, adjusted for carry-over effects, have a larger variance and worse bias than the unadjusted estimates. Senn's opinion is valid for the particular example given, in which he has made certain pharmacodynamic assumptions about the nature of the response obtained. Whether his opinion is valid in general is debatable. Abeyasekera and Curnow (1984) argue that it is always preferable to adjust for carry-over effects since, if very small carry-over effects exist but are ignored, then this can lead to a bias in the estimates of the direct treatment effects. This point of view is in direct contrast to that of Senn.

When designing experiments it is important to include in the model those terms which may account for substantial amounts of variation. The model selected for this purpose is a tool used to guide the experimenter to the most appropriate design to use. Different models will lead to different choices of design and this can have a profound effect on the design proposed for the study. Therefore, it is important that the model used during the planning stage is appropriate. We do not recommend the indiscriminate use of the simple carry-over model since there are circumstances in which it is not realistic. However, many experimental situations do exist, particularly in medicine, in which it is believed to be appropriate to consider the possibility of carry-over effects at the planning stage and to employ the simple carry-over model when planning the study.

1.4 The Problem of Subject Dropout and the Aims of the Thesis

Frequently in cross-over trials, particularly in medicine, subjects fail to complete a study, most commonly dropping out during the last one or two periods. It is generally acknowledged that dropouts are a frequently occurring problem in cross-over studies. Matthews (1987) states that dropouts are inevitable in clinical cross-over trials and Gough (1994) says that dropouts are a major concern in the design and analysis of clinical trials. Despite this fact, there are few examples in the literature of cross-over studies in which dropouts have occurred. One reason for this may be that such studies are not reported because no firm conclusions can be drawn due to the number of missing observations and their distribution across the treatment sequences, as in the following example.

Example 1.2 In Example 1.1 suppose that the simple carry-over model (1.1) is assumed for the observations. When the experiment is performed there are very many possible outcomes which may result. Two of these are shown in Table 1.2, where

the numbers shown in brackets denote the number of subjects present throughout the period and the absence of the number indicates that all subjects are present throughout the period.

Table 1.2: Two possible outcomes for Example 1.2.

Outcome 1

		Period			
		1	2	3	4
	1	A	B	D(4)	C(4)
Treatment	2	B	C	A(4)	D(3)
sequence	3	C	D	B(4)	A(3)
	4	D	A	C(4)	B(2)

Outcome 2

		Period			
		1	2	3	4
	1	A	B	D(4)	C(3)
Treatment	2	B	C	A(3)	D(0)
sequence	3	C	D	B(2)	A(0)
	4	D	A	C(3)	C(2)

In outcome 1, four dropouts occur during the final period, namely one from each of sequences 2 and 3 and two from sequence 4. If this outcome were realised from the experiment, then the resultant increase in the total variance of the least squares estimators of the direct and first-order carry-over treatment comparisons under model (1.1), compared with that of the design in Example 1.1 is small, being only 8.68% and 8.38% respectively. In contrast, the objective of the experiment could not be achieved if outcome 2 were realised, that is four subjects lost in the third period and a further seven subjects in the final period. The resulting design is disconnected with respect to the estimation of both the direct and first-order carry-over effects, that is under model (1.1) some of the pairwise direct and first-order carry-over treatment comparisons cannot be estimated.

When a trial is being planned, it is sometimes known that subjects may drop out in the later stages. There are often similar trials which have been undertaken earlier from which an estimate of the probability of a subject dropping out during a particular period can be made. Current practice fails to use this information in planning a trial. A design is chosen under the assumption that subjects will not

drop out. As Example 1.2 illustrates, this approach may have severe consequences for estimating the treatment comparisons of interest.

The reasons a subject withdraws from a study prematurely are not necessarily connected to the treatments under investigation. Diggle and Kenward (1994) list three types of dropout processes: completely random dropout, in which the probability of a subject dropping out is completely unrelated to the treatments administered; random dropout, that is the probability of dropout is related to the treatment(s) which precede the subject leaving the study; informative dropout, where the probability of dropout is related to the treatment being administered at the time of dropout. There can be serious ethical objections to using a particular treatment in a trial if it is believed at the outset that this treatment may lead to subjects dropping out, as in random or informative dropout. In this thesis we shall consider only completely random dropout. The methodology can be extended to consider situations involving random or informative dropout.

The aim of this thesis is to present a method for investigating the robustness of cross-over designs to dropouts in order that this information may be used in selecting a design. The method presented is not model-dependent and can be used in conjunction with any of the models described in Section 1.2. It is presented for experiments having $p \geq 3$ in which it is anticipated that subjects may drop out only in the final period. If dropout is an issue in a two period study then the same approach could be used. However, the use of a parallel group study is a more realistic proposition. In this thesis the method is developed under the assumption that the observations follow model (1.1). Designs are obtained for $t = p = 3$ and $t = p = 4$ which are more robust to dropouts than those currently favoured. The approach can be applied to designs having larger numbers of periods. However, when dropout is an issue, it is usually not sensible to plan studies involving more than four periods.

In Chapter 2 the method of assessing the robustness of cross-over designs is given. The implementation of this methodology is a daunting task, since the number of designs to be evaluated is considerable even for relatively small studies. In Chapter

3 we show how the size of the computational problem can be reduced by using results from combinatorial theory. In Chapters 4 and 5 we investigate designs in which $t = p = 4$ and $t = p = 3$ respectively. Chapter 6 outlines how the method can be extended to situations in which subjects may drop out during any period and identifies areas for further research.

In the final section of this chapter the estimation of treatment comparisons via least squares analysis for a cross-over trial is described for reference in later chapters of this thesis.

1.5 Analysis and Estimability

In this section an analysis is outlined for observations from a cross-over experiment to compare the effects of t treatments over p periods in which nm subjects are assigned to m treatment sequences so that n subjects receive each sequence.

We assume that the observations follow the model (1.1) which can be expressed in matrix form as

$$Y = 1_{mnp}\mu + X_s v_s + X_\pi v_\pi + X_\alpha v_\alpha + X_\lambda v_\lambda + \varepsilon, \tag{1.2}$$

where Y is an $mnp \times 1$ vector of observations, X_s, X_π, X_α and X_λ are matrices with mnp rows and mn, p, t and t columns respectively which hold the subject, period, direct and first-order carry-over treatment effects respectively and v_s, v_π, v_α and v_λ are vectors of length mn, p, t and t respectively which hold the parameters of interest for subjects, periods, direct and first-order carry-over treatment effects respectively. The vector 1_{mnp} of length mnp has every element unity and ε is a vector of length mnp which holds the random errors. Alternatively equation (1.2) can be written as:

$$Y = X\beta + \varepsilon, \tag{1.3}$$

where X is the partitioned matrix $(1_{mnp} \mid X_s \mid X_\pi \mid X_\alpha \mid X_\lambda)$ and β is the vector holding all the model parameters. The variance-covariance matrix of ε is I_{mnp}.

The following notation is used in this section and throughout the thesis. The transpose of a matrix M is denoted by M' and a generalised inverse of M is denoted M^-. The projection matrix onto the column space of M is denoted by $pr(M) = M(M'M)^-M'$. Let I_x denote the identity matrix of order x and $pr^\perp(M) = I_x - pr(M)$ denote the projection matrix onto the space which is orthogonal to the column space of M.

Suppose that, when the experiment is performed, there are $l_{i,j}$ dropouts on sequence j who complete i periods and then drop out during period $i+1$, where $i = 0, \ldots, p-1$ and $j = 1, \ldots, m$. Then the resulting incomplete set of observations and their model can be obtained by premultiplying each of the matrices in equation (1.2) by an $mnp - q \times mnp$ matrix L of zeros and ones, where q is the total number of missing observations. The matrix L acts on the terms in (1.2) in such a way that the rows corresponding to the missing observations are deleted and will be called the *loss matrix*. The purpose of introducing this matrix is to simplify the computations of the variance-covariance matrices, for all designs which may result from subjects dropping out, required for the investigations presented in Chapters 4 and 5.

Premultiplying equation (1.2) by the loss matrix gives the model

$$LY = L(X\beta + \varepsilon). \tag{1.4}$$

The ordinary least squares estimator of the vector of parameters β is obtained by minimising the error sum of squares, $\varepsilon'\varepsilon$, with respect to β. The normal equations can be shown to be

$$[X'L'LX]\hat{\beta} = X'L'LY. \tag{1.5}$$

On eliminating the subject and period effects from the normal equations we obtain

$$X'_\alpha K X_\alpha \hat{v}_\alpha + X'_\alpha K X_\lambda \hat{v}_\lambda = X'_\alpha KY \tag{1.6}$$

$$X'_\lambda K X_\alpha \hat{v}_\alpha + X'_\lambda K X_\lambda \hat{v}_\lambda = X'_\lambda KY \tag{1.7}$$

where

$$K = W[pr^{\perp}(WX_{\pi})]W$$
$$W = L'[pr^{\perp}(LX_{s})]L$$

1.5.1 Estimators of the Direct Treatment Effects

From equation (1.7) it can be shown that

$$\hat{v}_{\lambda} = [X'_{\lambda}KX_{\lambda}]^{-}\{X'_{\lambda}KY - X'_{\lambda}KX_{\alpha}\hat{v}_{\alpha}\}. \tag{1.8}$$

Substituting (1.8) into (1.6) we obtain

$$\{X'_{\alpha}KX_{\alpha} - X'_{\alpha}KX_{\lambda}[X'_{\lambda}KX_{\lambda}]^{-}X'_{\lambda}KX_{\alpha}\}\hat{v}_{\alpha} = \{X'_{\alpha}K - X'_{\alpha}KX_{\lambda}[X'_{\lambda}KX_{\lambda}]^{-}X'_{\lambda}K\}Y$$

Hence the reduced normal equations for estimating the direct treatment effects can be expressed as

$$A_{\alpha}\hat{v}_{\alpha} = Q_{\alpha} \tag{1.9}$$

where

$$A_{\alpha} = X'_{\alpha}KX_{\alpha} - X'_{\alpha}KX_{\lambda}[X'_{\lambda}KX_{\lambda}]^{-}X'_{\lambda}KX_{\alpha}$$

and

$$Q_{\alpha} = \{X'_{\alpha}K - X'_{\alpha}KX_{\lambda}[X'_{\lambda}KX_{\lambda}]^{-}X'_{\lambda}K\}Y.$$

A_{α} is the information matrix for estimating the direct treatment effects and Q_{α} is the vector of direct treatment totals after adjusting for subject, period and first-order carry-over treatment effects. Since $\text{rank}(A_{\alpha}) \leq t - 1$, there is no unique solution to (1.9). A solution can be obtained as

$$\hat{v}_{\alpha} = \Omega_{\alpha}Q_{\alpha}, \tag{1.10}$$

where Ω_{α} is a generalised inverse of A_{α}. The general solution of (1.10) is given by

$$\hat{v}_{\alpha} = \Omega_{\alpha}Q_{\alpha} + (\Omega_{\alpha}A_{\alpha} - I)Z, \tag{1.11}$$

where Z is an arbitrary vector, see Searle (1971, p 11).

1.5.2 Estimators of the First-order Carry-over Effects

From equation (1.6) it can be shown that

$$\hat{v}_\alpha = [X'_\alpha K X_\alpha]^- \{X'_\alpha K Y - X'_\alpha K X_\lambda \hat{v}_\lambda\} \tag{1.12}$$

On substituting (1.12) into (1.5) we obtain

$$\{X'_\lambda K X_\lambda - X'_\lambda K X_\alpha [X'_\alpha K X_\alpha]^- X'_\alpha K X_\lambda\}\hat{v}_\lambda = \{X'_\lambda K - X'_\lambda K X_\alpha [X'_\alpha K X_\alpha]^- X'_\alpha K\}Y.$$

Hence the reduced normal equations for estimating the first-order carry-over effects are

$$A_\lambda \hat{v}_\lambda = Q_\lambda, \tag{1.13}$$

where

$$A_\lambda = X'_\lambda K X_\lambda - X'_\lambda K X_\alpha [X'_\alpha K X_\alpha]^- X'_\alpha K X_\lambda$$

and

$$Q_\lambda = \{X'_\lambda K - X'_\lambda K X_\alpha [X'_\alpha K X_\alpha]^- X'_\alpha K\}Y.$$

A_λ is the information matrix for the carry-over treatment effects and Q_λ is the vector holding the totals of the carry-over treatment effects after adjusting for subject, period and direct treatment effects.

One solution to equations (1.13) is given by

$$\hat{v}_\lambda = \Omega_\lambda Q_\lambda, \tag{1.14}$$

where Ω_λ is a generalised inverse of A_λ. Again it is not possible to find a unique solution to the equations because rank $(A_\lambda) \leq t - 1$. The general solution of (1.13) is, for any specific Ω_λ,

$$\hat{v}_\lambda = \Omega_\lambda Q_\lambda + (\Omega_\lambda A_\lambda - I)Z, \tag{1.15}$$

where Z is an arbitrary vector.

1.5.3 Estimating Treatment Contrasts

In this section results on the estimators of treatment contrasts are given for reference later in the thesis.

Let C_α and C_λ be contrast matrices of rank at most $t-1$ in the treatment effects, that is $C_\alpha J_t = C_\lambda J_t = 0_t$, where $J_t = 1_t 1_t'$ and 0_t is a vector of length t with every element zero. Then the least squares estimators of $C_\alpha v_\alpha$ and $C_\lambda v_\lambda$ are, from (1.11) and (1.15),

$$C_\alpha \hat{v}_\alpha = C_\alpha \Omega_\alpha Q_\alpha + C_\alpha(\Omega_\alpha A_\alpha - I)Z \quad \text{and} \tag{1.16}$$

$$C_\lambda \hat{v}_\lambda = C_\lambda \Omega_\lambda Q_\lambda + C_\lambda(\Omega_\lambda A_\lambda - I)Z \tag{1.17}$$

where Z is an arbitrary vector.

The following lemmas give necessary and sufficient conditions for contrasts in the direct and carry-over treatment effects to be estimable and gives the properties of the estimators.

Lemma 1.1

(i) $C_\alpha \hat{v}_\alpha$ is unique and hence estimable if and only if

$$C_\alpha(\Omega_\alpha A_\alpha - I)Z = 0 \quad \text{for all } Z,$$

that is, if and only if

$$C_\alpha(\Omega_\alpha A_\alpha - I) = 0,$$

or

$$C_\alpha \Omega_\alpha A_\alpha = C_\alpha.$$

(ii) When $C_\alpha \Omega_\alpha A_\alpha = C_\alpha$

(a) $E(C_\alpha \hat{v}_\alpha) = C_\alpha v_\alpha$

(b) $\mathrm{Var}(C_\alpha \hat{v}_\alpha) = C_\alpha \Omega_\alpha C_\alpha' \sigma^2$.

Lemma 1.2

(i) $C_\lambda \hat{v}_\lambda$ is unique and hence estimable if and only if

$$C_\lambda(\Omega_\lambda A_\lambda - I)Z = 0 \quad \text{for all } Z,$$

that is, if and only if

$$C_\lambda(\Omega_\lambda A_\lambda - I) = 0,$$

or

$$C_\lambda \Omega_\lambda A_\lambda = C_\lambda.$$

(ii) When $C_\lambda \Omega_\lambda A_\lambda = C_\lambda$

(a) $E(C_\lambda \hat{v}_\lambda) = C_\lambda v_\lambda$

(b) $\text{Var}(C_\lambda \hat{v}_\lambda) = C_\lambda \Omega_\lambda C_\lambda' \sigma^2$.

1.5.4 The Case of No Dropouts

The case when all subjects successfully complete the study, that is when $L = I_{mnp}$, has been extensively researched. The purpose of this section is to outline briefly how the form of the information matrices given earlier in this section can be reconciled with the information matrices in the literature for the no dropout case.

When there are no dropouts the design matrices for subject and period effects are $X_s = I_{mn} \otimes 1_p$ and $X_\pi = 1_{mn} \otimes I_p$ respectively. Hence the reduced normal equations (1.9) for estimating the direct treatment effects become

$$A_\alpha \hat{v}_\alpha = Q_\alpha, \tag{1.18}$$

where

$$\begin{aligned} A_\alpha &= X_\alpha' K_1 X_\alpha - X_\alpha' K_1 X_\lambda [X_\lambda' K_1 X_\lambda]^- X_\lambda' K_1 X_\alpha \\ Q_\alpha &= \{X_\alpha' K_1 - X_\alpha' K_1 X_\lambda [X_\lambda' K_1 X_\lambda]^- X_\lambda' K_1\} Y \quad \text{and} \\ K_1 &= [I_{mn} - (1/mn) J_{mn}] \otimes [I_p - (1/p) J_p] \end{aligned}$$

Similarly, the reduced normal equations for estimating the first-order carry-over treatment effects become

$$A_\lambda \hat{v}_\lambda = Q_\lambda, \tag{1.19}$$

where

$$\begin{aligned} A_\lambda &= X'_\lambda K_1 X_\lambda - X'_\lambda K_1 X_\alpha [X'_\alpha K_1 X_\alpha]^- X'_\alpha K_1 X_\lambda \\ Q_\lambda &= \{X'_\lambda K_1 - X'_\lambda K_1 X_\alpha \ [X'_\alpha K_1 X_\alpha]^- X'_\alpha K_1\} Y \end{aligned}$$

where K_1 is as for (1.18). Equations (1.18) and (1.19) can be shown to be identical to different forms given in the literature, such as Kunert (1983) and Cheng and Wu (1980).

1.6 Design Selection Criteria

When assessing how well the contrasts of interest can be estimated from a cross-over design, it is common practice to assess the design under some appropriate optimality criterion $\psi[C\Omega C']$, as defined by Kiefer (1975), where C is C_α or C_λ and Ω is Ω_α or Ω_λ. In this section the design selection criteria used in this thesis are given for reference in later chapters.

Let D be the class of competing designs for a cross-over study, that is having the same numbers of subjects, periods and treatments.

Definition 1.1 - The A-criterion A design $d^* \in D$ is **A-optimal over** D for the estimation of the contrasts of interest $C_\alpha v_\alpha$ in the direct effects if d^* minimises the average variance of the contrast estimators; that is,

$$tr[C_\alpha \Omega_\alpha(d^*) C'_\alpha] \leq tr[C_\alpha \Omega_\alpha(d) C'_\alpha]$$

for all $d \in D$, where $\Omega_\alpha(d)$ is a generalised inverse of the direct treatment information matrix for design d.

The definition of an A-optimal design for the estimation of the contrasts of interest in the first-order carry-over effects can be obtained by replacing α by λ in Definition 1.1. When choosing between competing designs in D, in addition to calculating the average variance of the treatment contrasts of interest, it is also advisable to examine the maximum variance that is obtained from the designs. This is the purpose of the following criterion.

Definition 1.2 - The MV-criterion A design d^* is **MV-optimal over** D for the estimation of the contrasts of interest $C_\alpha v_\alpha$ in the direct effects if the maximum variance for the contrast estimators under d^* is less than or equal to the maximum variance for the estimators obtained for each $d \in D$.

An MV-optimal design for the estimation of first-order carry-over treatment contrasts is defined similarly.

It should be noted that, when the contrasts of interest $C_\alpha v_\alpha$ or $C_\lambda v_\lambda$ form a complete set of orthonormalised contrasts, such as the basic contrasts, the MV-criterion is identical to the E-criterion which seeks to minimise the maximum eigenvalue of the information matrix for the contrasts of interest.

For some experiments designs have been obtained which are optimal for a range of criteria. For example, Cheng and Wu (1980) identify designs for estimating direct and carry-over effects which are optimal in the following sense.

Definition 1.3 [Kiefer (1975)] - Universal optimality Suppose that $\mathcal{S}$ is the set of $t \times t$ non-negative definite matrices with zero row and column sums and $\Phi : \mathcal{S} \longrightarrow (-\infty, \infty]$ satisfies:

(i) Φ is convex,

(ii) $\Phi[bA_\alpha(d)]$ is nonincreasing in the scalar $b \geq 0$, (1.20)

(iii) Φ is invariant under each permutation of rows and of columns,

where $A_\alpha(d)$ is the information matrix for the direct treatment effects for design d.

A design d^* is **universally optimal** for the estimation of contrasts of interest $C_\alpha v_\alpha$ in the direct effects if $\Phi[A(d^*)] \leq \Phi[A(d)]$ for all $d \in D$ and every Φ satisfying (1.20).

A universally optimal design for the estimation of first-order carry-over treatment comparisons is defined similarly.

A design which is found to be universally optimal will also be A- and MV-optimal. However, a design which is optimal under one criterion is not necessarily optimal under the other criterion. Nevertheless, it is usual for a design which is optimal under one criterion to perform well under the other criterion.

Chapter 2

Design Assessment and Selection

2.1 Introduction

One of the most important decisions made during the planning stage of any cross-over trial is the design selection. Given that a study aims to compare t treatments over a maximum of p periods using s subjects, which of the many designs available should be chosen? An appropriate design is one which is efficient for estimating the treatment comparisons of interest, most commonly all comparisons of pairs of direct treatment effects or a set of orthogonal contrasts such as the orthogonal polynomial contrasts.

A summary of some of the most commonly used design selection criteria has been given in Chapter 1. These criteria do not, however, assess the performance of designs when subjects may drop out during the later stages of the trial. Hence current practice is to select a design on the basis of its performance assuming no subjects drop out. This is an unrealistic assumption for many medical applications. For example, in an arthritis trial over four treatment periods separated by wash-out periods, patients often leave the study in the final period (or even earlier) for reasons unconnected with the trial, such as leaving the geographical area.

Matthews (1988), in his discussion concerning the optimality properties of cross-over designs, queried how robust these properties might be to dropouts. Herzberg

and Andrews (1976) considered the effects of missing values and outliers on optimal response surface designs, and defined a measure of expected average precision to compare designs. In this chapter a similar approach is adopted for comparing designs for cross-over trials; the main difference is that designs are judged by how accurately the pairwise treatment comparisons can be estimated with the assumption that each subject has a fixed probability, θ, of dropping out in the final period, independent of treatment. These assumptions may be somewhat unrealistic. However, all the ideas introduced in this chapter can easily be extended to situations where subjects may drop out during any period or when it is anticipated that treatment-dependent dropout may occur, for instance when one treatment is a placebo.

In the following section the definitions and notation used to develop the method of assessing cross-over designs when subjects may drop out are given. In Section 2.3 the probability of realising any of the designs which can occur when one or more subjects drop out during the final period is formulated. Two ways of assessing the performance of cross-over designs subject to final period dropout are described in Section 2.4 and in Sections 2.5 and 2.6 the preferred method is illustrated. Finally, in Sections 2.7 and 2.8 the issues of sensitivity and robustness to the probability of final period dropout are discussed. Criteria for robust designs are formulated and used to establish a set of design selection criteria which is used, in latter chapters, to compare the robustness of competing designs to dropouts.

2.2 Definitions and Notation

To consider cross-over designs subject to patient dropout the following definitions and notation will be employed.

Definition 2.1 The **planned design** is a cross-over design adopted at the outset of a study to compare t treatments over p periods using mn subjects with n subjects allocated to each of m distinct treatment sequences.

The planned design is denoted by $d(t, m, n, p)$.

Definition 2.2 The design realised through applying the planned design in a cross-over trial is called the **implemented design**.

In order to list the possible implemented designs the following notation is used.

Notation When l_i dropouts occur in the final period of sequence i $(i = 1, \ldots, m)$, we denote the implemented design by $d_l(t, m, n, p)$, where l is the m-tuple $l_1 l_2 \ldots l_m$. Note that $d(t, m, n, p) = d_0(t, m, n, p)$, where $0 = 00 \ldots 0$.

Definition 2.3 The set of all **implementable designs**, D, is the set of all possible realisable designs, that is consisting of $d(t, m, n, p)$ together with all designs formed by dropping one or more subjects in the final period of $d(t, m, n, p)$

A more concise definition of D is achieved using the following notation.

Notation Let $G_{m,n} = \{\, l;\ l = l_1 \ldots l_m,\ l_i = 0, \ldots, n,\ i = 1, \ldots, m \,\}$
i.e. $G_{m,n}$ holds all possible dropout patterns which could be incurred by $d(t, m, n, p)$.
Then $D = \{\, d_l(t, m, n, p);\ l \in G_{m,n} \,\}$.
Note that the size of D, denoted by $|D|$, is given by $|D| = (n + 1)^m$.

In this chapter we consider only implementable designs resulting from dropouts occurring in the final period. An example of a set of implementable designs is given in Example 2.1. In the example, and throughout this thesis, designs will be shown as two-way layouts with rows corresponding to treatment sequences and columns corresponding to periods.

Example 2.1 Let $d(4, 4, 1, 4)$ be the Williams square of side four for treatment labels 0, 1, 2 and 3 and initial treatment sequence (0 1 3 2). This design is labelled d_{0000} in Table 2.1. The set, D, of $2^4 = 16$ implementable designs and its elements

are listed in Table 2.1. In Table 2.1, and throughout this thesis, * is used to indicate that a subject has dropped out.

In this chapter we assess the performance of any planned design, subject to a fixed probability, θ, of final period dropout, by assessing each member of its associated set of implementable designs, D. From these individual assessments it will be possible to produce summary measures of overall performance or robustness of the planned design.

2.3 Probability of Implementing d_l

Assume that each subject has a fixed probability, θ, of dropping out in the final period independent of treatment. For each $i = 1, \ldots, m$ let Z_i be the total number of subjects in the ith sequence group who drop out. Then the Z_i's are independently distributed binomial random variables with parameters n and θ, i.e $Z_i \sim$ B(n, θ). Hence the joint probability function for the numbers of final period dropouts on each treatment sequence is

$$\begin{aligned} P(l_1, \ldots, l_m|\theta) &= \prod_{i=1}^{m} P(Z_i = l_i|\theta) \\ &= \prod_{i=1}^{m} \binom{n}{l_i} \theta^{l_i}(1-\theta)^{n-l_i} \end{aligned} \tag{2.1}$$

This is the probability that the design d_l, where $l = l_1 \ldots l_m$, is implemented in practice. Hence, for each $d_l \in D$ we can calculate the probability that d_l is the realised design, as in the following example.

Example 2.2 Let $d(4, 4, 1, 4)$ be the Williams square given in Example 2.1, whose set of implementable designs is listed in Table 2.1. The probability of implementation for each design $d_l \in D$, calculated from equation (2.1), is given in Table 2.2.

Table 2.1: Full set, D, of implementable designs for Example 2.1.

0 Dropouts

d_{0000}

0 1 3 2

1 2 0 3

2 3 1 0

3 0 2 1

1 Dropout

d_{1000}	d_{0100}	d_{0010}	d_{0001}
0 1 3 *	0 1 3 2	0 1 3 2	0 1 3 2
1 2 0 3	1 2 0 *	1 2 0 3	1 2 0 3
2 3 1 0	2 3 1 0	2 3 1 *	2 3 1 0
3 0 2 1	3 0 2 1	3 0 2 1	3 0 2 *

2 Dropouts

d_{1100}	d_{1010}	d_{1001}	d_{0110}	d_{0101}	d_{0011}
0 1 3 *	0 1 3 *	0 1 3 *	0 1 3 2	0 1 3 2	0 1 3 2
1 2 0 *	1 2 0 3	1 2 0 3	1 2 0 *	1 2 0 *	1 2 0 3
2 3 1 0	2 3 1 *	2 3 1 0	2 3 1 *	2 3 1 0	2 3 1 *
3 0 2 1	3 0 2 1	3 0 2 *	3 0 2 1	3 0 2 *	3 0 2 *

3 Dropouts

d_{0111}	d_{1011}	d_{1101}	d_{1110}
0 1 3 2	0 1 3 *	0 1 3 *	0 1 3 *
1 2 0 *	1 2 0 3	1 2 0 *	1 2 0 *
2 3 1 *	2 3 1 *	2 3 1 0	2 3 1 *
3 0 2 *	3 0 2 *	3 0 2 *	3 0 2 1

4 Dropouts

d_{1111}

0 1 3 *

1 2 0 *

2 3 1 *

3 0 2 *

Table 2.2: Probabilities of implementation for each design $d_l \in D$ for Example 2.2.

Design	Probability of Implementation
d_{0000}	$(1-\theta)^4$
d_{1000}	$\theta(1-\theta)^3$
d_{0100}	$\theta(1-\theta)^3$
d_{0010}	$\theta(1-\theta)^3$
d_{0001}	$\theta(1-\theta)^3$
d_{1100}	$\theta^2(1-\theta)^2$
d_{1010}	$\theta^2(1-\theta)^2$
d_{1001}	$\theta^2(1-\theta)^2$
d_{0110}	$\theta^2(1-\theta)^2$
d_{0101}	$\theta^2(1-\theta)^2$
d_{0011}	$\theta^2(1-\theta)^2$
d_{0111}	$\theta^3(1-\theta)$
d_{1011}	$\theta^3(1-\theta)$
d_{1101}	$\theta^3(1-\theta)$
d_{1110}	$\theta^3(1-\theta)$
d_{1111}	θ^4

2.4 Assessment of Design Performance Subject to Final Period Dropout

To assess the performance of a cross-over design, d, subject to a fixed probability, θ, of final period dropout, we consider the set, D, of all possible implemented designs derived from d.

We assume that the purpose of the experiment is to estimate contrasts of interest defined by $C_\alpha v_\alpha$ and $C_\lambda v_\lambda$, where C_α and C_λ are contrast matrices holding the coefficients of the contrasts of interest and v_α and v_λ are vectors holding the direct and first-order carry-over treatment effects respectively.

2.4.1 Design Requirements

In some practical cases D contains one or more disconnected designs, that is designs where it is no longer possible to estimate all the contrasts of interest. For example, if the aim of a study is to compare all the pairwise direct treatment effects and the design given in Example 2.1 is considered, it can be shown that any design in D which contains two or more final period dropouts is disconnected. From Table 2.2 the probability of implementing a disconnected design is $[6\theta^2(1-\theta)^2+4\theta^3(1-\theta)+\theta^4]$.

Let $D_0 \subseteq D$ denote the set of disconnected designs derived from d and $\mathrm{P}(D_0)$ denote the probability of implementing a disconnected design. Ideally, a design d should generate no disconnected implementable designs. However, for small values of n and p this is not always achievable. In this case, a desirable property of d is that the probability of implementing a disconnected design should be small. If the probability is unacceptably large then an alternative design should be adopted.

Let $D_c \subseteq D$ denote the set of connected designs derived from d. An additional requirement of any planned design is that each of the designs in D_c estimates, as accurately as possible, the contrasts of interest in the direct and first-order carry-over treatment effects. In the next section ways of assessing how well a planned

design meets this requirement will be considered.

2.4.2 Methods of Assessment

Current practice when assessing the performance of a cross-over design is to consider how well it performs under one or more appropriate optimality criteria. Kiefer (1975) gave a general formulation of a criterion as $\psi[C\,\Omega(d)\,C']$ where $C = C_\alpha$ or C_λ, and $\Omega(d)$ is a generalised inverse of the information matrix for the direct or for the first-order carry-over treatment effects for the planned design, see Section 1.6. For example, when $\psi[C\,\Omega(d)\,C'] = \mathrm{tr}[C\,\Omega(d)\,C']$ or the maximum eigenvalue of $C\,\Omega(d)\,C'$, then the criterion is A- or E- respectively, applied to the direct or first-order carry-over treatment effects. In this section two approaches for assessing cross-over designs subject to dropout are described.

1. The Direct Approach

Each of the connected implementable designs can be assessed individually using any of the above optimality criteria. One approach, therefore, is to calculate the **expected average variance** of the direct and first-order carry-over treatment effects for the set of connected designs. This expectation must be restricted to the set of connected implementable designs, since the average variance of the treatment effects is not obtainable for any of the disconnected designs.

This method of assessment has the advantage of being a simple extension of the procedure employed when the probability of dropout is not an issue. The disadvantage of using this method, however, is that the performance measures obtained are conditional expectations which cannot take into account the disconnected designs. When D contains one or more disconnected designs the performance measures will give a misleadingly good impression of the design's expected performance in repeated use in experiments. This will cause a problem when trying to compare the relative performance of two competing designs which have different numbers of disconnected

implementable designs.

2. Alternative Approach

In order to overcome these problems, an alternative method of assessing a planned design, d, is required which involves performance measures for both the connected and disconnected designs in D. These measures are now defined for general sets of contrasts in the direct and in the first-order carry-over treatment effects. In the following definitions $\Re^+$ denotes the non-negative real numbers.

Definition 2.4 Let $X_d{:}D \rightarrow \Re^+$ be a random variable such that

$$X_d(d_l) = \begin{cases} [\psi(C_\alpha\, \Omega_\alpha(d_l)\, C_\alpha')]^{-1} & \text{for} \quad d_l \in D \backslash D_0 \\ 0 & \text{for} \quad d_l \in D_0 \quad , \end{cases} \tag{2.2}$$

where ψ is a measure of design performance, C_α is a matrix holding the coefficients of the contrasts of interest in the direct treatment effects and $\Omega_\alpha(d_l)$ is a generalised inverse of the information matrix of d_l for estimating the direct treatment effects.

Thus the random variable X_d takes values which are reciprocals of the value of the function ψ for the direct treatment effects for the individual implementable designs.

Definition 2.5 Let $Y_d{:}D \rightarrow \Re^+$ be a random variable such that

$$Y_d(d_l) = \begin{cases} [\psi(C_\lambda\, \Omega_\lambda(d_l)\, C_\lambda')]^{-1} & \text{for} \quad d_l \in D \backslash D_0 \\ 0 & \text{for} \quad d_l \in D_0 \quad , \end{cases} \tag{2.3}$$

where ψ is a measure of design performance, C_λ is a matrix holding the coefficients of the contrasts of interest in the first-order carry-over treatment effects and $\Omega_\lambda(d_l)$ is a generalised inverse of the information matrix of d_l for estimating the first-order carry-over treatment effects.

Similarly, the random variable Y_d takes values which are the reciprocals of the function ψ for the first-order carry-over treatment effects for the individual implementable designs.

The probability distribution of X_d for a constant probability, θ, of final period dropout is obtained from (2.1) as

$$P(X_d = x|\theta) = \sum_{l \in L} P(l|\theta) \ ,$$

where $L = \{l \in G_{m,n} \ ; d_l \in D, \ X_d(d_l) = x\}$.

Similarly, the probability distribution of Y_d for a constant probability, θ, of final period dropout is

$$P(Y_d = y|\theta) = \sum_{l \in L} P(l|\theta) \ ,$$

where $L = \{l \in G_{m,n} \ ; d_l \in D, \ Y_d(d_l) = y\}$.

In order to summarise the performance of a planned design under repeated use in experiments with probability θ of final period dropout, the mean and variance of the random variables X_d and Y_d can be examined.

Definition 2.6 For the planned design d define the mean of X_d by

$$E[X_d|\theta] = \sum_{d_l \in D} X_d(d_l) P(l|\theta) = \sum_{d_l \in D \setminus D_0} [\psi(C_\alpha \, \Omega_\alpha(d_l) \, C'_\alpha)]^{-1} \, P(l|\theta) \ , \qquad (2.4)$$

where ψ, C_α and $\Omega_\alpha(d_l)$ are as in Definition 2.4 and X_d is the random variable defined in (2.2) for the planned design d.

Definition 2.7 For the planned design d the variance of X_d is

$$\mathrm{Var}[X_d|\theta] = E\{X_d - E[X_d|\theta]\}^2 \ . \qquad (2.5)$$

A similar summary of design performance for estimating the first-order carry-over treatment effects can be made using the following definitions.

Definition 2.8 For the planned design d define the mean of Y_d by

$$E[Y_d|\theta] = \sum_{d_l \in D} Y_d(d_l) P(l|\theta) = \sum_{d_l \in D \setminus D_0} [\psi(C_\lambda \, \Omega_\lambda(d_l) \, C'_\lambda)]^{-1} \, P(l|\theta) \ , \qquad (2.6)$$

where ψ, C_λ and $\Omega_\lambda(d_l)$ are given in Definition 2.5 and Y_d is the random variable defined in (2.3) for the planned design d.

Definition 2.9 For the planned design d the variance of Y_d is

$$\mathrm{Var}[Y_d|\theta] = E\{Y_d - E[Y_d|\theta]\}^2 . \tag{2.7}$$

In the next section the second approach is illustrated for the A-criterion and a Williams square design.

2.5 Illustration using the A-criterion

Cross-over trials are commonly used in studies which aim to compare the efficacy of two or more treatments. A natural criterion for design selection and assessment in cross-over trials is, therefore, the A-criterion which seeks to minimise the average variance of the contrasts of interest, see Section 1.6. The criterion can be applied to each of the direct and first-order carry-over treatment effects through equations (2.4) and (2.6) where C_α, C_λ, $\Omega_\alpha(d_l)$ and $\Omega_\lambda(d_l)$ are as defined in the previous section and ψ is $tr[C_\alpha\Omega_\alpha(d_l)C'_\alpha]$ or $tr[C_\lambda\Omega_\lambda(d_l)C'_\lambda]$. Since $tr[C_\alpha\Omega_\alpha(d_l)C'_\alpha]$ and $tr[C_\lambda\Omega_\lambda(d_l)C'_\lambda]$ appear as reciprocals in (2.4) and (2.6) a "good" design is one which has $E[X_d|\theta]$ as large as possible; similarly for $E[Y_d|\theta]$. Example 2.3 illustrates the use of this method for assessing a planned design, d, in the presence of final period dropout.

Example 2.3 Let $d(4,4,4,4)$ be the Williams square of side four for treatment labels 0, 1, 2 and 3 and initial treatment sequence (0 1 3 2). Assume we wish to estimate

1. the pairwise direct treatment comparisons, $\alpha_i - \alpha_j$ for $1 \leq i < j \leq t$ and,
2. the pairwise first-order carry-over treatment comparisons, $\lambda_i - \lambda_j$ for $1 \leq i < j \leq t$.

The number of implementable designs realisable from d is

$$|D| = (n+1)^m = 5^4 = 625$$

An examination of all implementable designs shows that the number of disconnected implementable designs is $|D_0| = 113$.

As some of the potentially implementable designs are disconnected, it is necessary to examine $\mathrm{P}(D_0)$, the probability of implementing a disconnected design. Applying equation (2.1) to each of the disconnected implementable designs and considering the whole range of possible values for θ, $0 \leq \theta \leq 1$, we obtain the distribution of $\mathrm{P}(D_0)$ given in Table 2.3.

Table 2.3: Probabilities of implementing a disconnected design for a Williams square of side four with 16 subjects, $d(4,4,4,4)$.

Probability of Dropout (θ)	Probability d_l Disconnected $\mathrm{P}(D_0)$
0.0	0.00
0.1	6.0×10^{-8}
0.2	2.0×10^{-4}
0.3	3.9×10^{-4}
0.4	3.8×10^{-3}
0.5	2.2×10^{-2}
0.6	8.4×10^{-2}
0.7	0.25
0.8	0.54
0.9	0.88
1.0	1.00

It is reasonable to assume that any design having a probability greater than 0.2 of producing a disconnected implementable design would not be acceptable to an

experimenter. If we examine the probabilities given in Table 2.3 we observe that, for the small practical values of θ, the probability of implementing a disconnected design is very small. For $\theta \leq 0.6$ this probability is less than 0.09. When $\theta > 0.6$, however, the probability rapidly increases beyond 0.2, at which point an experimenter may reasonably consider that this represents too great a risk to proceed with the planned design. In these circumstances we could either increase the number of subjects allocated to each of the sequences sequences or choose an alternative design with a smaller probability of realising a disconnected design. For this example, if the probability of subjects dropping out in the final period of a four period study is anticipated to be as high as 0.7, then there is a strong case for adopting a three period design instead.

Applying equations (2.4), (2.5), (2.6) and (2.7) to the A-criterion we can obtain the mean and variance of the performance measures X_d and Y_d, over the range of possible θ values, $0 \leq \theta \leq 1$. These are given in Table 2.4 and they provide a summary of the performance of the average variance of the direct and first-order carry-over treatment effects for the planned design. Figures 2.1 and 2.2 show how $E[X_d|\theta]$ and $E[Y_d|\theta]$ change with θ. Note that the bars represent $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$ and $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$ which give an indication of the spread of the distribution of X_d and Y_d respectively. In these figures and throughout the thesis, when $E[X_d|\theta] - \sqrt{\text{Var}[X_d|\theta]}$ or $E[Y_d|\theta] - \sqrt{\text{Var}[Y_d|\theta]}$ is less than zero the bars are not shown below the θ axis.

From Figure 2.1 we observe that as θ increases there is a gradual reduction in the mean of X_d from a maximum value of 7.27, when $\theta = 0.0$, to a minimum of 0.00, when $\theta = 1.0$. Note that, when the probability of final period dropout is anticipated to be $\theta = 0.0$, the set of implementable designs D contains only one design, that is the planned design. When the probability of final period dropout is anticipated to be $\theta = 1.0$, D contains only one design, that is the planned design with the entire final period deleted which in this case is disconnected.

Table 2.4: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for the Williams square of side four with 16 subjects, $d(4,4,4,4)$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.90	0.10	4.68	0.07
0.2	6.46	0.26	4.32	0.17
0.3	5.95	0.53	3.91	0.32
0.4	5.33	0.97	3.43	0.53
0.5	4.58	1.62	2.87	0.79
0.6	3.63	2.40	2.22	1.03
0.7	2.50	2.86	1.48	1.10
0.8	1.26	2.21	0.73	0.77
0.9	0.28	0.61	0.16	0.20
1.0	0.00	0.00	0.00	0.00

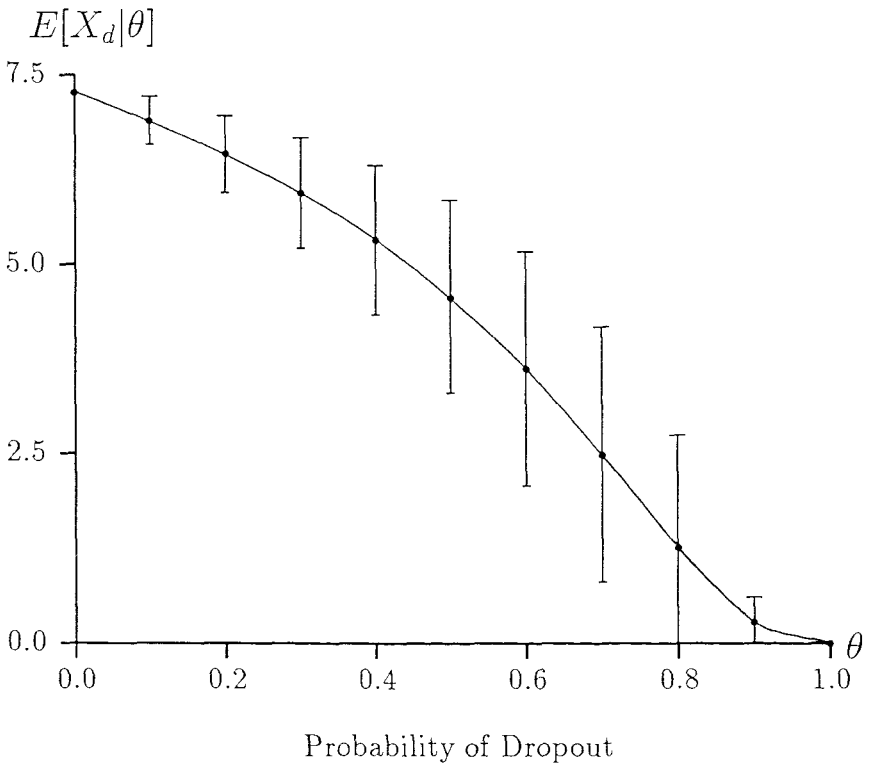


Figure 2.1: Summary of the performance of the Williams square of Example 2.3 in estimating the pairwise direct treatment effects under the A-criterion, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

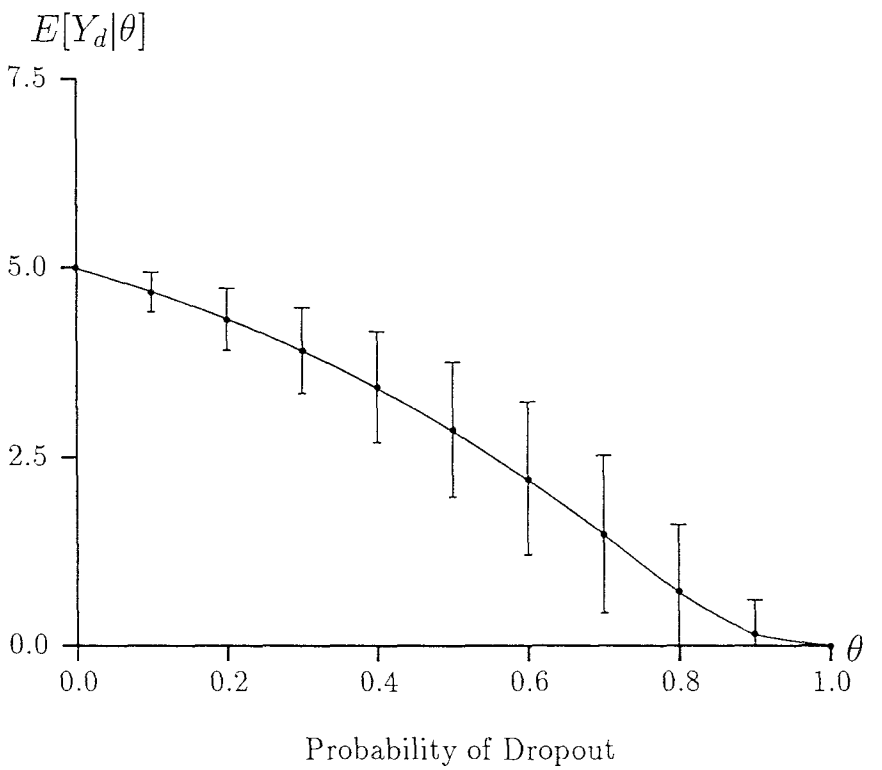


Figure 2.2: Summary of the performance of the Williams square of Example 2.3 in estimating the pairwise first-order carry-over treatment effects under the A-criterion, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

An examination of all the designs in D shows that, for this example, all implementable designs with two or more sequence groups containing $l = n = 4$ dropouts in the final period are disconnected. There are 113 disconnected implementable designs each containing at least eight dropouts in the final period. This partly explains the sudden reduction in the mean value of both performance measures at around $\theta = 0.5$, which produces an increase in the gradient of each of the curves in Figures 2.1 and 2.2 when $\theta \geq 0.5$.

The graphs showing the mean of the performance measures X_d and Y_d are similar but, as expected, the amount of information in the direct treatment effects is consistently greater than that in the first-order carry-over treatment effects. We can conclude, therefore, that the rate of loss of information due to final period dropout is fairly consistent for both the direct and first-order carry-over treatment effects for this design.

Note that, regardless of the value of θ, each of the distributions for X_d will have a maximum value of 7.27 contributed by the planned design and a minimum value of 0.00. Similarly, each of the distributions for Y_d, regardless of the value of θ, will have a maximum value of 5.00 contributed by the planned design and a minimum value of 0.00.

The spread of a distribution will always be smaller when θ is very small or very large. When θ is very large the distributions for X_d and Y_d will be positively skewed and when θ is very small the distributions will be negatively skewed. The skews occur because the distributions will be dominated by the performance measures for the designs with the greatest probability of being implemented. When θ is small, the most probable designs are those with high performance measures. When θ is large, the most probable designs are those with the poorer performance measures; in our particular example, the disconnected designs. The distributions with the greatest spread will be those with values of θ close to 0.5 because each of the implementable designs will then have a similar probability of being realised. All these properties are illustrated in Figures 2.1 and 2.2.

It can be argued that when a distribution is skewed it may be appropriate to consider the median value of the distribution in addition to the mean. For very small values of θ, therefore, the median of the distribution of X_d and Y_d could be used as an additional summary measure. This would not necessarily be appropriate for large values of θ since, as mentioned previously, if the probability of dropouts occurring in the final period is anticipated to be large then an alternative design with fewer periods should be sought.

2.6 Illustration using the MV-criterion

Frequently, a cross-over trial is planned with the aim of estimating all the pairwise treatment comparisons within the direct effects with equal precision, that is so that $\text{Var}(\hat{\alpha}_i - \hat{\alpha}_j) = k\sigma^2$, where k is constant for all $i \neq j$. A design having this property is called **variance balanced** for the estimation of direct treatment effects. Unfortunately, even if variance balance has been achieved in the planned design, the property is rarely retained when dropouts occur. If variance balance is desired, but dropouts are anticipated, it is desirable to obtain an indication of the spread in the variances of the pairwise treatment comparisons which might reasonably be expected in the implemented experiment. One approach is to calculate the difference between the highest and lowest pairwise variances achieved for each design in D and then summarise this in some way.

Ideally, designs which have high variances on some of the contrasts should be avoided. By considering only the A-criterion, these designs will not always be detected. If the performance of designs is considered under the MV-criterion which seeks to minimise the maximum variance of the contrasts of interest, designs will be discovered which give rise to high variances on some of the contrasts.

The MV-criterion can be applied to each of the direct and first-order carry-over treatment effects through equations (2.2) and (2.3) by letting $\psi[C_i\Omega_i(d_l)C_i']$ equal the maximum variance of $C_i\Omega_i(d_l)C_i'$ over all $d_l \in D\backslash D_0$ for $i = \alpha$ and λ. Since the value

of the maximum variance appears as a reciprocal in the performance measures given by equations (2.4) and (2.6), a "good" design for estimating the direct treatment effects will be one which has $E[X_d|\theta]$ as large as possible. Similarly, a "good" design for estimating the first-order carry-over effects will be one which has $E[Y_d|\theta]$ as large as possible. By considering these performance measures, in addition to those obtained using the A-criterion, we have a further way of assessing a planned design under dropout.

Example 2.4 We assess the Williams square of Example 2.3 using performance measures based on the MV-criterion. Using equations (2.4), (2.5), (2.6) and (2.7) with the MV-criterion we can obtain the mean and variance of the performance measures X_d and Y_d, over the range of possible θ values, $0 \leq \theta \leq 1$. These are given in Table 2.5 and they provide a summary of the performance of the maximum variance of the direct and carry-over treatment effects for the planned design. The results are illustrated in Figures 2.3 and 2.4 respectively.

Discussion of Examples 2.3 and 2.4. The observations made concerning the probability of implementing a disconnected design and the general trend with θ for the mean of X_d and Y_d for the design assessed under the A-criterion (see Example 2.3) all apply to Example 2.4. In Example 2.3, the effect of final period dropout on the estimation of the average variance of the treatment comparisons was examined. In Example 2.4, since the MV-criterion is used to calculate the performance measures X_d and Y_d, the effect of final period dropout on the maximum variance of the treatment comparisons can be observed. These two estimates are not unconnected. When a design is variance balanced the performance measures obtained using the A- or MV-criterion will be identical. When there are no final period dropouts the design of the example is variance balanced. When final period dropouts occur, however, the design is only varianced balanced if the number of subjects dropping out is the same for each treatment sequence. By examining the mean of the performance measures

Table 2.5: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for a Williams square of side four with 16 subjects, $d(4,4,4,4)$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.60	0.26	4.45	0.17
0.2	5.97	0.56	3.94	0.34
0.3	5.30	0.96	3.41	0.52
0.4	4.55	1.43	2.85	0.69
0.5	3.70	1.90	2.26	0.81
0.6	2.77	2.18	1.65	0.85
0.7	1.77	2.00	1.04	0.72
0.8	0.83	1.20	0.48	0.40
0.9	0.17	0.25	0.10	0.08
1.0	0.00	0.00	0.00	0.00

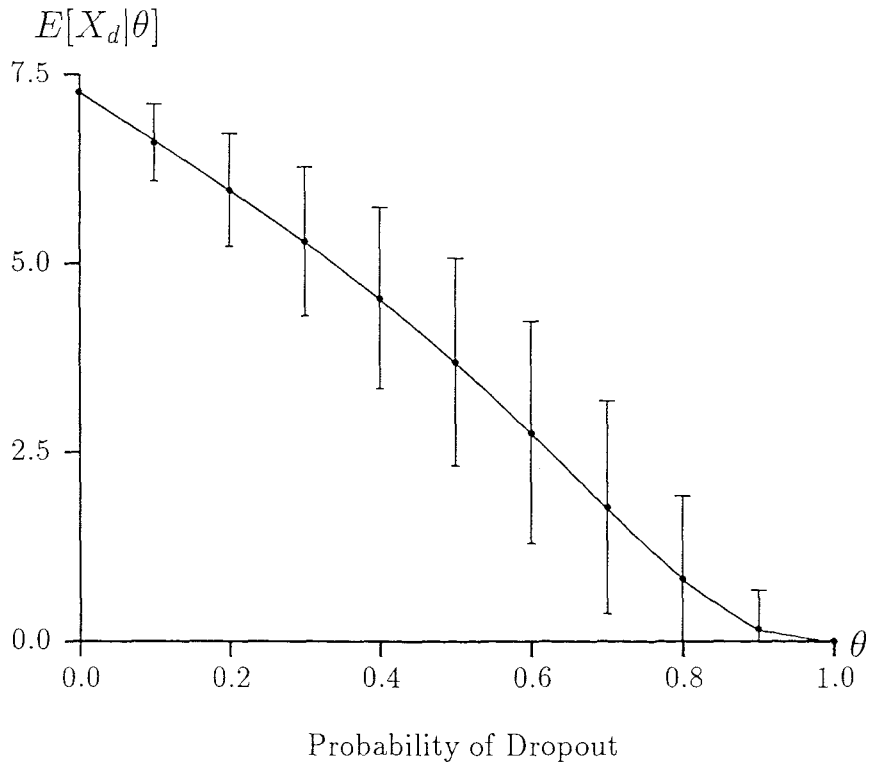


Figure 2.3: Summary of the performance of the Williams square of Example 2.4 in estimating the pairwise direct treatment effects under the MV-criterion, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

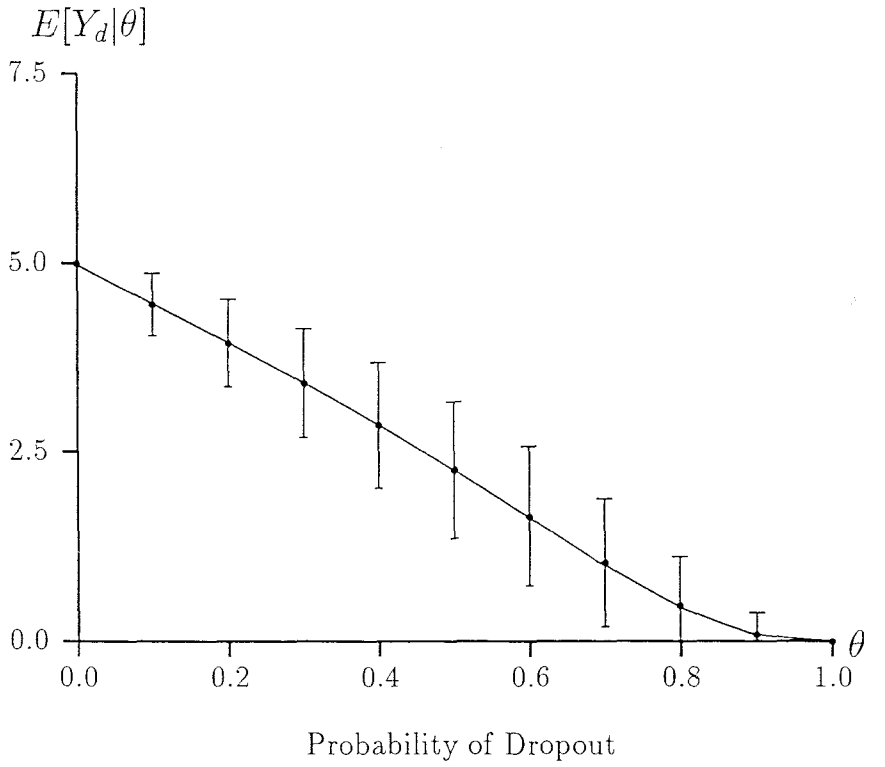


Figure 2.4: Summary of the performance of the Williams square of Example 2.4 in estimating the pairwise first-order carry-over treatment effects under the MV-criterion, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

X_d and Y_d, under the MV-criterion, an insight into whether the planned design may give rise to an implemented design having high variances for some contrasts but not for others will be obtained.

Comparing Tables 2.4 and 2.5 it can be observed that the mean values of X_d and Y_d, under the MV-criterion are always smaller than their corresponding A-criterion values. This is true for any design by definition, since the MV-criterion considers only the maximum variance of the contrasts of interest while the A-criterion finds their average. Hence, the reciprocal value of the A-criterion value will always be larger than the corresponding MV-criterion value. For the small practical values of θ, however, they do not differ by very much. From this it is possible to conclude that, although variance balance is unlikely to occur in the implemented experiment, the spread across the different treatment comparisons should not be too great, particularly for small values of θ. For example, when $\theta = 0.2$ the mean performance measures for estimating the direct treatment comparisons under the A- and MV-criteria are 6.46 and 5.97, respectively. Since the performance measure obtained from the maximum variance of the treatment comparisons for each of the implementable designs does not differ from that achieved by considering the average value of all the treatment comparisons it is reasonable to conclude that the spread in the variances of the pairwise treatment comparisons for each of the implementable designs can never be large.

Examining the trends of the mean performance measures for the Williams square of side four given in Examples 2.3 and 2.4 we observe that there is a gradual loss of information expected in the direct and carry-over treatment effects as θ increases. Even if only one subject is lost in the final period of an experiment the information available in the resulting implemented design will be less than that of the planned design. These examples illustrate that even for small anticipated numbers of dropouts the information available in the implemented experiment may differ considerably from that of the original planned design. In the design assessment procedures in

current use, little account is taken of this reduction in performance. By using the proposed assessment procedures it is possible to assess any planned design more realistically for trials in which final period dropout is anticipated.

2.7 Sensitivity to Choice of θ

A feature of the proposed design assessment procedure is its Bayesian nature. The probability of final period dropout, used to obtain the performance measures X_d and Y_d, is an anticipated value obtained by considering information provided from previous similar studies. Ideally, a study will have a proportion of dropouts in the final period which is close to, but not necessarily identical to, the anticipated number of dropouts used at the early planning stage to guide design choice. It is important, therefore, that the sensitivity of a planned design's performance to changes in θ is investigated. For this reason it is proposed that the probability of disconnectivity, and the mean and variance of the performance measures are examined over the entire range of possible θ values. By doing this it will be possible to detect whether a design is sensitive to changes in θ and, if so, the region of θ values which lead to large changes in design performance. This is particularly important if a planned design gives rise to disconnected implementable designs for certain patterns of final period dropout. For instance, in Example 2.3, when $\theta = 0.6$ the probability of implementing a disconnected design is less than 0.02. However, when $\theta = 0.7$, an increase of only 0.1, this probability rises dramatically to 0.25. The recommendation concerning whether to proceed with this planned design would be different in each case.

Clearly, it is very important that a design be robust to a mis-specification of the probability of final period dropout. Ideally, designs which are robust to changes in θ are preferable to those which are very sensitive. This is because the performance measures considered at the planning stage will give a more realistic impression of what might reasonably be achieved in the implemented experiment.

If it is undesirable to consider the entire range of θ values, it may be appropriate to consider only those values of θ close to the anticipated value. For example, if the investigators are confident that the probability of subjects dropping out in the final period is in the region of 0.2 then it would be reasonable to only consider $0.1 \leq \theta \leq 0.3$.

2.8 Robust Designs

Any implemented design containing final period dropouts will contain less information in the direct and first-order carry-over treatment effects than its parent planned design. It is important, therefore, to establish the characteristics of designs which indicate whether or not they are robust to final period dropout.

An additional, highly desirable, property of any planned design is that for the contrasts of interest in the direct and carry-over treatment effects, the design outperforms all other candidate designs over the entire range of possible values for θ. In this section criteria for robust designs are formulated and then used to establish a set of design selection criteria.

Consider the set, S, of all possible cross-over designs which compare t treatments over p periods using mn subjects with n subjects allocated to each of m treatment sequences. The design in S which is most robust for estimating contrasts $C_\alpha v_\alpha$ in the direct effects, under final period dropout with fixed probability θ, is the design with maximum value of $E[X_d|\theta]$ and minimum $\mathrm{Var}[X_d|\theta]$. Similarly, the design most robust for estimating contrasts $C_\lambda v_\lambda$ in the first-order carry-over treatment effects is that with maximum value of $E[Y_d|\theta]$ and minimum $\mathrm{Var}[Y_d|\theta]$.

As discussed in Section 2.4, under certain patterns of final period dropout, some implemented designs are disconnected. An important criterion for robustness is that the probability of the implemented design being disconnected is zero or, at worst, acceptably small to the experimenter.

The task of obtaining an appropriate design can now be considered to be that

of obtaining a design which satisfies the above conditions. Hence, we propose the following design selection criteria.

2.8.1 Design Selection Criteria

Given a fixed probability θ of final period dropout, select a design d^* for which:

(i) $P(D_0^*)$ is zero or very close to zero, where D_0^* is the set of disconnected implementable designs derived from d^*,

(ii) $E[X_{d^*}|\theta]$ is as large as possible with minimum $\text{Var}[X_{d^*}|\theta]$ provided (i) is satisfied, and

(iii) $E[Y_{d^*}|\theta]$ is as large as possible with minimum $\text{Var}[Y_{d^*}|\theta]$ provided (i) is satisfied.

Usually the estimation of the direct treatment effects is more important than the carry-over effects. In these circumstances it may be more appropriate either to disregard part (iii) or allow a decrease in $E[Y_{d^*}|\theta]$ and/or an increase in $\text{Var}[Y_{d^*}|\theta]$ in order to achieve an increase in $E[X_{d^*}|\theta]$ and/or a decrease in $\text{Var}[X_{d^*}|\theta]$.

The use of these design selection criteria is now illustrated in the following example which compares the relative performance of two different designs built from Williams squares of side four.

Example 2.5 We wish to compare the performance of the following planned designs under repeated use in cross-over trials with some probability, θ, of final period dropout.

Design (a) Single Williams square $d(4,4,4,4)$ with treatment labels 0, 1, 2 and 3 and initial treatment sequence (0 1 3 2). This is the design considered in Examples 2.3 and 2.4.

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1

Note: All Williams squares of side four obtained under some permutation of the treatment labels are isomorphic. However, of the 24 isomorphic squares there are six different squares, if the order of the treatment sequences is unimportant.

Design (b) Complementary pair of Williams squares $d(4, 8, 2, 4)$ with treatment labels 0, 1, 2 and 3 and initial sequences (0 1 3 2) and (0 3 1 2).

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1
0	3	1	2
1	0	2	3
2	1	3	0
3	2	0	1

Note: For each of the six possible arrangements of a Williams square of side four each square has only one complement i.e. there are three complementary pairs.

The second square complements the first in the sense of not replicating any of the ordered pairs of treatments in the third and fourth periods found in the first square. In design (b) there are eight distinct ordered pairs of treatments in the final two periods, while in design (a) there are only four. Design (b), therefore, achieves a better spread of treatment pairs in the final two periods than design (a). Designs formed by replicating the sequences of a complementary pair of Williams squares of side four are discussed in more detail in Chapter 4.

Assume that we wish to compare all the pairwise direct and first-order carry-over treatment effects. Using equations (2.4), (2.5), (2.6) and (2.7) with the A-criterion we can obtain a summary of the average variance of the direct and first-order carry-over treatment effects for each design. Summary measures for designs (a) and (b) are given in Table 2.4 and Table 2.6, respectively.

Table 2.6: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for a complementary pair of Williams squares with 16 subjects, $d(4, 8, 2, 4)$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.92	0.07	4.69	0.06
0.2	6.54	0.15	4.37	0.11
0.3	6.16	0.22	4.04	0.16
0.4	5.75	0.28	3.70	0.20
0.5	5.32	0.34	3.34	0.23
0.6	4.86	0.36	2.98	0.23
0.7	4.39	0.35	2.60	0.21
0.8	3.90	0.27	2.23	0.16
0.9	3.45	0.12	1.89	0.06
1.0	3.20	0.00	1.71	0.00

Comparisons of the graphs of the mean of X_d and Y_d against θ, for $0 \leq \theta \leq 1$, for designs (a) and (b), are given in Figures 2.5 and 2.6 respectively. From these we observe that the range of values for the mean of X_d or Y_d as θ varies is not as large for design (b) as for design (a). For design (a) the ranges are 7.27-0.00 and 5.00-0.00, respectively, while for design (b) they are 7.27-3.20 and 5.00-1.71, respectively. Consequently, design (b) is less sensitive to the choice of θ than design (a).

Using the design selection criteria of Section 2.8.1 with the A-criterion leads to the choice of design (b) for any probability of final period dropout $0 \leq \theta \leq 1$. This is because (b) does not produce any disconnected implementable designs, unlike design (a). Examining the mean performance measures given in Tables 2.4 and 2.6 and illustrated in Figures 2.5 and 2.6 shows that

$$\mathrm{E}[X_{d_b}|\theta] \geq \mathrm{E}[X_{d_a}|\theta] \text{ and } \mathrm{E}[Y_{d_b}|\theta] \geq \mathrm{E}[Y_{d_a}|\theta]$$

for each probability of final period dropout θ, given in the tables.

Similarly, if we compare the respective variances for X_d and Y_d for each design from Tables 2.4 and 2.6 we observe that

$$\mathrm{Var}[X_{d_b}|\theta] \leq \mathrm{Var}[X_{d_a}|\theta] \text{ and } \mathrm{Var}[Y_{d_b}|\theta] \leq \mathrm{Var}[Y_{d_a}|\theta]$$

for each probability of final period dropout θ, given in the tables.

Alternatively, we could compare the relative performance of each design using the summary measures provided by equations (2.4), (2.5), (2.6) and (2.7) with the MV-criterion. Summary measures for designs (a) and (b) under the MV-criterion are given in Tables 2.5 and 2.7, respectively.

Comparisons of the graphs of the mean of X_d and Y_d against θ, for $0 \leq \theta \leq 1$, for designs (a) and (b) are given in Figures 2.7 and 2.8 respectively.

Using the design selection criteria of Section 2.8.1 we would again select design (b). This is because, as before, it does not produce any disconnected implementable designs, unlike design (a). Also the mean values of X_d and Y_d are always larger for

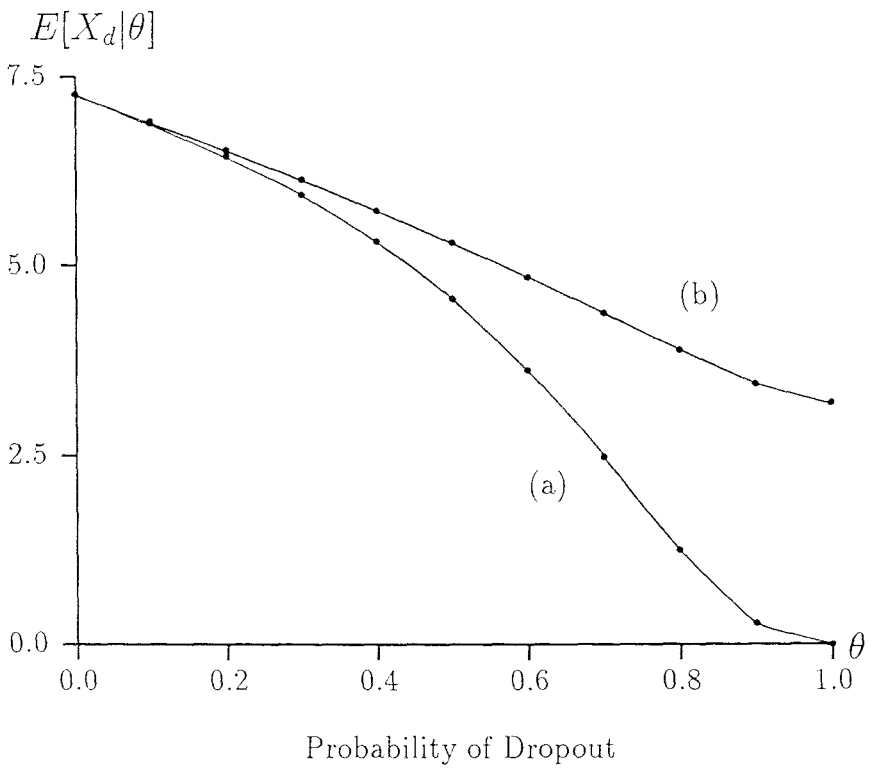


Figure 2.5: Comparison of the graphs showing the mean of X_d, under the A-criterion, for designs (a) and (b).

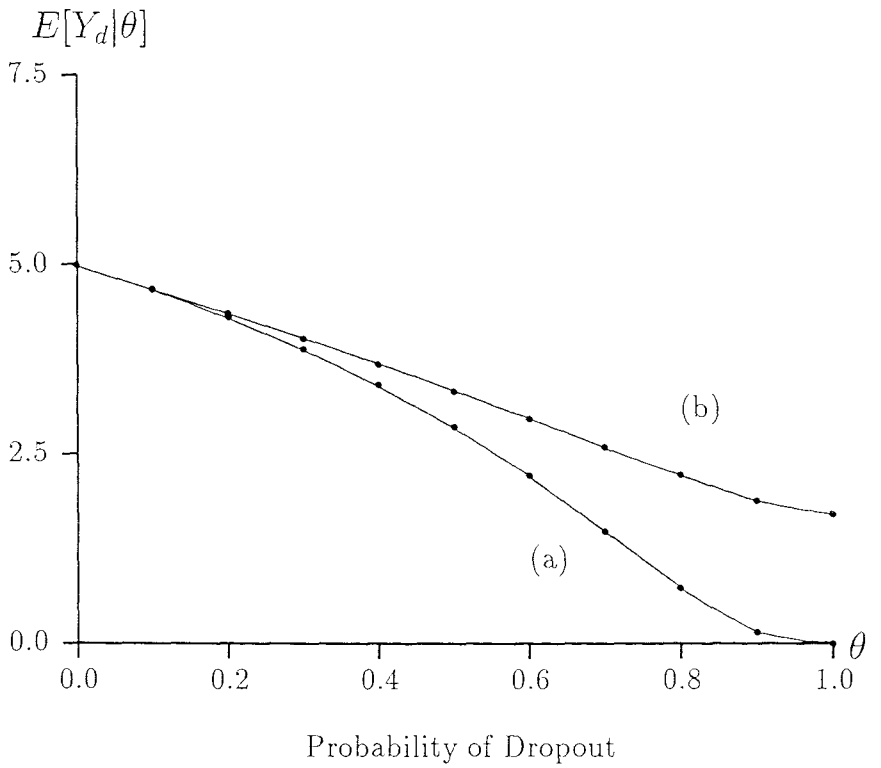


Figure 2.6: Comparison of the graphs showing the mean of Y_d, under the A-criterion, for designs (a) and (b).

Table 2.7: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for a complementary pair of Williams squares with 16 subjects, $d(4, 8, 2, 4)$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.63	0.20	4.47	0.14
0.2	6.11	0.30	4.02	0.21
0.3	5.62	0.37	3.60	0.26
0.4	5.14	0.43	3.19	0.29
0.5	4.66	0.46	2.80	0.29
0.6	4.17	0.46	2.41	0.27
0.7	3.68	0.42	2.03	0.22
0.8	3.21	0.29	1.69	0.14
0.9	2.82	0.10	1.43	0.04
1.0	2.67	0.00	1.33	0.00

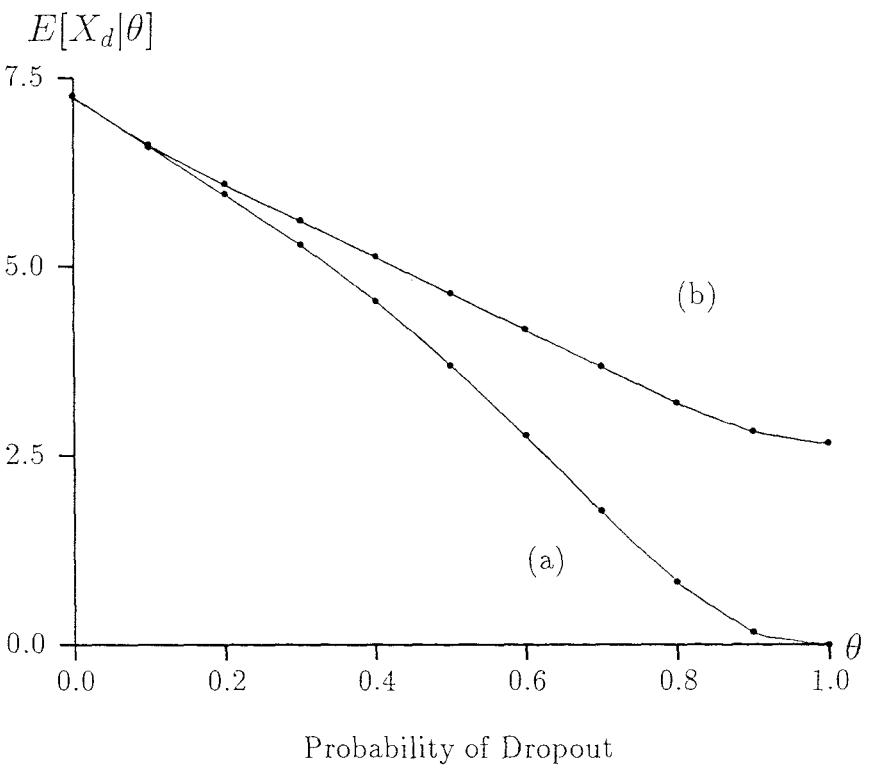


Figure 2.7: Comparison of the graphs showing the mean of X_d, under the MV-criterion, for designs (a) and (b).

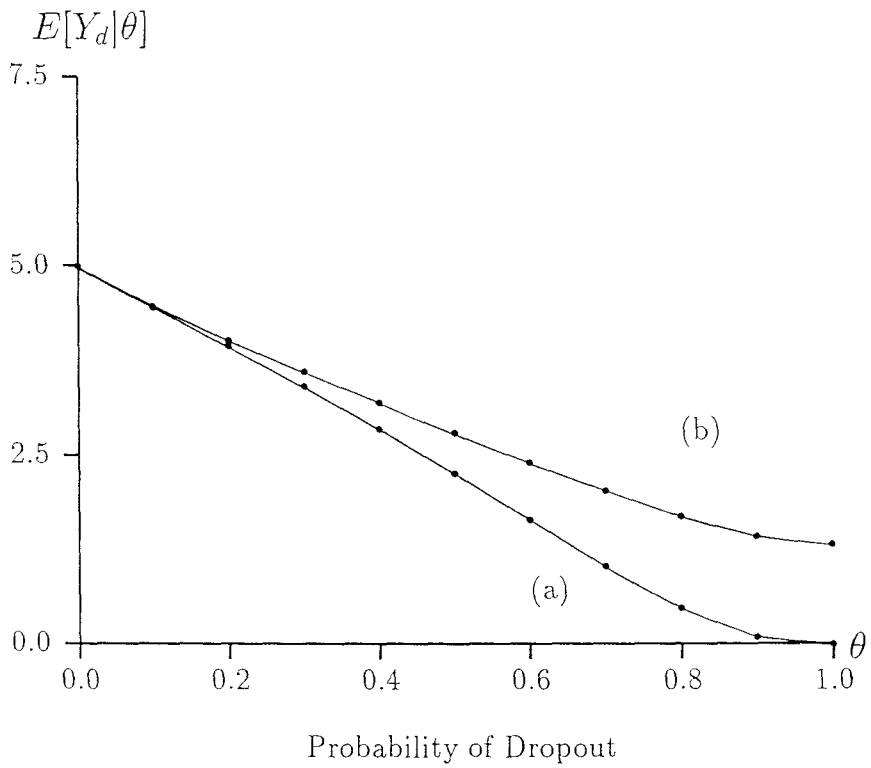


Figure 2.8: Comparison of the graphs showing the mean of Y_d, under the MV-criterion, for designs (a) and (b).

design (b), with corresponding smaller variances, for each probability of final period dropout considered.

To summarise, the advantages of selecting design (b) rather than design (a) are that the probability of producing a disconnected implemented design is eliminated while the mean values of the performance measures X_d and Y_d are increased, for all values of $\theta > 0$, with a corresponding reduction in their variances. This is true whether we use the A- or MV-criterion when calculating the performance measures X_d and Y_d.

If the probability of final period dropout had not been considered when comparing the relative performance of these planned designs, the conclusion reached would have been that there was no difference between the designs because the planned designs possess the same performance measures for the estimation of both the direct and first-order carry-over treatment comparisons. This is true whether the A- or MV-criterion is employed.

2.9 Discussion

In this chapter methods for the assessment and selection of cross-over designs in the presence of final period dropout have been presented. These methods have been illustrated using examples involving Williams squares of side four. This design was chosen because it is of particular practical importance. It is widely used in experiments because, when the possibility of dropouts is ignored, it is known to be universally optimal, over the class of uniform designs in which $t = p$, for estimating the direct and first-order carry-over treatment effects under the simple carry-over model (1.1). In Example 2.5 we have shown that, by carefully combining two "different" Williams squares of side four, it is possible to produce a design which is more robust to final period dropout than two copies of a single Williams square. These results suggest the following questions: How do different designs perform subject

to the assessment criteria? Which designs are robust to final period dropout? In Chapters 4 and 5 these questions are considered for cross-over designs in which $t = p$ involving four and three treatments respectively.

The task of identifying an *optimal* design under the selection criteria given in Section 2.8.1 is difficult. This is because it is necessary to obtain performance measures for each of the implementable designs in D. The set of implementable designs is large, even for relatively small studies, and so the computational problems involved are considerable. For example, a design involving 24 subjects in an eight sequence design will give rise to 65,536 different implementable designs. It is for this reason that all the examples presented in this chapter have involved fewer subjects than would usually be used in practice. Fortunately, the size of the computational problem can be reduced using results from combinatorial theory. In Chapter 3 these combinatorial results are presented together with examples illustrating their role in reducing the size of the computational problem.

Chapter 3

Computational Reductions

3.1 Introduction

To assess the performance of a cross-over design $d(t, m, n, p)$ using the criteria proposed in Chapter 2 it is necessary to obtain performance measures for each of the connected implementable designs in D. The number of implementable designs may be large even for relatively small studies. Hence, there are considerable computational difficulties involved in evaluating designs using this approach. In this chapter ways of reducing the amount of computation required are presented, using results from combinatorial theory.

3.2 Equivalence

In order to reduce the computation involved it is necessary to consider which designs in D give rise to the same performance measures under (2.2) and (2.3). Any designs which do this are equivalent in the following sense:

Definition 3.1 Consider the planned design $d(t, m, n, p)$ and its associated set of implementable designs D. Designs d_{l_1}, $d_{l_2} \in D \backslash D_0$ are **performance equivalent** with respect to direct and first-order carryover effects if and only if

$$[\psi(C_\alpha \Omega_\alpha(d_{l_1}) C'_\alpha)]^{-1} = [\psi(C_\alpha \Omega_\alpha(d_{l_2}) C'_\alpha)]^{-1}$$

and

$$[\psi(C_\lambda \Omega_\lambda(d_{l_1}) C'_\lambda)]^{-1} = [\psi(C_\lambda \Omega_\lambda(d_{l_2}) C'_\lambda)]^{-1},$$

where ψ, C_α, C_λ, $\Omega_\alpha(d_l)$ and $\Omega_\lambda(d_l)$ are as in Definitions 2.4 and 2.5.

Remark: This definition can be extended to include performance equivalence with respect to higher order carry-over effects, see Section 1.2.3, if these are present in the model.

If it can be established which of the implementable designs are performance equivalent, then D can be partitioned into equivalence classes and it will be unnecessary to calculate performance measures for every implementable design. Instead, it will be sufficient to obtain performance measures for one member from each equivalence class and then combine these to obtain summary measures for the planned design by multiplying each of the performance measures by the size of the corresponding equivalence class.

In order to do this we seek answers to the following questions. Given a planned design $d(t, m, n, p)$ with an associated set of implementable designs, D, of size $(n+1)^m$,

(i) how can we identify the performance equivalent designs without calculating the performance measures,

(ii) how many equivalence classes are there,

(iii) what is the size of each equivalence class, and

(iv) how can we identify one member from each class?

This set of questions is analogous to those considered by a particular colouring problem found in combinatorics. In the next section the colouring problem is reviewed and in Section 3.4 the theory of the colouring problem is used to provide answers to questions (ii)-(iv) above. In the remainder of this section we define a combinatorial equivalence relationship between designs which is sufficient to establish performance equivalent designs.

In order that real reductions in computational effort can be achieved by using the techniques developed for the colouring problem (see Section 3.3), it is necessary to establish conditions for equivalence which do not require the performance measures of the implementable designs to be calculated. Cross-over designs which can be shown to be the same under some permutation of either the treatment labels or the treatment sequences are regarded as equivalent in the sense that their design properties are identical with respect to the estimation of treatment comparisons. Using this process of establishing equivalence via permutations of the treatment labels and/or the treatment sequences, we now define a combinatorial equivalence relationship between designs which is sufficient to establish performance equivalent designs and has the computational advantage that its equivalence classes can be found without calculating the performance measures of the designs.

Definition 3.2 Consider the planned design d, and its associated set of implementable designs D. Designs $d_{l_1}, d_{l_2} \in D$ are **combinatorially equivalent** if d_{l_2} can be obtained from d_{l_1} by permuting firstly the order of the treatment sequences and, secondly, the treatment labels.

To illustrate combinatorial equivalence we consider the following examples.

Example 3.1 Let $d(2,2,1,3)$ be the planned design with treatment labels A and B listed, together with its associated set of implementable designs, in Figure 3.1.

We wish to establish whether the implementable designs d_{10} and d_{01} are combinatorially equivalent. They will be combinatorially equivalent if there exists some

Figure 3.1: Full set of implementable designs for Example 3.1.

d_{00}			d_{10}			d_{01}			d_{11}		
A	B	B	A	B	$*$	A	B	B	A	B	$*$
B	A	A	B	A	A	B	A	$*$	B	A	$*$

permutation of the treatment labels and/or treatment sequences which when applied to design d_{10} gives design d_{01}. A permutation of just the treatment labels will not achieve this, since all such permutations will leave the position of the dropout unchanged. To obtain a design in which the dropout occurs on a different treatment sequence it is necessary to permute the order of the treatment sequences. For this design only one permutation of the treatment sequences, other than the identity, exists. This is the permutation which reverses the order of the treatment sequences. If we let treatment sequence i be defined as r_i then this operation is defined, in cycle notation, by the permutation $(r_1 r_2)$.

If we apply the permutation $(r_1 r_2)$ to the implementable design d_{10} we obtain the following.

$$\begin{array}{ccc} A & B & * \\ B & A & A \end{array} \quad \xrightarrow{(r_1 r_2)} \quad \begin{array}{ccc} B & A & A \\ A & B & * \end{array}$$

Thus the application of this permutation does not give d_{01}. However, if we now relabel the treatments using the permutation (AB), defined on the treatment labels, it is possible to obtain the design d_{01} as follows

$$\begin{array}{ccc} B & A & A \\ A & B & * \end{array} \quad \xrightarrow{(AB)} \quad \begin{array}{ccc} A & B & B \\ B & A & * \end{array}$$

Therefore, we conclude that the implementable designs d_{10} and d_{01} are combinatorially equivalent.

A permutation of the treatment sequences will not necessarily act on an implementable design d_{l_1}, where $d_{l_1} \in D$, to give a design which can be shown to

be equivalent, under some relabelling of the treatments, to another implementable design $d_{l_2} \in D$. This is illustrated in the following example.

Example 3.2 Let $d(2,2,1,3)$ be the planned design with treatment labels A and B listed, together with its associated set of implementable designs, in Figure 3.2.

Figure 3.2: Full set of implementable designs for Example 3.2.

d_{00}			d_{10}			d_{01}			d_{11}		
A	B	A	A	B	$*$	A	B	A	A	B	$*$
B	A	A	B	A	A	B	A	$*$	B	A	$*$

We wish to establish whether the designs d_{10} and d_{01}, in Figure 3.2, are combinatorially equivalent. Taking d_{10} and permuting the order of the treatment sequences we obtain the following

$$\begin{array}{ccc} A & B & * \\ B & A & A \end{array} \quad \xrightarrow{(r_1 r_2)} \quad \begin{array}{ccc} B & A & A \\ A & B & * \end{array}$$

In this case it is not possible to act on the treatment labels of the resultant design to obtain the implementable design d_{01}. Therefore, we conclude that, in this case, the implementable designs d_{10} and d_{01} are not combinatorially equivalent.

In Examples 3.1 and 3.2 it has been shown that, by considering appropriate permutations of the treatment sequences and treatment labels, it is possible to determine which, if any, of the implementable designs $d_l \in D$ are combinatorially equivalent and hence performance equivalent, without having to calculate their respective performance measures. Note that if two, or more, designs are combinatorially equivalent they will always be performance equivalent. However, it is possible for designs to be performance equivalent without being combinatorially equivalent.

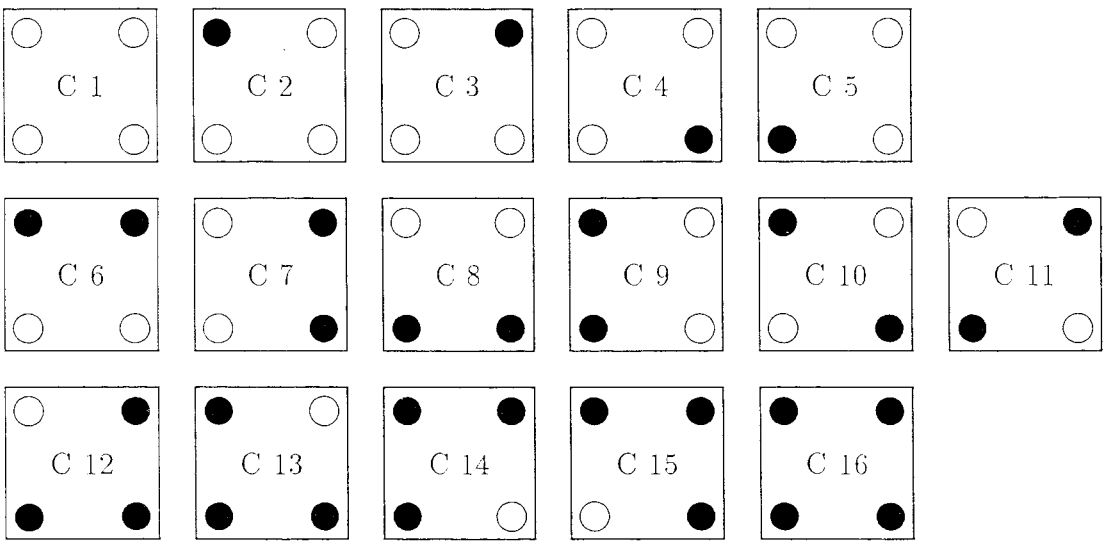
3.3 Review of the Colouring Problem

To illustrate the colouring problem we shall consider the following example.

Example 3.3 Suppose we have a square tray and we seek to place a coloured disc at each of the four corners. If each disc may be either black or white, how many different patterns can we form?

Since there is a choice of two colours for each disc, and each tray contains four discs, the total number of colourings, provided the tray remains fixed, is given by $2^4 = 16$. These 16 colourings are illustrated in Figure 3.3.

Figure 3.3: Sixteen colourings of a square tray using black and white discs.



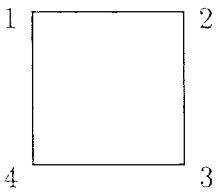
Having found the total number of possible colourings, it is necessary to establish whether or not any of these are equivalent. If the tray must remain fixed all 16 colourings are different. If, however, the tray may be rotated clockwise then some of these can be shown to be equivalent. For example, C2 is equivalent to C3, since we can obtain C3 from C2 by rotating the tray through an angle of $\pi/2$.

In this example, there are four distinct rotational symmetries of the tray, namely the identity and rotations through angles of $\pi/2$, π and $3\pi/2$. If we label the four

vertices 1, 2, 3, and 4, as in Figure 3.4, then the action of the group of rotations on the four vertices can be represented by the following permutations, written in cycle notation.

$$\begin{aligned} g_1 &= (1)(2)(3)(4), \text{ the identity} \\ g_2 &= (1234) \\ g_3 &= (13)(24) \\ g_4 &= (1432) \end{aligned}$$

Figure 3.4: Square tray with vertices labelled 1, 2, 3 and 4.



The action of this group permutes the vertices 1, 2, 3 and 4, while the action we are interested in rotates the 16 colourings $C1, \ldots, C16$. Each of these colourings can be regarded as a mapping from the set of vertices $\{1, 2, 3, 4\}$ to the set of colours $\{black, white\}$ as follows

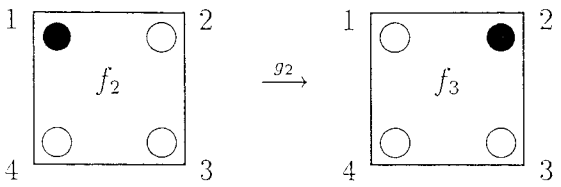
$$f : \{1, 2, 3, 4\} \longrightarrow \{black, white\}\,.$$

For example, if we let Ci be represented by the mapping f_i $(i = 1, \ldots, 16)$ then C2, in Figure 3.1 corresponds to the mapping f_2 given by

$$f_2(1) = black, \quad f_2(2) = f_2(3) = f_2(4) = white.$$

In this way the set of colourings, $\{C1, \ldots, C16\}$, can be represented by the set of mappings, $\{f_1, \ldots, f_{16}\}$.

It is now possible to apply the group of permutations to the set of colourings by considering the effect of each permutation upon each mapping. For example, the permutation g_2 transforms f_2 into f_3 as follows.



It is conventional, however, to denote the set of mappings in terms of their corresponding colourings and to consider the colouring as representing the relevant mapping.

Applying the set of permutations to the colourings in this way we observe that there are only six different patterns amongst the 16 different colourings. We can divide the colourings into the following sets, where all the colourings belonging to the same set have the same pattern.

1. {C1}
2. {C2, C3, C4, C5}
3. {C6, C7, C8, C9}
4. {C10, C11}
5. {C12, C13, C14, C15}
6. {C16}

The underlying theory of the colouring problem concerns the relationship between a group and the members of a set. In Example 3.3 the group is the rotational symmetries of a square and the 16 colourings form the set of mappings of interest. In this example, the symmetries of the square interact with each of the colourings to produce another, possibly different, colouring. In general this situation can be described as one in which an element from a group G acts on any member of a set D to give another, not necessarily different, member of the set D.

In order to explain the general theory behind the colouring problem and apply it to evaluating the number and type of equivalence classes in any given set it is necessary to recall some general results from group theory. Further details are given in Slomson (1991) and Cohen (1978).

We adopt the following notation for two operations:

Notation If D is a set and G any group then for $g_1,\ g_2 \in G$ and $d \in D$ we denote the product of the elements g_1 and g_2 under the group operation by $g_1 g_2$ and the action of the element g_1 on the set member d by $g_1 \cdot d$.

Definition 3.3 Let D be a set and G any group. We say that G **acts on** D if for each $d \in D$ and each $g \in G$, there exists an element $g \cdot d \in D$ which has the following properties.

1. For each $d \in D,\ I_G \cdot d = d$, where I_G represents the identity of G.

2. For all $g_i,\ g_j \in G$, and each $d \in D$

$$g_i \cdot (g_j \cdot d) = (g_i g_j) \cdot d.$$

The situation in which a group acts on a set is called a **group action** and the properties given in Definition 3.3 are called the **axioms for a group action.**

Definition 3.4 Let G be a group which acts on the set D. For each $g \in G$ and all $d_i \in D$ if $g \cdot d_i = d_i$ then d_i is a **fixed point** of g and the set

$$Fix(g) = \{d_i \in D\ ;\ g \cdot d_i = d_i\}$$

is called the **fixed point set** of g.

Example 3.4 Suppose we wish to obtain the fixed point set for each $g_i \in G$ for the colouring problem of Example 3.3. In this case, G is the group of rotational

symmetries of a square, given previously, and D is the set of 16 colourings given in Figure 3.3.

Using Definition 3.4, the fixed point set of g_1 is given by

$$\begin{aligned} Fix(g_1) &= \{d_i \in D \; ; \; g_1 \cdot d_i = d_i\} \\ &= \{C1, \; C2, \; C3, \; C4, \; C5, \; C6, \; C7, \; C8, \\ &\quad C9, \; C10, \; C11, \; C12, \; C13, \; C14, \; C15, \; C16\} \end{aligned}$$

Similarly,

$$\begin{aligned} Fix(g_2) &= \{C1, \; C16\} \\ Fix(g_3) &= \{C1, \; C10, \; C11, \; C16\} \\ Fix(g_4) &= \{C1, \; C16\} \end{aligned}$$

Definition 3.5 Let G be a group which acts on the set D. We can define a relation $\sim_G$ on D such that

$$\text{for all } d_i, d_j \in D, \; d_i \sim_G d_j \text{ if and only if for some } g \in G, \; g \cdot d_i = d_j,$$

where $\sim_G$ is an equivalence relation on D.

The two elements d_i and d_j are said to be **equivalent** with respect to this equivalence relation. The relationship of equivalence splits D into disjoint equivalence classes which are frequently referred to as the **orbits of the group action**.

Definition 3.6 If $d_i \in D$ then the orbit of the element d_i, denoted by $Orb(d_i)$, is defined by

$$Orb(d_i) = \{d_j \in D \; ; \; d_j = g \cdot d_i \; , \; g \in G\}$$

As the orbits represent the equivalence classes which partition the set D then the question posed in Example 3.3, namely how many different patterns can we form by placing black and white discs at the corners of a square tray, can be answered by finding the number of different orbits contained in the set of 16 colourings. The following result enables the number of orbits to be determined.

Theorem 3.1 (Burnsides's Lemma) If G is a finite group which acts on the finite set D, then the number of distinct orbits is given by

$$\frac{1}{|G|}\sum_{g\in G}|Fix(g)|.$$

In order to apply Burnsides's Lemma to Example 3.3 it is necessary to obtain $|Fix(g)|$ for each element $g \in G$. Now, from Example 3.4, we have that g_1, the identity, fixes all 16 colourings, g_2 fixes $C1$ and $C16$, g_3 fixes $C1$, $C10$, $C11$ and $C16$ and g_4 fixes $C1$ and $C16$. Therefore, applying Burnside's Lemma we find that the total number of distinct orbits is given by $(16 + 2 + 4 + 2)/4 = 6$.

In this example there are only two choices of coloured discs available. Hence, the total number of possible colourings is small enough to list individually. When three or more colours of disc are available, however, it will be impracticable to list all the possible colourings. Nevertheless, it is still possible to find the total number of orbits by using a generalised version of Burnside's Lemma. Also, by applying a theorem developed by Polya, it is possible to obtain an algebraic expression, known as the pattern inventory, which identifies the number of orbits with each possible combination of coloured discs. Before stating these theorems the following definitions and notation are needed:

Definition 3.7 A cycle $(x_1x_2\ldots x_n)$ has length n and is called an **n-cycle**. We denote an n-cycle by s_n.

Notation If G is a group of permutations, for any element $g \in G$ let $N(g)$ denote the total number of cycles in g. For example, if $g = (12)(34)$ then g consists of two cycles each of length two.

Definition 3.8 For any permutation $g \in G$, written in cycle notation, the **cycle monomial** is the single term

$$s_1^{j_1(g)}s_2^{j_2(g)}s_3^{j_3(g)}\ldots,$$

where s_n is an n-cycle and $j_n(g)$ is the number of n-cycles in the cycle representation of g.

Example 3.5 The element (12)(34) consists of two cycles each of length two and therefore has cycle monomial s_2^2; the element (1)(23)(4) consists of two cycles of length one and one cycle of length two and therefores has cycle monomial $s_1^2 s_2$.

Definition 3.9 If G is a group of permutations, the **cycle index** of G, denoted by Z_G, is defined to be the polynomial

$$\frac{1}{|G|}\sum_{g\in G} s_1^{j_1(g)} s_2^{j_2(g)} s_3^{j_3(g)} \ldots .$$

Definition 3.10 For any set C a **weight function** on the set C is a mapping ω which assigns an algebraic symbol $\omega(c)$ to each $c \in C$.

Theorem 3.2 (Burnside's Lemma -generalised) If G is a finite group which acts on the finite set D and C denotes the set of colours, then the number of distinct orbits is given by

$$\frac{1}{|G|}\sum_{g\in G} |C|^{N(g)}.$$

In order to apply Theorem 3.2 to Example 3.3 it is necessary to obtain the total number of cycles, $N(g)$, for each element $g \in G$. Now g_1, the identity, consists of four cycles, g_2 one cycle, g_3 two cycles and g_4 one cycle. Therefore, applying Theorem 3.2, we find that the total number of distinct orbits can be represented by

$$\frac{1}{|G|}\sum_{g\in G} |C|^{N(g)} = \frac{1}{4}(c^4 + c^2 + 2c), \tag{3.1}$$

where $c = |C|$.

In Example 3.3 the number of colours available is two, namely black and white. Substituting $c = 2$ into equation (3.1) we find that the number of distinct orbits is $(1/4)(2^4 + 2^2 + 4) = 6$, which agrees with our previous findings.

Equation (3.1) can be used to calculate the number of different patterns, or orbits, for examples using a range of different values of c suitable for application to our design problem, see Section 3.4. This is shown in Table 3.1.

Table 3.1: The number of colourings and orbits that can be formed from a square tray and a range of different coloured discs.

Number of colours.	Number of colourings.	Number of orbits.
1	1	1
2	16	6
3	81	24
4	256	70
5	625	165
6	1296	336
7	2401	616
8	4096	1044
9	6561	1665
10	10000	2530
11	14641	3696
12	20736	5226

The following theorem allows the pattern inventory, which gives the number of orbits with each possible combination of coloured discs, to be determined.

Theorem 3.3 (Polya's Theorem) Let D be the set of all mappings from the set V to the set C, and let ω be a weight function on C. Let G be a group of permutations of V which acts on D. If the cycle index of G is $Z_G(s_1, s_2, \ldots)$, then the pattern inventory is

$$Z_G\left(\sum_{c\in C}\omega(c), \sum_{c\in C}\omega(c)^2, \sum_{c\in C}\omega(c)^3, \ldots\right).$$

The application of Theorem 3.3 means that the pattern inventory is obtained as follows.

1. Find the cycle index for the action G on D.
2. Replace each s_1 in the cycle index by the sum of the weights, each s_2 by the sum of the squares of the weights, each s_3 by the sum of the cubes of the weights, and so on.
3. Expand the resultant expression to obtain a polynomial in the weight functions.

The pattern inventory is a polynomial in terms of the weight functions. Each term of the polynomial denotes a particular type of pattern, or combination of the weight functions. For example, for the two weights a and b the term ab^3 denotes a pattern containing one element of type a and three elements of type b. The coefficient of each term denotes the number of patterns, or orbits, which contain the particular combination of elements it describes. For example, the term $2ab^3$ would represent two distinct patterns, or orbits, containing one element of type a and three elements of type b.

The following example illustrates the use of Polya's Theorem (Theorem 3.3).

Example 3.6 Suppose we wish to obtain the pattern inventory for the colouring problem of Example 3.3.

In this case V is the set of vertices $\{1, 2, 3, 4\}$ and C is the set of colours $\{black,\ white\}$. The weight function ω is given by

$$\omega(black) = b, \quad \omega(white) = w.$$

The group G is the group of permutations of V which correspond to the rotational symmetries of a square. That is

$$G = \{(1)(2)(3)(4),\ (1234),\ (13)(24),\ (1432)\}.$$

The cycle index of G is given by

$$Z_G(s_1, s_2, s_4) = \frac{1}{4}(s_1^4 + s_2^2 + 2s_4). \tag{3.2}$$

Applying Polya's Theorem (Theorem 3.3), the pattern inventory is obtained by substituting $s_1 = b+w$, $s_2 = b^2+w^2$ and $s_4 = b^4+w^4$ into equation (3.2). Therefore, the pattern inventory is

$$\frac{1}{4}[(b+w)^4 + (b^2+w^2)^2 + 2(b^4+w^4)] = b^4 + b^3w + 2b^2w^2 + bw^3 + w^4. \tag{3.3}$$

Examining each of the terms in the pattern inventory, (3.3), we find that, for this example, there exists:

one pattern with four black discs,

one pattern with three black and one white disc,

two patterns with two white and two black discs,

one pattern with one black and three white discs, and

one pattern with four white discs.

This agrees with our previous findings, in Figure 3.3. Note that if we sum the coefficients of each of the terms in the pattern inventory we will always obtain the total number of orbits, since this polynomial is an algebraic representation of each of the distinct orbits.

In addition to finding the number of orbits, or equivalence classes, the following theorem enables the number of colourings belonging to the same equivalence class as any given colouring to be determined. Hence, the size of each equivalence class can be found without listing all the individual colourings.

Definition 3.11 Let G be a finite group which acts on a set D. If $d \in D$ then the **stabilizer** of d, denoted by $Sta(d)$, is the set of all permutations $g \in G$ which leave d fixed. That is

$$Sta(d) = \{g \in G; g \cdot d = d\}.$$

Theorem 3.4 (The Orbit-Stabilizer Theorem) Let G be a finite group which acts on the finite set D. Then for each $d \in D$ the number of elements in $Orb(d)$ multiplied by the number of elements in $Sta(d)$ is equal to the number of elements in the group G. That is

$$|Orb(d)| \times |Sta(d)| = |G|.$$

Corollary 3.1 The number of elements in $Orb(d)$, that is the size of the equivalence class containing d, is given by

$$|Orb(d)| = \frac{|G|}{|Sta(d)|}.$$

In the following example the corollary to the Orbit-Stabilizer Theorem is used to establish the size of each equivalence class formed from the 16 colourings of a square tray given in Figure 3.3.

Example 3.7 Suppose we wish to calculate the size of each orbit for the colouring problem of Example 3.3. It has already been established that there are 16 different colourings which can be partitioned into six equivalence classes. Given that C1, C2, C6, C10, C12 and C16, from Figure 3.3, all belong to different equivalence classes we wish to establish the size of their respective equivalence classes.

In order to apply Corollary 3.1 we require the elements of the group G and the stabilizers of each colouring. The group G is the set of rotational symmetries of a square given previously. The stabilizers of each of the colourings C1, C2, C6, C10, C12 and C16 are given in Table 3.2.

Applying Corollary 3.1 we find that the equivalence classes containing C1 and C16 have one member each, C2, C6 and C12 have four members each and C10 has two members.

Table 3.2: Description of the stabilizers for the six different patterns given in Example 3.7.

Colouring	Stabilizer	$\|Sta\|$
$C1$	$\{g_1, g_2, g_3, g_4\}$	4
$C2$	$\{g_1\}$	1
$C6$	$\{g_1\}$	1
$C10$	$\{g_1, g_3\}$	2
$C12$	$\{g_1\}$	1
$C16$	$\{g_1, g_2, g_3, g_4\}$	4

3.4 The Colouring Problem and Cross-over Designs Subject to Final Period Dropout

In this section techniques for obtaining the number of equivalence classes contained within a set, D, of implementable designs and their respective sizes are established using results developed from the theory of the colouring problem.

In order to apply the theory of the colouring problem, outlined in Section 3.3, to our design problem it is necessary to define a group, G, which will act on the set of implementable designs in the same way as the rotational symmetries of a square acts on the colourings given in Figure 3.3. Each of these rotational symmetries acts on the empty square tray, that is the tray without any coloured discs, to give an image which is identical to itself. Therefore, the group we require will consist only of those elements which, when they act on the planned design, will transform it into itself. Also, in the colouring problem, since the coloured discs are to be placed at the corners of the tray, each of the rotational symmetries is defined in terms of its action on the set of vertices. This enables each of the colourings to be represented by a mapping from the set of vertices to the set of colours. For our design problem to be similar, since the dropouts occur in the final period of each treatment sequence,

we require the group G to consist of elements defined in terms of an action on the distinct treatment sequences and each implementable design to be represented by a mapping from the set of distinct treatment sequences to the set of final period dropouts.

Definition 3.12 Let the planned design $d(t, m, n, p)$ be defined as the set R of m distinct treatment sequences $r_1, r_2, \ldots, r_m$ where each of these treatment sequences consists of a different arrangement of treatment labels.

Definition 3.13 Let the set of **dropout identifiers** be the set C whose elements are all the possible dropout totals which can occur in the final period of each distinct treatment sequence of the planned design. Then $C = \{0, 1, \ldots, n\}$, where n is the number of subjects allocated to each distinct treatment sequence of the planned design.

Using this notation each implementable design can be represented by a mapping from the set R to the set C as follows

$$f : \{r_1, r_2, \ldots, r_m\} \longrightarrow \{0, 1, \ldots, n\}.$$

For instance, in Example 3.1 the mapping can be defined as follows

$$f : \{r_1, r_2\} \longrightarrow \{0, 1\}.$$

If we let the implementable design d_l be represented by the mapping f_l then the design d_{01} corresponds to the mapping f_{01} given by

$$f_{01}(r_1) = 0, \quad f_{01}(r_2) = 1.$$

As in the colouring problem, it is now possible to apply the appropriate group G to the set of implementable designs by considering the effect of each element $g \in G$ upon each mapping. This leads to the following definition of the group of permutations we require.

Definition 3.14 Let G be a permutation group which acts on the set R of distinct treatment sequences of a planned design. A permutation g, of the sequences, belongs to G if and only if the planned design can be preserved by applying g to the sequences and then relabelling the treatments.

The following example illustrates how to find the elements of the group G.

Example 3.8 Let $d(4,4,1,4)$ be the Williams square with treatment labels 0, 1, 2 and 3, given in Example 2.1.

Consider the following operations.

1. Permute the order of the treatment sequences using the permutation $(r_1r_2r_3r_4)$, then

2. relabel the treatments using the permutation of the treatment labels (0123).

If we apply these operations to the planned design we obtain the following.

$$\begin{matrix} 0 & 1 & 3 & 2 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 1 & 0 \\ 3 & 0 & 2 & 1 \end{matrix} \xrightarrow{(r_1r_2r_3r_4)} \begin{matrix} 3 & 0 & 2 & 1 \\ 0 & 1 & 3 & 2 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 1 & 0 \end{matrix} \xrightarrow{(0123)} \begin{matrix} 0 & 1 & 3 & 2 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 1 & 0 \\ 3 & 0 & 2 & 1 \end{matrix}$$

The image of these operations is the planned design. Therefore, the permutation of the treatment sequences $(r_1r_2r_3r_4)$ is an element of the group G.

Example 3.9 For the design in Example 3.8, consider the following operations.

1. Permute the order of the treatment sequences using the permutation $(r_1r_2)(r_3)(r_4)$, then

2. relabel the treatments using the permutation of the treatment labels (0321).

If we apply these operations to the planned design we obtain the following.

$$\begin{matrix} 0 & 1 & 3 & 2 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 1 & 0 \\ 3 & 0 & 2 & 1 \end{matrix} \xrightarrow{(r_1r_2)(r_3)(r_4)} \begin{matrix} 1 & 2 & 0 & 3 \\ 0 & 1 & 3 & 2 \\ 2 & 3 & 1 & 0 \\ 3 & 0 & 2 & 1 \end{matrix} \xrightarrow{(0321)} \begin{matrix} 0 & 1 & 3 & 2 \\ 3 & 0 & 2 & 1 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 1 & 0 \end{matrix}$$

Unlike the situation in Example 3.8, the image of the operations is not the planned design. In addition, no further relabelling of the treatments can possibly give the planned design. Therefore, we conclude that the permutation of the treatment labels $(r_1r_2)(r_3)(r_4)$ is not an element of the group G.

When all the elements of the group G have been identified it is possible to determine the implementable designs which will give rise to identical performance measures using the following definition of equivalence.

Definition 3.15 For the planned design $d(t, m, n, p)$ any two implementable designs $d_{l(x)}$ and $d_{l(y)} \in D$ will be equivalent if and only if

$$d_{l(y)} = g \cdot d_{l(x)} \text{ for some } g \in G.$$

If we replace the square tray by the planned design, the set of vertices by the set of distinct treatment sequences, the group of rotational symmetries by the group of permutations of the distinct treatment sequences given in Definition 3.15 and the set of colours by the set of dropout identifiers, then we can define a problem which is directly analogous to the colouring problem. Therefore, the following results can be derived directly.

Theorem 3.5 Let $d(t, m, n, p)$ be a planned design, D its set of implementable designs and G a finite permutation group of the distinct treatment sequences of the planned design which satisfies Definition 3.14. If the group G acts on the set D then the number of equivalence classes contained in D is given by

$$\frac{1}{|G|} \sum_{g \in G} |Fix(g)|.$$

The proof follows directly from Burnside's Lemma (Theorem 3.1).

Theorem 3.6 Let $d(t, m, n, p)$ be a planned design, D its set of implementable designs and G a finite group consisting of permutations of the distinct treatment sequences of the planned design which satisfies Definition 3.14. If the group G acts on the set D then the number of equivalence classes contained in D is given by

$$\frac{1}{|G|}\sum_{g\in G}(n+1)^{N(g)},$$

where n denotes the number of subjects allocated to each distinct treatment sequence and $N(g)$ the number of cycles in g.

The proof follows directly from Theorem 3.2.

Theorem 3.7 Let D be the set of implementable designs formed by dropping one or more subject in the final period of the planned design $d(t, m, n, p)$. Each design $d_l \in D$ can be represented by a set of mappings from the set of distinct treatment sequences, R, to the set of dropout identifiers, C. Let ω be a weight function on C and G be a group of permutations of R which acts on D and satisfies Definition 3.14. If the cycle index of G is $Z_G(s_1, s_2, \ldots)$, then the polynomial which determines the number of equivalence classes with each possible combination of dropout identifiers is,

$$Z_G\left(\sum_{c\in C}\omega(c), \sum_{c\in C}\omega(c)^2, \sum_{c\in C}\omega(c)^3, \ldots\right),$$

and will be referred to as the **equivalence class inventory**.

The proof follows directly from that of Polya's Theorem (Theorem 3.3).

Theorem 3.8 Let D be the set of implementable designs formed by dropping one or more subject in the final period of the planned design $d(t, m, n, p)$ and let G be a group of permutations which acts on D and satisfies Definition 3.14. Then for each

$d_l \in D$ the number of elements in $Orb(d_l)$, that is the size of the equivalence class containing d_l, is given by

$$|Orb(d_l)| = \frac{|G|}{|Sta(d_l)|}.$$

The proof follows directly from that of the Orbit-Stabilizer Theorem (Theorem 3.4).

In the next section these results are illustrated using designs of particular importance when examining the robustness of cross-over designs to final period dropout.

3.5 Illustrations

Example 3.10 Let $d(4, 4, n, 4)$ be the design formed from the sequences of a Williams square of side four having treatment labels 0, 1, 2 and 3 and initial treatment sequence (0 1 3 2) in which n subjects are allocated to each treatment sequence.

The set, R, of distinct treatment sequences consists of the following elements.

$$r_1 = 0\ 1\ 3\ 2,\ \ r_2 = 1\ 2\ 0\ 3,\ \ r_3 = 2\ 3\ 1\ 0 \text{ and } r_4 = 3\ 0\ 2\ 1$$

The group G of permutations of the distinct treatment sequences which satisfy the conditions of Definition 3.14 is as follows.

$$G = \{(r_1)(r_2)(r_3)(r_4), (r_1r_2r_3r_4), (r_1r_3)(r_2r_4), (r_1r_4r_3r_2)\}.$$

Hence,

$$|G| = 4,\ N(g_1) = 4,\ N(g_2) = 1,\ N(g_3) = 2 \text{ and } N(g_4) = 1.$$

Applying Theorem 3.6, the total number of equivalence classes is given by

$$\frac{1}{4}[(n+1)^4 + (n+1)^2 + 2(n+1)]. \tag{3.4}$$

Table 3.3 lists the number of equivalence classes, calculated from expression (3.4), for a range of values for n, the number of subjects allocated to each treatment sequence.

Table 3.3: The number of equivalence classes and implementable designs for designs based on a single Williams square of side four with up to 12 subjects per treatment sequence.

Number of subjects per treatment sequence, n.	Number of implementable designs, $\|D\|$.	Number of equivalence classes.
1	16	6
2	81	24
3	256	70
4	625	165
5	1296	336
6	2401	616
7	4096	1044
8	6561	1665
9	10000	2530
10	14641	3696
11	20736	5226
12	28561	7189

From Table 3.3 we observe that, by only having to obtain performance measures for one member from each equivalence class, it is possible to reduce substantially the amount of computation involved in evaluating the mean and variance of the performance measures for each planned design. The reduction is particularly great for large values of n. In fact, for the designs based on a single Williams square of side four given in this example, we find that the computational reduction is approximately three quarters.

In order to evaluate the performance measures for the planned design by only evaluating the individual performance measures of non-equivalent designs, it is necessary to identify one design from each of the equivalence classes and determine the size of each class. The following example outlines how this can be achieved using the equivalence class inventory defined in Theorem 3.7.

Example 3.11 Suppose we wish to find the equivalence class inventory for the Williams design $d(4,4,1,4)$ whose set of implementable designs is given in Table 2.1.

The number of subjects allocated to each treatment sequence is $n = 1$. Therefore, the set of dropout identifiers $C = \{0,1\}$ and the weight function ω is given by

$$\omega(0) = x_0, \;\; \omega(1) = x_1.$$

The group G is the group of permutations given in Example 3.10 and the cycle index of G is given by

$$Z_G(s_1, s_2, s_4) = \frac{1}{4}(s_1^4 + s_2^2 + 2s_4). \tag{3.5}$$

Applying Theorem 3.7, the equivalence class inventory is obtained by substituting $s_1 = x_0 + x_1$, $s_2 = x_0^2 + x_1^2$ and $s_4 = x_0^4 + x_1^4$ into the cycle index, equation (3.5). Hence, the equivalence class inventory is

$$\frac{1}{4}[(x_0 + x_1)^4 + (x_0^2 + x_1^2)^2 + 2(x_0^4 + x_1^4) = x_0^4 + x_0^3 x_1 + 2x_0^2 x_1^2 + x_0 x_1^3 + x_1^4.$$

Examining each of the terms in this expression we find that the design based on a single Williams square $d(4,4,1,4)$, subject to some probability of final period

dropout, has a set of implementable designs with the following inventory of equivalence classes.

1. One class of designs with four sequences each containing zero dropouts,
2. one class of designs with three complete sequences and one sequence containing one dropout,
3. two classes of designs with two complete sequences and two sequences containing one dropout each,
4. one class of designs with one complete sequence and three sequences containing one dropout each, and
5. one class of designs with four sequences each containing one dropout.

Having found the equivalence class inventory it is necessary to find one member from each equivalence class and the size of the class to which each of the designs belong. For a small set of implementable designs, as in this example, it is possible to list all the designs and establish whether they are equivalent using the conditions for equivalence given in Definition 3.15. Doing this we find that the set of 16 implementable designs consists of the following six equivalence classes.

1. $\{d_{0000}\}$
2. $\{d_{1000}, d_{0100}, d_{0010}, d_{0001}\}$
3. $\{d_{1100}, d_{0110}, d_{0011}, d_{1001}\}$
4. $\{d_{1010}, d_{0101}\}$
5. $\{d_{1110}, d_{1101}, d_{1011}, d_{0111}\}$
6. $\{d_{1111}\}$

For planned designs involving larger numbers of subjects it will be impractical to list all the implementable designs. An alternative approach, which is illustrated in the following example, is to search for the appropriate number of non-equivalent designs and find the size of the equivalence class to which each of these belongs using Theorem 3.8.

Example 3.12 Consider the Williams design $d(4,4,1,4)$ of Example 3.11. We wish to identify one member from, and the size of, each of the equivalence classes into which its set of implementable designs can be partitioned.

There are 16 designs in the set of implementable designs and from Table 3.3 we observe that these can be partitioned into six equivalence classes. From the equivalence class inventory obtained in Example 3.11, we observe that these equivalence classes fall into five distinct categories, one for each of the possible combinations of dropout identifiers. A description of the equivalence class inventory has been given in Example 3.11.

Using this information we find that the following implementable designs all belong to different equivalence classes.

$$d_{0000}\ \ d_{0001}\ \ d_{0011}\ \ d_{0101}\ \ d_{0111}\ \ d_{1111}.$$

In order to apply Theorem 3.8 and establish the size of the six equivalence classes to which each of the designs belongs it is necessary to find the stabilizers of each of these designs with respect to the group G. The group G is the group of permutations of the distinct treatment sequences given in Example 3.10. The stabilizers of each of these implementable designs and the size of their respective equivalence classes is given in Table 3.4 whose results all agree with our previous findings.

Example 3.13 Let $d(4,8,n,4)$ be the design based on a complementary pair of Williams squares with treatment labels 0, 1, 2 and 3, initial sequences (0 1 3 2) and (3 2 0 1) and n subjects allocated to each treatment sequence.

The set, R, of distinct treatment sequences consists of the following elements.

Table 3.4: Table showing the stabilizers and the size of the equivalence classes for the six non-equivalent designs given in Example 3.12.

Implementable design, d_l.	$Sta(d_l)$	$\lvert Sta(d_l)\rvert$	Size of equivalence class containing d_l.
d_{0000}	$\{g_1, g_2, g_3, g_4\}$	4	1
d_{0001}	$\{g_1\}$	1	4
d_{0011}	$\{g_1\}$	1	4
d_{0101}	$\{g_1, g_3\}$	2	2
d_{0111}	$\{g_1\}$	1	4
d_{1111}	$\{g_1, g_2, g_3, g_4\}$	4	1

$$r_1 = 0\ 1\ 3\ 2,\ \ r_2 = 1\ 2\ 0\ 3,\ \ r_3 = 2\ 3\ 1\ 0,\ \ r_4 = 3\ 0\ 2\ 1,$$

$$r_5 = 3\ 2\ 0\ 1,\ \ r_6 = 0\ 3\ 1\ 2,\ \ r_7 = 1\ 0\ 2\ 3,\ \text{and}\ r_8 = 2\ 1\ 3\ 0.$$

The group G is the group of permutations of the treatment sequences which act on the elements $r_1, \ldots, r_8$, in such a way that the planned design can be preserved by applying the permutation and then relabelling the treatments. Therefore, the group G consists of the following permutations, written in cycle notation.

$$\begin{aligned}
g_1 &= (r_1)(r_2)(r_3)(r_4)(r_5)(r_6)(r_7)(r_8)\\
g_2 &= (r_1r_2r_3r_4)(r_5r_6r_7r_8)\\
g_3 &= (r_1r_3)(r_2r_4)(r_5r_7)(r_6r_8)\\
g_4 &= (r_1r_4r_3r_2)(r_5r_8r_7r_6)\\
g_5 &= (r_1r_5)(r_2r_8)(r_3r_7)(r_4r_6)\\
g_6 &= (r_1r_6)(r_2r_5)(r_3r_8)(r_4r_7)\\
g_7 &= (r_1r_7)(r_2r_6)(r_3r_5)(r_4r_8)\\
g_8 &= (r_1r_8)(r_2r_7)(r_3r_6)(r_4r_5)
\end{aligned}$$

Hence,

$|G| = 8,\ N(g_1) = 8,\ N(g_2) = 2,\ N(g_3) = 4,\ N(g_4) = 2,$
$N(g_5) = 4,\ N(g_6) = 4,\ N(g_7) = 4$ and $N(g_8) = 4.$

Applying Theorem 3.6 we find that the total number of equivalence classes is given by

$$\frac{1}{8}[(n+1)^8 + 5(n+1)^4 + 2(n+1)^2]. \tag{3.6}$$

Using this expression it is possible to find the total number of equivalence classes for experiments using designs based on a complementary pair of Williams squares with any suitable number of subjects allocated to each of the distinct treatment sequences. The number of equivalence classes together with the total number of implementable designs which exist for each size of experiment considered are given in Table 3.5. Examining these we observe that there will be a considerable saving when calculating the performance measures if only one member from each of the equivalence classes need be used. In fact, we observe that for designs based on a complementary pair of Williams squares of side four the total number of equivalence classes is approximately one eighth of the total number of implementable designs for each size of experiment considered.

As before, in order to calculate performance measures for the planned design it is necessary to identify one design from each equivalence class and the size of the equivalence class to which each of these designs belongs. For example, when $n = 1$ we observe from Table 3.5 that the number of implementable designs is 256 and that these can be partitioned into 43 equivalence classes. Therefore, in order to evaluate the performance measures as efficiently as possible, it is necessary to find 43 non-equivalent designs and the size of the equivalence class to which each of these belongs.

To assist this process it is useful to obtain the equivalence class inventory. In order to obtain this we require the cycle index of G. This is given by

$$Z_G(s_1, s_2, s_4) = \frac{1}{8}(s_1^8 + 5s_2^4 + 2s_4^2). \tag{3.7}$$

Table 3.5: The number of equivalence classes and implementable designs for designs based on a complementary pair of Williams squares with up to six subjects per treatment sequence.

Number of subjects per treatment sequence, n.	Number of implementable designs, $\|D\|$.	Number of equivalence classes.
1	256	43
2	6561	873
3	65536	8356
4	390625	49225
5	1679616	210771
6	5764801	722113

If C is the set of dropout identifiers $\{0, 1\}$ and the weight function ω is given by

$$\omega(0) = x_0, \quad \omega(1) = x_1,$$

then applying Theorem 3.7 the equivalence class inventory is obtained by substituting $s_1 = x_0 + x_1$, $s_2 = x_0^2 + x_1^2$ and $s_4 = x_0^4 + x_1^4$ into equation (3.7). Therefore, the equivalence class inventory is

$$\frac{1}{8}[(x_0 + x_1)^8 + 5(x_0^2 + x_1^2)^4 + 2(x_0^4 + x_1^4)^2 =$$
$$x_0^8 + x_0^7 x_1 + 6x_0^6 x_1^2 + 7x_0^5 x_1^3 + 13x_0^4 x_1^4 + 7x_0^3 x_1^5 + 6x_0^2 x_1^6 + x_0 x_1^7 + x_1^8.$$

Examining each of the terms in this expression we find that the 43 equivalence classes can be divided into nine categories as shown in Table 3.6.

It is necessary to calculate one member from each of the 43 equivalence classes described in Table 3.6. Having done this the size of each of their respective classes can be determined by obtaining the stabilizer for each design and applying Theorem 3.8.

Table 3.6: The interpretation of the equivalence class inventory of Example 3.13.

Number of equivalence classes.	Number of treatment sequences containing	
	zero dropouts	one dropout.
1	8	0
1	7	1
6	6	2
7	5	3
13	4	4
7	3	5
6	2	6
1	1	7
1	0	8

Example 3.14 Let $d(4, 12, n, 4)$ be the design based on a complete set of mutually orthogonal Latin squares of side four with treatment labels 0, 1, 2 and 3 and n subjects allocated to each distinct treatment sequence.

The set, R, of distinct treatment sequences consists of the following elements.

$$r_1 = 0\ 1\ 2\ 3,\ \ r_2 = 1\ 0\ 3\ 2,\ \ r_3 = 2\ 3\ 0\ 1,\ \ r_4 = 3\ 2\ 1\ 0,$$

$$r_5 = 0\ 3\ 1\ 2,\ \ r_6 = 1\ 2\ 0\ 3,\ \ r_7 = 2\ 1\ 3\ 0,\ \ r_8 = 3\ 0\ 2\ 1,$$

$$r_9 = 0\ 2\ 3\ 1,\ \ r_{10} = 1\ 3\ 2\ 0,\ \ r_{11} = 2\ 0\ 1\ 3 \text{ and } r_{12} = 3\ 1\ 0\ 2.$$

The group G consists of the following permutations of the distinct treatment sequences, written in cycle notation.

$$\begin{aligned}
g_1 &= (r_1)(r_2)(r_3)(r_4)(r_5)(r_6)(r_7)(r_8)(r_9)(r_{10})(r_{11})(r_{12}) \\
g_2 &= (r_1r_2)(r_3r_4)(r_5r_6)(r_7r_8)(r_9r_{10})(r_{11}r_{12}) \\
g_3 &= (r_1r_3)(r_2r_4)(r_5r_7)(r_6r_8)(r_9r_{11})(r_{10}r_{12}) \\
g_4 &= (r_1r_4)(r_2r_3)(r_5r_8)(r_6r_7)(r_9r_{12})(r_{10}r_{11})
\end{aligned}$$

$$
\begin{aligned}
g_5 &= (r_1r_5r_9)(r_2r_8r_{11})(r_3r_6r_{12})(r_4r_7r_{10}) \\
g_6 &= (r_1r_6r_{11})(r_2r_7r_9)(r_3r_5r_{10})(r_4r_8r_{12}) \\
g_7 &= (r_1r_7r_{12})(r_2r_6r_{10})(r_3r_8r_9)(r_4r_5r_{11}) \\
g_8 &= (r_1r_8r_{10})(r_2r_5r_{12})(r_3r_7r_{11})(r_4r_6r_9) \\
g_9 &= (r_1r_9r_5)(r_2r_{11}r_8)(r_3r_{12}r_6)(r_4r_{10}r_7) \\
g_{10} &= (r_1r_{10}r_8)(r_2r_{12}r_5)(r_3r_{11}r_7)(r_4r_9r_6) \\
g_{11} &= (r_1r_{11}r_6)(r_2r_9r_7)(r_3r_{10}r_5)(r_4r_{12}r_8) \\
g_{12} &= (r_1r_{12}r_7)(r_2r_{10}r_6)(r_3r_9r_8)(r_4r_{11}r_5)
\end{aligned}
$$

Hence,

$|G| = 12,\ N(g_1) = 12,\ N(g_2) = 6,\ N(g_3) = 6,\ N(g_4) = 6,\ N(g_5) = 4,\ N(g_6) = 4,$
$N(g_7) = 4,\ N(g_8) = 4,\ N(g_9) = 4,\ N(g_{10}) = 4,\ N(g_{11}) = 4$ and $N(g_{12}) = 4.$

Applying Theorem 3.8, we find that the total number of equivalence classes is given by

$$\frac{1}{12}[(n+1)^{12} + 3(n+1)^6 + 8(n+1)^4]. \tag{3.8}$$

The number of implementable designs which can be formed by dropping one or more subjects in the final period of the planned design of this example is given by $|D| = (n+1)^{12}$. Using expression (3.8) the number of equivalence classes into which the implementable designs are partitioned can be evaluated. These are listed in Table 3.7 for a range of values of n, the number of subjects allocated to each sequence, which might realistically be used in cross-over experiments.

Examining Table 3.7 we observe that, as before, there will be a considerable saving if the mean and variance of the performance measures for each of the planned designs is evaluated by considering only one member from each of the equivalence classes rather than evaluating the individual performance measures for all the implementable designs. In each of the planned designs listed in Table 3.7 the total number of equivalence classes is approximately one twelfth of the total number of implementable designs.

Table 3.7: The number of equivalence classes and implementable designs for designs based on a complete set of mutually orthogonal Latin squares of side four and up to three subjects per treatment sequence.

Number of subjects per treatment sequence, n.	Number of implementable designs, $\|D\|$.	Number of equivalence classes.
1	4096	368
2	531441	44523
3	16777216	1399296

3.6 Discussion

In this chapter methods have been presented for reducing the amount of computation required to assess the performance of a cross-over design, when final period dropout may occur, using the criteria proposed in Chapter 2. These allow the mean and variance of the individual performance measures of the implementable designs to be evaluated by obtaining the size of each of the equivalence classes into which the implementable designs can be partitioned and an assessment of only one member from each. Consequently, it is possible to reduce the amount of computation required. However, for certain designs involving relatively small numbers of subjects, this reduction is not as large as we would like it to be. For instance, consider the complete set of mutually orthogonal Latin squares of Example 3.14. For a relatively small study, that is one involving just 24 subjects, even after reducing the necessary computation to considering only the non-equivalent designs, to obtain the summary measures for the design's performance, it will be necessary to calculate performance measures for 44,523 different implementable designs. Therefore, even after the reduction made available by using the results presented in this chapter, a great deal of computation is required to obtain the summary measures for this particular design.

This will be true for any planned design which involves a large number of distinct

treatment sequences. The total number of implementable designs is considerably larger for designs involving the same numbers of subjects but larger numbers of distinct treatment sequences. For instance, when 24 subjects are allocated to a single Williams square, a complementary pair of Williams squares or a complete set of mutually orthogonal Latin squares of side four, the total numbers of implementable designs are respectively, 2401, 65536, or 531441. Consequently, even after reducing the necessary calculations to just the non-equivalent designs, the amount of computation required to evaluate the summary measures for designs involving a large number of distinct treatment sequences is considerable.

There is an additional difficulty involved in evaluating the performance measures for planned designs which give rise to a large number of implementable designs. To obtain the mean and variance of the individual performance measures it is necessary to obtain the individual probabilities for each of the implementable designs, for the particular probability of final period dropout being considered. If there are a large number of implementable designs, then each of these individual probabilities will be very small. In fact for certain designs these probabilities may be so small that it is difficult to accurately compute them, given the limitations regarding the smallest possible number that a computer can accurately store.

Considerable computational reductions can be achieved by using the methods described in this chapter when evaluating the mean and variance of the performance measures of a planned design. A computer program incorporating the methods described in this chapter and developed at my instigation by B. D. McKay (Australian National University) has been used in the comparative studies presented in Chapters 4 and 5. However, for designs involving large numbers of distinct treatment sequences, or large numbers of subjects allocated to a relatively small number of distinct treatment sequences, this method of assessing a planned design still incurs a large amount of computation.

Chapter 4

Four Treatment, Four Period Designs

4.1 Introduction

When planning any cross-over trial, before selecting an appropriate design for the study it is usual to consider the following:

(i) the number of treatments to be investigated,

(ii) the maximum number of treatment periods available,

(iii) the maximum number of treatment sequences which can reasonably be expected to be administered correctly, and

(iv) the maximum number of subjects available for the study.

Experiments in which cross-over designs are employed may be very different in terms of the number of treatments to be compared and the resources available. Much work has already been carried out to determine the most appropriate design(s) to employ for experiments with different requirements. Frequently, the recommended design is the design which is optimal, over the set of competing designs, for the

estimation of the direct and/or carry-over treatment comparisons under some appropriate optimality criterion such as A-optimality.

As stated previously in Section 2.1, selecting the appropriate design to use is one of the most crucial decisions made during the planning stage of any cross-over trial. At present, this decision is made without considering how the efficiency of the treatment comparisons, obtained for the various competing designs, might be affected if subjects drop out during the study. This happens even when information from previous, similar studies indicates that there is a real possibility of subjects dropping out during the later stages of the trial.

In Chapter 2, a method for assessing the performance of cross-over designs has been established, when it is believed a priori that dropouts may occur during the final period. In Chapter 3 we have shown how the considerable computation involved in evaluating these performance measures can be reduced using results from combinatorial theory. In the next two chapters, this methodology is used to study the performance of frequently employed cross-over designs when final period dropout may occur. It is a widely held belief that the longer a clinical trial lasts, the greater will be the probability of subjects dropping out, for reasons unrelated to the treatments administered. Consequently, it is uncommon and unwise for cross-over trials involving a large number of treatment periods to be contemplated. Often the maximum number of treatment periods believed to be viable is four. In this chapter we consider the performance of cross-over designs involving four treatment periods since, of the designs most commonly used in medical trials, it is these designs which appear to be most vulnerable to subjects dropping out during the final period. In Chapter 5 the performance of three period designs subject to final period dropout is addressed.

The main purposes of the study described in the current and next chapter are as follows:

1. To examine the robustness of the most frequently employed cross-over designs to final period dropout with probability, θ. Also to consider the sensitivity of

the mean performance measures to slight increases or decreases in the value of θ.

2. To compare the relative performance measures of competing designs and hence to recommend which designs should be used for studies in which final period dropouts are anticipated.

3. To investigate the "best" last period to employ in designs recommended for studies in which final period dropouts are anticipated. By "best" we mean the final period which enables the design to possess the maximum mean performance measures and the minimum variances for estimating the direct and/or carry-over treatment comparisons.

4. To investigate the properties of those designs which are more robust to final period dropout than others.

Throughout these chapters we adopt the simple carry-over model (1.1) for the observations. If a different model is adopted, the design assessments will yield different performance measures and this may alter the recommendations concerning design selection. In addition, we assume throughout that the purpose of the experiments in which these designs are to be employed is to estimate all the pairwise differences amongst the t treatments giving equal importance to each comparison. Hence, the contrasts of interest used are all the pairwise direct and first-order carry-over treatment comparisons. If different sets of contrasts are considered, or if some contrasts are given more importance than others, then the overall conclusions concerning design selection may again be different.

We begin, in the following section, by defining uniformly balanced designs. In subsequent sections we examine designs for four treatments and four periods of three different types: designs derived from a single Williams square, designs built from pairs of such squares and designs based upon sets of mutually orthogonal Latin squares. We investigate the performance of the designs for different values of θ. A

comparison is then made of the different types of designs and recommendations are given on which designs to employ when final period dropouts may occur.

4.2 Uniform Balanced Designs

In Sections 2.5 and 2.6 a Williams square of side four and 16 subjects was used to illustrate the design assessment procedures proposed in this thesis. This design possesses each of the following properties:

Definition 4.1 A design is **uniform** on the **periods** if, in each period, each treatment occurs equally often.

Definition 4.2 A design is **uniform** on the **subjects** if each subject receives each treatment equally often.

Definition 4.3 A design is **uniform** if it is uniform on both subjects and periods.

Definition 4.4 A design is **balanced** if each treatment is preceded equally often by every other treatment but never by itself.

Hedayat and Afsarinejad (1978) showed that, over the class of uniform designs in which $p = t$, a uniformly balanced design (that is, both uniform and balanced) is universally optimal for the estimation of the direct and first-order carry-over treatment effects under the simple carry-over model (1.1). Thus, the particular Williams design of Sections 2.5 and 2.6 is universally optimal because it is a uniformly balanced design. For this reason it is a design frequently employed in cross-over studies.

The class of cross-over designs over which Hedayat and Afsarinejad were able to establish universal optimality is, as commented on by several authors, somewhat restrictive. Cheng and Wu (1980) and Kunert (1983, 1984) have attempted to relax these conditions. Unfortunately there has been only limited success, particularly for the estimation of direct treatment effects which is often of greater importance

than the estimation of carry-over effects. Furthermore, the classes of designs over which these additional optimality results have been established involve additional treatment periods. Since it is undesirable to consider designs which have a large number of treatment periods for studies in which dropouts are anticipated, these results are of little benefit to the problem addressed in this thesis.

For studies involving four treatments there are good reasons for considering uniform balanced designs. They do not involve too many treatment periods and, if the probability of dropout is very small, there is a fairly high probability that the implemented experiment will be the planned design and hence a universally optimal design.

In the following sections, we examine the robustness to final period dropout of a variety of uniform balanced designs of practical size. In addition, for designs of an equal size we recommend which of the designs should be used when final period dropouts are anticipated.

4.3 Examination of Williams Square Designs

In this section, the results are presented of an investigation into the performance of designs based on a Williams square of side four involving up to 48 subjects, for estimating direct and carry-over treatment effects. There are 12 different designs to consider, namely the designs in which n, the number of subjects allocated to each treatment sequence, takes each of the values $1, \ldots, 12$. For each design Tables 4.1.1 - 4.1.12 (given on pages 101-106) contain the mean and variance of the performance measures X_d and Y_d under the A-criterion, over the range of possible θ values $0 \leq \theta \leq 1$ in steps of 0.1. The tables provide a summary of the performance of the average variance of the direct and carry-over treatment effects for each of the 12 planned designs.

Figures 4.1 and 4.2 show how $E[X_d|\theta]$ and $E[Y_d|\theta]$ change with θ for two of these designs, namely, those involving 24 and 48 subjects. The bars represent $E[X_d|\theta] \pm$

Table 4.1: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for designs based on a Williams square of side four and ≤ 48 subjects.

Table 4.1.1: $t = 4$, $m = 4$, $n = 1$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	1.82	0.00	1.25	0.00
0.1	1.50	0.24	1.01	0.13
0.2	1.18	0.43	0.78	0.21
0.3	0.87	0.49	0.57	0.23
0.4	0.60	0.45	0.39	0.20
0.5	0.38	0.34	0.24	0.15
0.6	0.21	0.21	0.13	0.09
0.7	0.09	0.10	0.06	0.04
0.8	0.03	0.03	0.02	0.01
0.9	4.0×10^{-3}	4.3×10^{-3}	2.5×10^{-3}	1.7×10^{-3}
1.0	0.00	0.00	0.00	0.00

Table 4.1.2: $t = 4$, $m = 4$, $n = 2$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	3.64	0.00	2.50	0.00
0.1	3.38	0.13	2.28	0.08
0.2	3.05	0.37	2.03	0.21
0.3	2.65	0.71	1.73	0.36
0.4	2.18	1.06	1.39	0.49
0.5	1.65	1.27	1.03	0.55
0.6	1.11	1.21	0.68	0.49
0.7	0.61	0.85	0.37	0.32
0.8	0.24	0.37	0.14	0.14
0.9	0.04	0.06	0.02	0.02
1.0	0.00	0.00	0.00	0.00

Table 4.1.3: $t = 4$, $m = 4$, $n = 3$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	5.45	0.00	3.75	0.00
0.1	5.15	0.10	3.49	0.07
0.2	4.78	0.28	3.19	0.18
0.3	4.33	0.59	2.84	0.33
0.4	3.79	1.06	2.43	0.53
0.5	3.13	1.61	1.96	0.74
0.6	2.34	2.04	1.43	0.84
0.7	1.47	1.96	0.88	0.74
0.8	0.66	1.17	0.39	0.41
0.9	0.13	0.25	0.07	0.08
1.0	0.00	0.00	0.00	0.00

Table 4.1.4: $t = 4$, $m = 4$, $n = 4$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.90	0.10	4.68	0.07
0.2	6.46	0.26	4.32	0.17
0.3	5.95	0.53	3.91	0.32
0.4	5.33	0.97	3.43	0.53
0.5	4.58	1.62	2.87	0.79
0.6	3.63	2.40	2.22	1.03
0.7	2.50	2.86	1.48	1.10
0.8	1.26	2.21	0.73	0.77
0.9	0.28	0.61	0.16	0.20
1.0	0.00	0.00	0.00	0.00

Table 4.1.5: $t = 4$, $m = 4$, $n = 5$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	9.09	0.00	6.25	0.00
0.1	8.65	0.11	5.87	0.08
0.2	8.14	0.27	5.44	0.18
0.3	7.54	0.53	4.96	0.33
0.4	6.84	0.94	4.41	0.54
0.5	5.98	1.58	3.76	0.81
0.6	4.91	2.49	3.00	1.12
0.7	3.57	3.42	2.11	1.35
0.8	1.97	3.26	1.13	1.14
0.9	0.49	1.13	0.27	0.36
1.0	0.00	0.00	0.00	0.00

Table 4.1.6: $t = 4$, $m = 4$, $n = 6$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.91	0.00	7.50	0.00
0.1	10.39	0.12	7.06	0.08
0.2	9.80	0.29	6.56	0.20
0.3	9.12	0.54	6.01	0.35
0.4	8.32	0.94	5.37	0.56
0.5	7.36	1.57	4.63	0.83
0.6	6.16	2.52	3.77	1.18
0.7	4.64	3.73	2.75	1.51
0.8	2.73	4.16	1.56	1.46
0.9	0.76	1.80	0.42	0.57
1.0	0.00	0.00	0.00	0.00

Table 4.1.7: $t = 4$, $m = 4$, $n = 7$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.73	0.00	8.75	0.00
0.1	12.14	0.13	8.25	0.09
0.2	11.47	0.31	7.68	0.21
0.3	10.70	0.58	7.05	0.37
0.4	9.80	0.97	6.33	0.59
0.5	8.72	1.59	5.50	0.87
0.6	7.39	2.55	4.53	1.23
0.7	5.70	3.90	3.38	1.62
0.8	3.53	4.87	2.01	1.73
0.9	1.08	2.57	0.59	0.80
1.0	0.00	0.00	0.00	0.00

Table 4.1.8: $t = 4$, $m = 4$, $n = 8$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	14.55	0.00	10.00	0.00
0.1	13.88	0.14	9.43	0.10
0.2	13.13	0.34	8.80	0.23
0.3	12.27	0.62	8.09	0.40
0.4	11.27	1.02	7.29	0.63
0.5	10.07	1.64	6.36	0.92
0.6	8.61	2.60	5.28	1.29
0.7	6.75	4.03	4.01	1.71
0.8	4.33	5.42	2.47	1.95
0.9	1.44	3.39	0.79	1.05
1.0	0.00	0.00	0.00	0.00

Table 4.1.9: $t = 4$, $m = 4$, $n = 9$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	16.36	0.00	11.25	0.00
0.1	15.63	0.16	10.62	0.11
0.2	14.79	0.37	9.92	0.25
0.3	13.84	0.66	9.13	0.44
0.4	12.73	1.08	8.24	0.67
0.5	11.42	1.71	7.22	0.97
0.6	9.81	2.67	6.03	1.35
0.7	7.79	4.14	4.63	1.80
0.8	5.14	5.83	2.93	2.12
0.9	1.82	4.22	0.99	1.31
1.0	0.00	0.00	0.00	0.00

Table 4.1.10: $t = 4$, $m = 4$, $n = 10$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	18.18	0.00	12.50	0.00
0.1	17.37	0.17	11.81	0.12
0.2	16.46	0.40	11.03	0.28
0.3	15.41	0.71	10.17	0.47
0.4	14.20	1.15	9.19	0.71
0.5	12.76	1.79	8.07	1.03
0.6	11.01	2.77	6.78	1.42
0.7	8.82	4.25	5.24	1.88
0.8	5.95	6.16	3.39	2.27
0.9	2.23	5.04	1.21	1.56
1.0	0.00	0.00	0.00	0.00

Table 4.1.11: $t = 4$, $m = 4$, $n = 11$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	20.00	0.00	13.75	0.00
0.1	19.12	0.18	12.99	0.13
0.2	18.12	0.43	12.15	0.30
0.3	16.98	0.76	11.21	0.50
0.4	15.66	1.22	10.14	0.76
0.5	14.10	1.88	8.93	1.08
0.6	12.21	2.87	7.52	1.49
0.7	9.85	4.38	5.86	1.97
0.8	6.75	6.43	3.85	2.40
0.9	2.66	5.82	1.45	1.80
1.0	0.00	0.00	0.00	0.00

Table 4.1.12: $t = 4$, $m = 4$, $n = 12$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	21.82	0.00	15.00	0.00
0.1	20.86	0.20	14.18	0.14
0.2	19.78	0.45	13.27	0.32
0.3	18.55	0.80	12.25	0.54
0.4	17.12	1.29	11.10	0.81
0.5	15.44	1.98	9.78	1.15
0.6	13.40	2.99	8.26	1.56
0.7	10.87	4.52	6.47	2.06
0.8	7.55	6.66	4.31	2.52
0.9	3.10	6.55	1.68	2.03
1.0	0.00	0.00	0.00	0.00

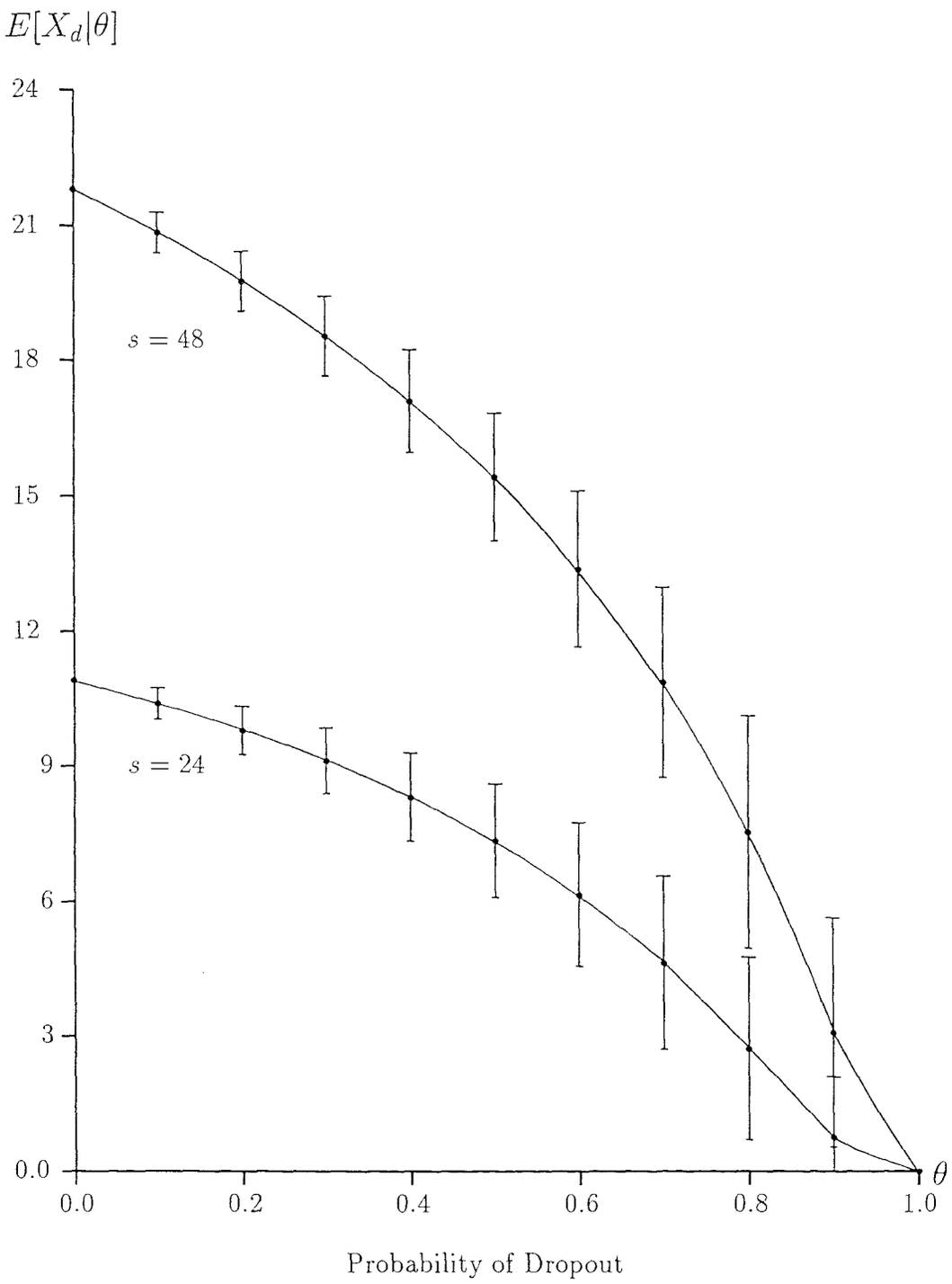


Figure 4.1: Performance for direct treatment comparisons under the A-criterion of single Williams square designs for 24 and 48 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

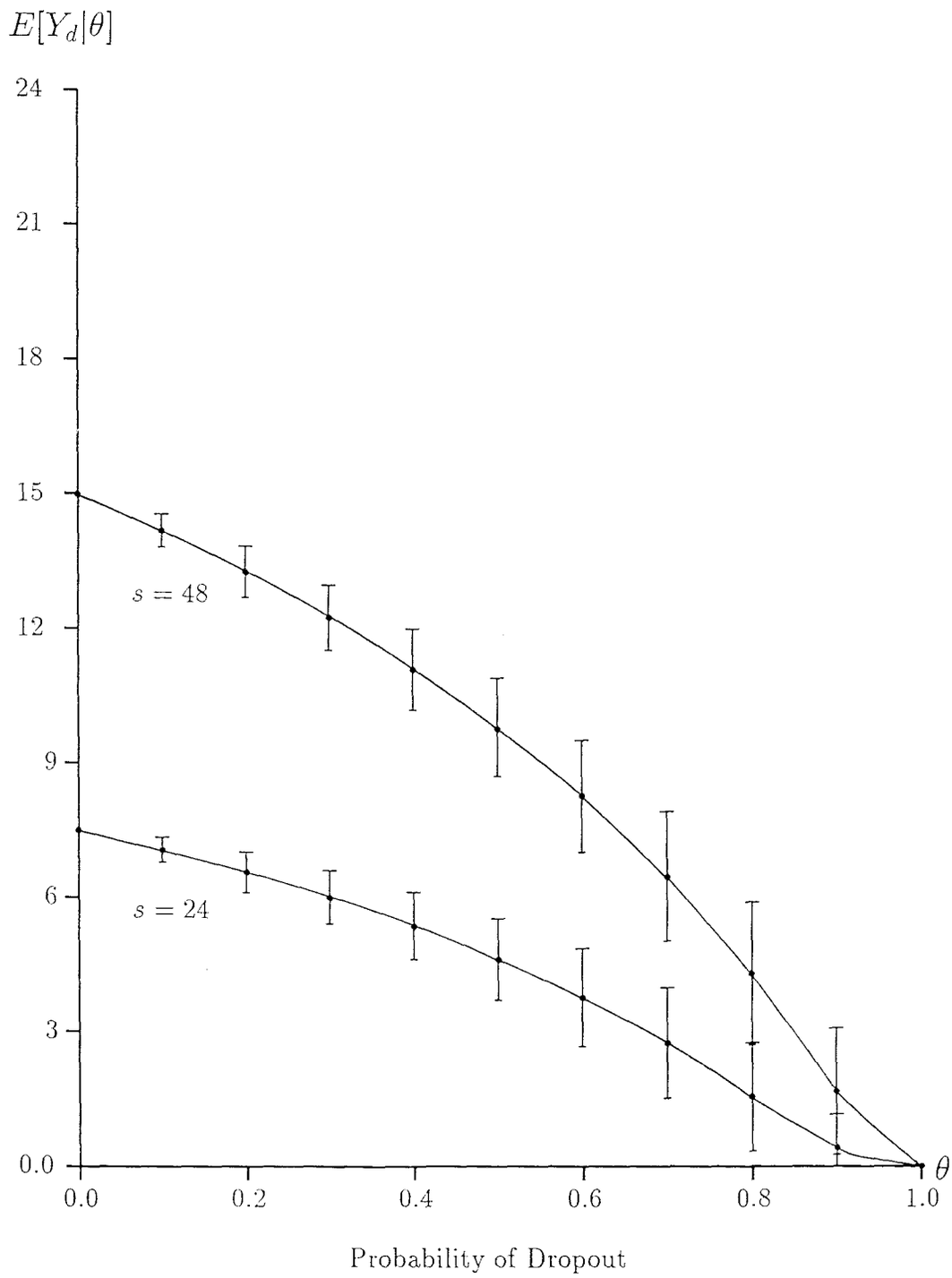


Figure 4.2: Performance for carry-over treatment comparisons under the A-criterion of single Williams square designs for 24 and 48 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

$\sqrt{Var[X_d|\theta]}$ and $E[Y_d|\theta] \pm \sqrt{Var[Y_d|\theta]}$ respectively for each design. These have been included to give an indication of the spread of the distributions for each value of θ. For larger values of θ, the full extent of the bar is not given since it would extend below the θ axis.

From Figures 4.1 and 4.2 we observe that, as expected, the mean values of both X_d and Y_d increase as the number of subjects, n, allocated to each treatment sequence is increased. In each case we observe that, as θ increases, there is a gradual reduction in the mean values of X_d and Y_d for each design. Note that, when the probability of final period dropout is anticipated to be $\theta = 0.0$, the set of implementable designs, D, contains only one design which is the planned design. When the probability of final period dropout is anticipated to be $\theta = 1.0$, D again consists of only one design, namely the planned design with the entire final period deleted, which is a disconnected design. The sudden and, in the case of the larger designs, very rapid reduction in the mean values of both performances measures as the value of θ approaches 1.0 is explained by the fact that, in every case, the set of implementable designs, D, contains a number of disconnected designs.

The spread of the distributions is always smaller when θ is either very small or very large. This is because, irrespective of the number of subjects allocated to the design, the distribution of X_d and Y_d will always be dominated by those designs with the greatest probability of being implemented. When θ is small these are the designs with the fewest number of dropouts and consequently the higher performance measures. When θ is large these are the designs containing a greater number of final period dropouts and consequently the poorer performance measures; for a Williams square of side four these are mainly disconnected designs.

Tables 4.2.1 - 4.2.12 (given on pages 111-116) contain the mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for the same values of θ and for each size of design in turn. The tables provide a summary of the performance of the maximum variance of the direct and carry-over treatment effects for each of the 12 planned designs considered. As mentioned in Section 2.6, the

performance measures obtained using either the A- or MV-criterion are related and will be identical if the particular implementable design being evaluated is variance balanced. A uniform balanced design is always variance balanced for the direct and first-order carry-over pairwise treatment comparisons. Hence, when $\theta = 0.0$, all the designs considered in this section will give equal mean performance measures for both X_d and Y_d based on either the A- or MV-criterion. However, when $\theta > 0$ the set of implementable designs will contain only a small number of variance balanced designs. For example, for a design based on a Williams square of side four and $\theta > 0$, the variance balanced designs are those designs in which the number of subjects dropping out during the final period is the same for each treatment sequence. These form only a very small proportion of all the implementable designs. It is therefore, important to examine the performance of designs under both the A- and MV-criteria.

The mean values of X_d and Y_d, obtained under the MV-criterion, will always be smaller than their corresponding A-criterion values irrespective of the planned design being assessed, see Section 2.6. By comparing the distributions of X_d and Y_d obtained using each criterion we can examine the differences between the mean average variance of the treatment contrasts and the mean maximum variance of the treatment contrasts. If these do not differ by much we can conclude that the differences amongst the variances of the treatment comparisons obtained from the implemented experiment should not be great. Comparing Tables 4.1 and 4.2 we observe that, for small values of θ, say $\theta < 0.2$, this difference is always small. For example, for the design $d(4, 4, 12, 4)$ when $\theta = 0.2$, the mean performance measures for estimating the direct treatment comparisons under the A- and MV-criteria are 19.78 and 18.80 respectively. Note that the relative difference between the respective measures becomes smaller as the number of subjects allocated to the design increases. This is true for both direct and carry-over treatment effects.

Figures 4.3 and 4.4 show how $E[X_d|\theta]$ and $E[Y_d|\theta]$, obtained under the MV-criterion, change with θ for two of these designs namely those involving 24 and 48 subjects. As previously, the bars represent $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$ and $E[Y_d|\theta] \pm$

Table 4.2: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for designs based on a Williams square of side four and ≤ 48 subjects.

Table 4.2.1: $t = 4$, $m = 4$, $n = 1$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	1.82	0.00	1.25	0.00
0.1	1.39	0.36	0.94	0.19
0.2	1.03	0.49	0.68	0.25
0.3	0.72	0.47	0.47	0.22
0.4	0.47	0.37	0.31	0.17
0.5	0.29	0.24	0.18	0.11
0.6	0.15	0.13	0.10	0.06
0.7	0.07	0.06	0.04	0.02
0.8	0.02	0.02	0.01	0.01
0.9	2.6×10^{-3}	2.0×10^{-3}	1.6×10^{-3}	7.8×10^{-4}
1.0	0.00	0.00	0.00	0.00

Table 4.2.2: $t = 4$, $m = 4$, $n = 2$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	3.64	0.00	2.50	0.00
0.1	3.19	0.28	2.14	0.16
0.2	2.74	0.59	1.80	0.31
0.3	2.26	0.89	1.46	0.42
0.4	1.77	1.07	1.12	0.48
0.5	1.27	1.06	0.79	0.44
0.6	0.81	0.85	0.49	0.33
0.7	0.42	0.50	0.26	0.19
0.8	0.16	0.19	0.09	0.07
0.9	0.02	0.03	0.01	0.01
1.0	0.00	0.00	0.00	0.00

Table 4.2.3: $t = 4$, $m = 4$, $n = 3$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	5.45	0.00	3.75	0.00
0.1	4.90	0.25	3.30	0.16
0.2	4.37	0.55	2.88	0.31
0.3	3.80	0.92	2.44	0.47
0.4	3.17	1.33	1.99	0.61
0.5	2.48	1.62	1.52	0.68
0.6	1.75	1.64	1.05	0.63
0.7	1.03	1.26	0.61	0.46
0.8	0.43	0.61	0.25	0.21
0.9	0.08	0.10	0.04	0.03
1.0	0.00	0.00	0.00	0.00

Table 4.2.4: $t = 4$, $m = 4$, $n = 4$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.60	0.26	4.45	0.17
0.2	5.97	0.56	3.94	0.34
0.3	5.30	0.96	3.41	0.52
0.4	4.55	1.43	2.85	0.69
0.5	3.70	1.90	2.26	0.81
0.6	2.77	2.18	1.65	0.85
0.7	1.77	2.00	1.04	0.72
0.8	0.83	1.20	0.48	0.40
0.9	0.17	0.25	0.10	0.08
1.0	0.00	0.00	0.00	0.00

Table 4.2.5: $t = 4$, $m = 4$, $n = 5$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	9.09	0.00	6.25	0.00
0.1	8.31	0.28	5.60	0.19
0.2	7.58	0.59	5.00	0.37
0.3	6.79	1.02	4.37	0.57
0.4	5.90	1.54	3.70	0.76
0.5	4.91	2.09	3.00	0.92
0.6	3.80	2.54	2.26	1.00
0.7	2.57	2.61	1.49	0.93
0.8	1.30	1.85	0.74	0.61
0.9	0.30	0.48	0.17	0.15
1.0	0.00	0.00	0.00	0.00

Table 4.2.6: $t = 4$, $m = 4$, $n = 6$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.91	0.00	7.50	0.00
0.1	10.01	0.30	6.76	0.21
0.2	9.18	0.64	6.05	0.41
0.3	8.27	1.10	5.32	0.63
0.4	7.25	1.66	4.55	0.84
0.5	6.11	2.27	3.74	1.02
0.6	4.82	2.83	2.87	1.14
0.7	3.38	3.09	1.96	1.11
0.8	1.82	2.48	1.03	0.82
0.9	0.46	0.77	0.26	0.24
1.0	0.00	0.00	0.00	0.00

Table 4.2.7: $t = 4$, $m = 4$, $n = 7$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.73	0.00	8.75	0.00
0.1	11.72	0.32	7.91	0.23
0.2	10.78	0.70	7.11	0.45
0.3	9.75	1.20	6.28	0.69
0.4	8.60	1.79	5.40	0.92
0.5	7.31	2.46	4.47	1.12
0.6	5.84	3.10	3.48	1.26
0.7	4.20	3.49	2.43	1.26
0.8	2.37	3.06	1.34	1.00
0.9	0.65	1.12	0.36	0.34
1.0	0.00	0.00	0.00	0.00

Table 4.2.8: $t = 4$, $m = 4$, $n = 8$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	14.55	0.00	10.00	0.00
0.1	13.43	0.35	9.06	0.25
0.2	12.38	0.76	8.17	0.49
0.3	11.23	1.29	7.24	0.75
0.4	9.95	1.93	6.25	1.00
0.5	8.50	2.65	5.20	1.23
0.6	6.86	3.35	4.08	1.38
0.7	5.01	3.84	2.90	1.40
0.8	2.94	3.57	1.65	1.17
0.9	0.87	1.50	0.48	0.46
1.0	0.00	0.00	0.00	0.00

Table 4.2.9: $t = 4$, $m = 4$, $n = 9$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	16.36	0.00	11.25	0.00
0.1	15.14	0.38	10.22	0.27
0.2	13.99	0.82	9.23	0.54
0.3	12.72	1.39	8.19	0.81
0.4	11.29	2.07	7.10	1.08
0.5	9.69	2.84	5.93	1.33
0.6	7.88	3.61	4.69	1.50
0.7	5.83	4.17	3.37	1.54
0.8	3.51	4.02	1.97	1.32
0.9	1.11	1.91	0.61	0.58
1.0	0.00	0.00	0.00	0.00

Table 4.2.10: $t = 4$, $m = 4$, $n = 10$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	18.18	0.00	12.50	0.00
0.1	16.85	0.40	11.37	0.29
0.2	15.59	0.89	10.29	0.58
0.3	14.20	1.49	9.15	0.88
0.4	12.64	2.22	7.94	1.17
0.5	10.88	3.04	6.66	1.43
0.6	8.90	3.86	5.29	1.62
0.7	6.65	4.49	3.84	1.67
0.8	4.09	4.44	2.29	1.46
0.9	1.36	2.33	0.75	0.70
1.0	0.00	0.00	0.00	0.00

Table 4.2.11: $t = 4$, $m = 4$, $n = 11$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	20.00	0.00	13.75	0.00
0.1	18.56	0.43	12.52	0.32
0.2	17.19	0.95	11.34	0.62
0.3	15.68	1.60	10.11	0.94
0.4	13.98	2.37	8.79	1.26
0.5	12.08	3.24	7.39	1.54
0.6	9.92	4.12	5.90	1.74
0.7	7.46	4.81	4.31	1.80
0.8	4.67	4.82	2.61	1.59
0.9	1.63	2.75	0.89	0.82
1.0	0.00	0.00	0.00	0.00

Table 4.2.12: $t = 4$, $m = 4$, $n = 12$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	21.82	0.00	15.00	0.00
0.1	20.28	0.46	13.68	0.34
0.2	18.80	1.01	12.40	0.67
0.3	17.16	1.70	11.06	1.01
0.4	15.33	2.51	9.64	1.34
0.5	13.27	3.44	8.13	1.64
0.6	10.93	4.38	6.51	1.86
0.7	8.28	5.13	4.78	1.92
0.8	5.25	5.19	2.94	1.71
0.9	1.91	3.16	1.04	0.94
1.0	0.00	0.00	0.00	0.00

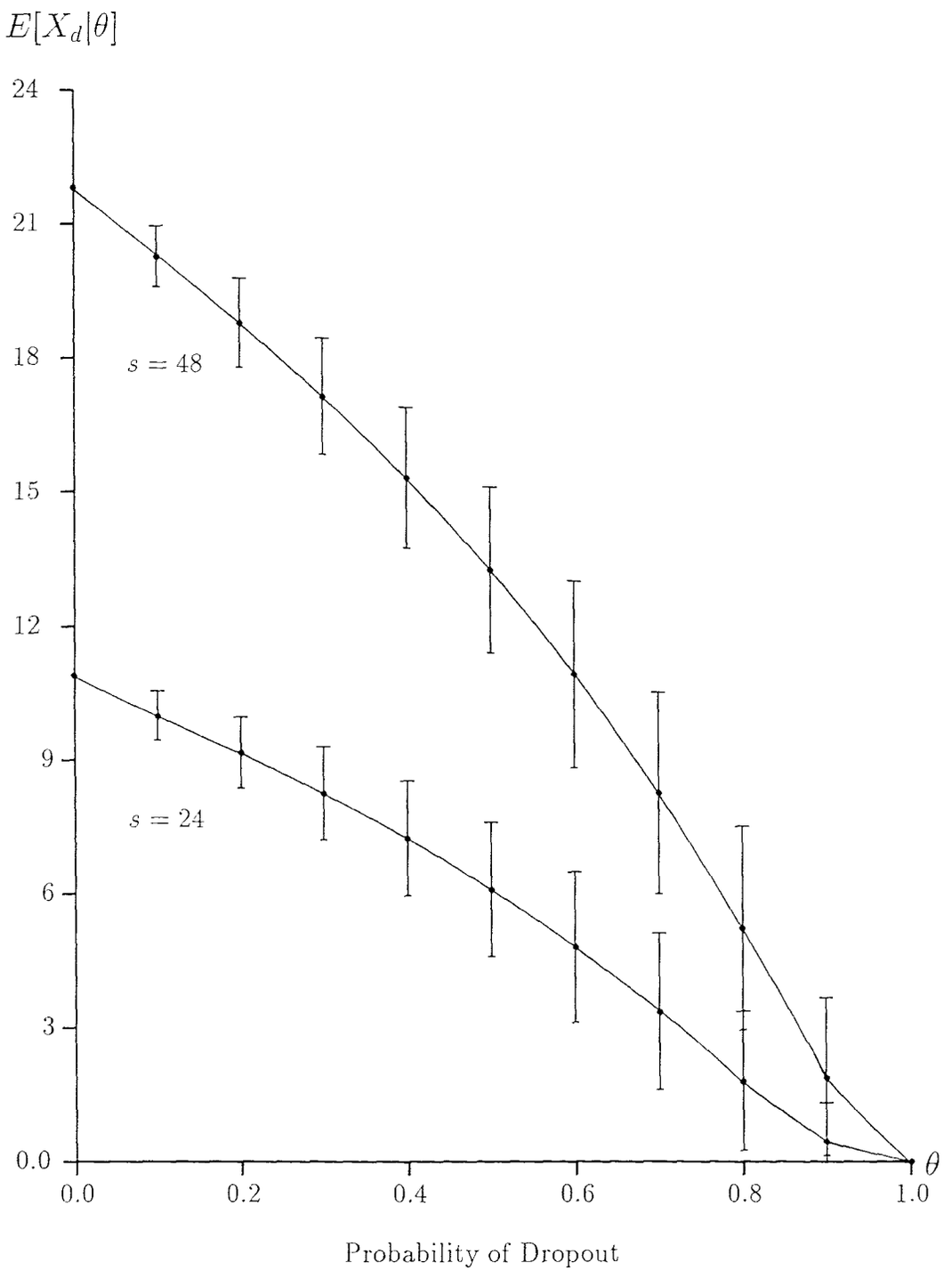


Figure 4.3: Performance for direct treatment comparisons under the MV-criterion of single Williams square designs for 24 and 48 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

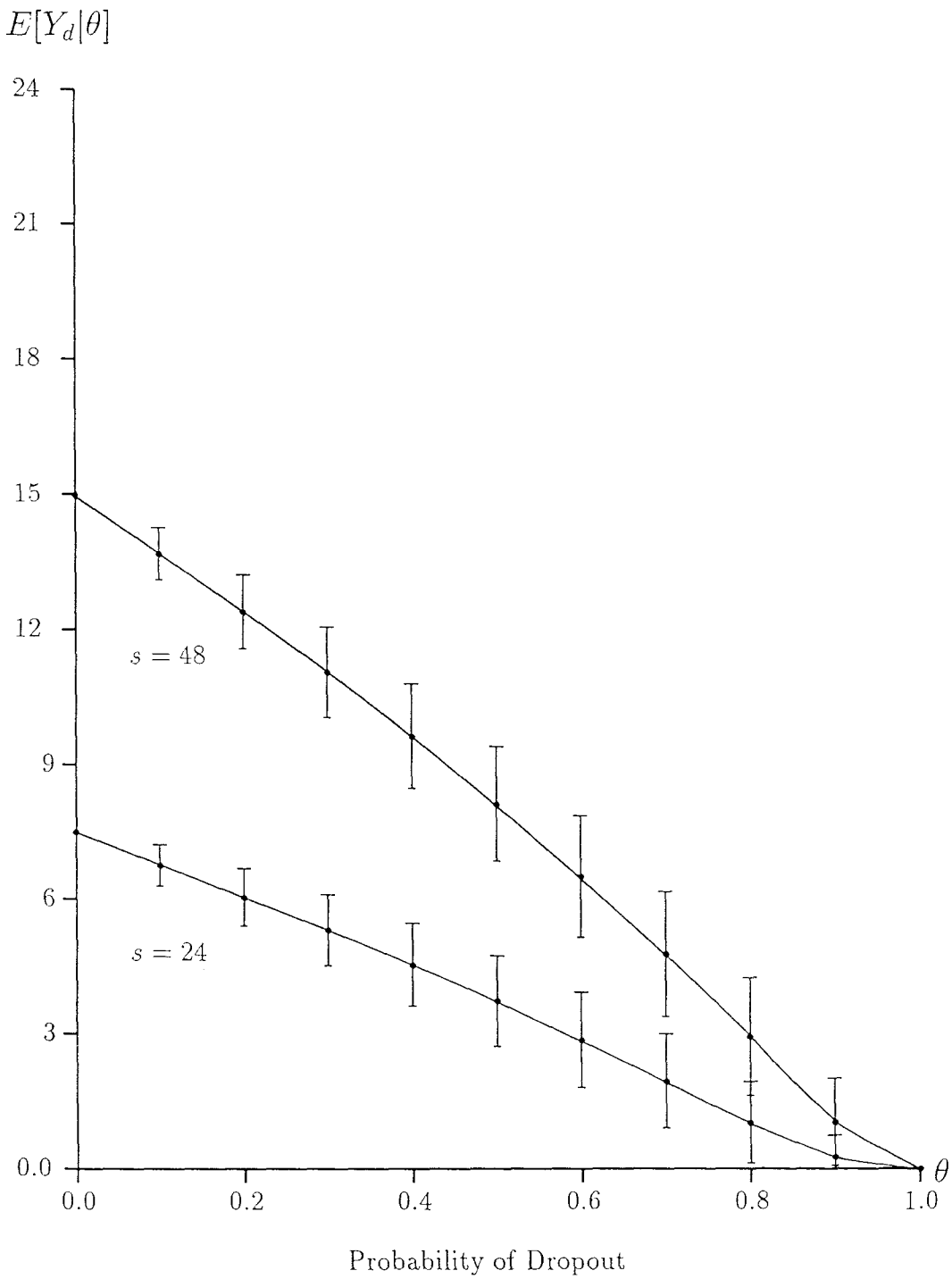


Figure 4.4: Performance for carry-over treatment comparisons under the MV-criterion of single Williams square designs for 24 and 48 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

$\sqrt{\text{Var}[Y_d|\theta]}$, and are included to give an indication of the spread of the distributions for the associated value of θ. Apart from the fact that the values obtained for the mean performance measures for X_d and Y_d under the MV-criterion are always less than for the A-criterion, the observations made concerning the general trend of X_d and Y_d as θ changes when obtained under the A-criterion all continue to apply here.

Each of the planned designs obtained by allocating an increasing number of subjects to the treatment sequences of a Williams square of side four gives rise to a number of disconnected implementable designs, D_0. It is necessary, therefore, to examine $P(D_0)$, the probability of implementing a disconnected design, when considering the robustness to final period dropout of each of these designs. The distribution of $P(D_0)$ can be investigated for each design by applying equation (2.1) to the set of disconnected designs arising from each planned design and then considering values of θ, in steps of 0.1 across $0 \leq \theta \leq 1$. The values of $P(D_0)$ are given in Tables 4.3. in which all probabilities are listed correct to two decimal places.

It is a reasonable assumption that, any design in which the probability of implementing a disconnected design is greater than 0.2 represents too great a risk to an experimenter. Examining the probabilities given in Table 4.3 we observe that, when $\theta \leq 0.2$, the probability of realising an implemented design is below 0.2. When $\theta \geq 0.9$, however, the probability is consistently above 0.2 and in most cases considerably greater. Whenever the probability of subjects dropping out during the final period of a four period study is anticipated to be as high as 0.9, it is unwise to consider running a four period study. A three period study is a safer option. In fact, it is very unlikely that a four period study would be seriously considered if the expected number of subjects to be lost during the final period exceeds the expected number completing the study, that is for $\theta > 0.5$. Examining Table 4.3 we observe that, provided the number of subjects available for the study is greater than or equal to 12, then when $\theta < 0.5$ the probability of implementing a disconnected design never exceeds 0.08.

Table 4.3: Probability, $P(D_0)$, of implementing a disconnected design for designs based on a Williams square of side four and ≤ 48 subjects.

	$P(D_0)$					
θ	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
0.0	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.05	0.00	0.00	0.00	0.00	0.00
0.2	0.18	0.01	0.00	0.00	0.00	0.00
0.3	0.35	0.04	0.00	0.00	0.00	0.00
0.4	0.52	0.12	0.02	0.00	0.00	0.00
0.5	0.69	0.26	0.08	0.02	0.01	0.00
0.6	0.82	0.45	0.21	0.08	0.03	0.01
0.7	0.92	0.67	0.42	0.25	0.13	0.07
0.8	0.97	0.86	0.71	0.54	0.40	0.28
0.9	0.99	0.98	0.94	0.88	0.81	0.73
1.0	1.00	1.00	1.00	1.00	1.00	1.00
	$P(D_0)$					
θ	$n=7$	$n=8$	$n=9$	$n=10$	$n=11$	$n=12$
0.0	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.00	0.00	0.00	0.00	0.00	0.00
0.2	0.00	0.00	0.00	0.00	0.00	0.00
0.3	0.00	0.00	0.00	0.00	0.00	0.00
0.4	0.00	0.00	0.00	0.00	0.00	0.00
0.5	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.00	0.00	0.00	0.00	0.00	0.00
0.7	0.04	0.02	0.01	0.00	0.00	0.00
0.8	0.20	0.13	0.09	0.05	0.04	0.03
0.9	0.65	0.58	0.50	0.43	0.37	0.32
1.0	1.00	1.00	1.00	1.00	1.00	1.00

4.4 Combining Williams Squares

A design based on a Williams square of side four is the design, for four treatments and four periods, which has the fewest number of distinct treatment sequences to achieve uniform balance. (See Definitions 4.3 and 4.4.) If the total number of subjects available is only four, which is inconceivable in practice, this is the only uniform balanced design available. However, when the number of subjects available is $4n$, for some integer $n > 1$, this design consists of the same four treatment sequences each replicated n times. If, rather than allocating increasing numbers of subjects to the same set of treatment sequences, it is possible to increase the total number of treatment sequences then other uniform balanced designs are available. In this section we consider the performance under final period dropout of uniform balanced designs formed by combining two Williams squares of side four which are "different" in the way described below.

It is known that all Williams squares of side four obtained under some permutation of the treatment labels are isomorphic. If we consider all the possible isomorphic designs formed by permuting the treatment labels 0, 1, 2 and 3 of the Williams square with initial treatment sequence (0 1 3 2) there are 24 different possible arrangements. In a cross-over design the order of the treatment sequences is unimportant since it does not affect the overall structure of the design. If the order of the treatment sequences is ignored, there are only six arrangements which have distinct sets of treatment sequences. These six squares, arranged such that the treatment labels in the first column of each square appears in lexicographical order, are given in Table 4.4.

Note that, each of the squares in Table 4.4 is a uniform balanced design. Therefore, when the probability of final period dropout is not considered they are all universally optimal over the class of uniform designs for estimating the direct and carry-over treatment effects. In addition, when assessing the performance of each design subject to final period dropout, the mean and variance of the performance

Table 4.4: Six "different" Williams squares of side four.

(i)	0	1	3	2	(ii)	0	3	1	2	(iii)	0	2	1	3
	1	2	0	3		1	0	2	3		1	0	3	2
	2	3	1	0		2	1	3	0		2	3	0	1
	3	0	2	1		3	2	0	1		3	1	2	0
(iv)	0	1	2	3	(v)	0	3	2	1	(vi)	0	2	3	1
	1	3	0	2		1	2	3	0		1	3	2	0
	2	0	3	1		2	0	1	3		2	1	0	3
	3	2	1	0		3	1	0	2		3	0	1	2

measures X_d and Y_d, obtained using either the A- or MV-criterion, are identical over the entire range of θ values, for designs involving equal numbers of subjects. The same is true of the respective probabilities of implementing a disconnected design. These results are unsurprising since the designs are all isomorphic under permutations of the treatment labels. The mean and variance of the performance measures X_d and Y_d and the probabilities of implementing a disconnected design have already been given in Section 4.3 for design (i) involving up to a maximum of 48 subjects.

Any eight sequence design formed as the union of the sequences from any two of the Williams squares in Table 4.4 is uniform balanced. Hence, when the probability of final period dropout is not considered, each of these designs is universally optimal. The same is true for any design formed by combining in this way any number of squares. The question we address in the remainder of this section is: when combining two or more "different" Williams squares, does the choice of squares affect the robustness of the design to final period dropout?

There are $\begin{pmatrix} 6 \\ 2 \end{pmatrix} = 15$ different possible designs which can be formed by combining any two of the Williams squares given in Table 4.4. Example 4.1 compares

the relative performance of two of these designs.

Example 4.1 The performance of the following two planned designs will be compared under their repeated use in cross-over trials with given probability, θ, of final period dropout.

Design (b): The pair of Williams squares $d(4, 8, 2, 4)$ with treatment labels 0, 1, 2 and 3 and initial sequences (0 1 3 2) and (0 3 1 2). This is the design considered previously in Example 2.5.

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1
0	3	1	2
1	0	2	3
2	1	3	0
3	2	0	1

Note: The design is the union of the sequences from squares (i) and (ii) of Table 4.4.

Design (c): The pair of Williams squares $d(4, 8, 2, 4)$ with treatment labels 0, 1, 2 and 3 and initial sequences (0 1 3 2) and (0 2 1 3).

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1
0	2	1	3
1	0	3	2
2	3	0	1
3	1	2	0

Note: The design is the union of the sequences from squares (i) and (iii) of Table 4.4.

The performance of each design in comparing all the direct and first-order carry-over treatment effects can be summarised by applying equations (2.4), (2.5), (2.6)

and (2.7) with the A-criterion to obtain the average variance of the direct and first-order carry-over treatment effects. Summary measures for design (b) have already been given in Table 2.6. Summary measures for design (c) are given in Table 4.5.

Table 4.5: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for design (c), with 16 subjects.

θ	$E[X_d\|\theta]$	$\text{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\text{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.92	0.08	4.69	0.06
0.2	6.54	0.17	4.38	0.12
0.3	6.14	0.28	4.04	0.18
0.4	5.70	0.42	3.69	0.26
0.5	5.22	0.59	3.31	0.34
0.6	4.68	0.79	2.90	0.42
0.7	4.06	0.96	2.46	0.47
0.8	3.36	0.94	1.96	0.43
0.9	2.61	0.51	1.46	0.22
1.0	2.15	0.00	1.16	0.00

Comparisons of the graphs of the mean of the performance measures X_d and Y_d (defined in Definitions 2.4 and 2.5) against θ, for $0 \leq \theta \leq 1$, for designs (b) and (c), are given in Figures 4.5 and 4.6 respectively. From these we observe that each design gives rise to identical mean performance measures correct to two decimal places for $\theta \leq 0.2$. As the value of θ increases beyond $\theta = 0.4$, however, each of the graphs depicting the mean of X_d and Y_d begins to diverge. In each case it is design (b) which gives rise to the larger mean performance measures. Consequently, the range of values for the mean of X_d or Y_d as θ varies is not as large for design (b) as it is for design (c). For design (b) the ranges are 7.27-3.20 and 5.00-1.71 respectively, while for design (c) they are 7.27-2.15 and 5.00-1.16 respectively. Note that when θ is small, and particularly when $\theta = 0.3$, the mean of Y_d for design (c) marginally

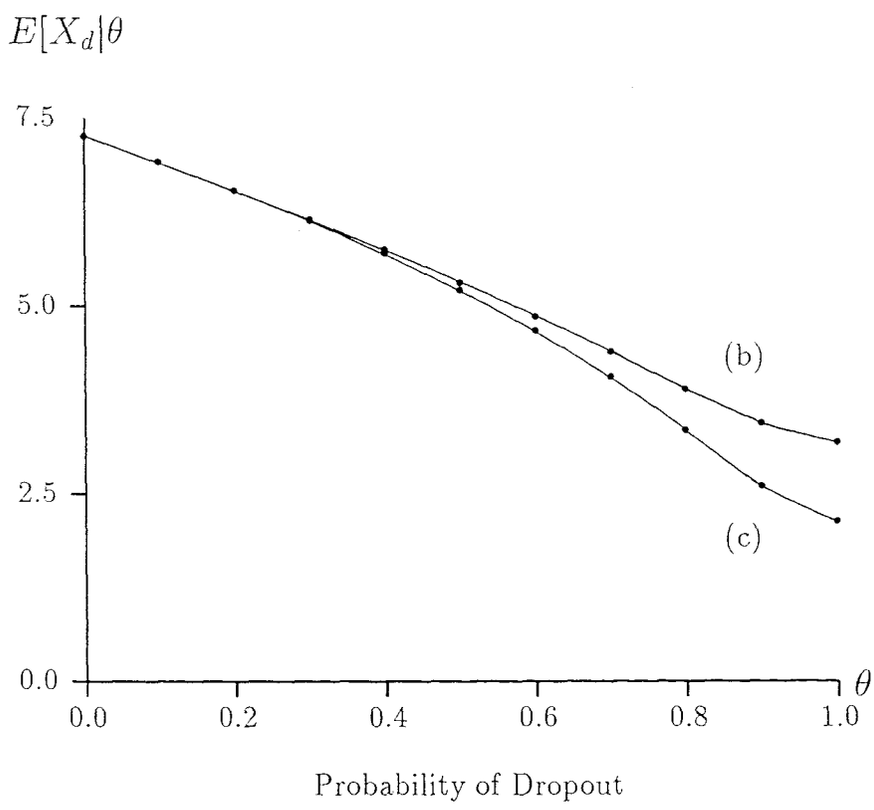


Figure 4.5: Comparisons of the graphs showing the mean of X_d, under the A-criterion, for designs (b) and (c).

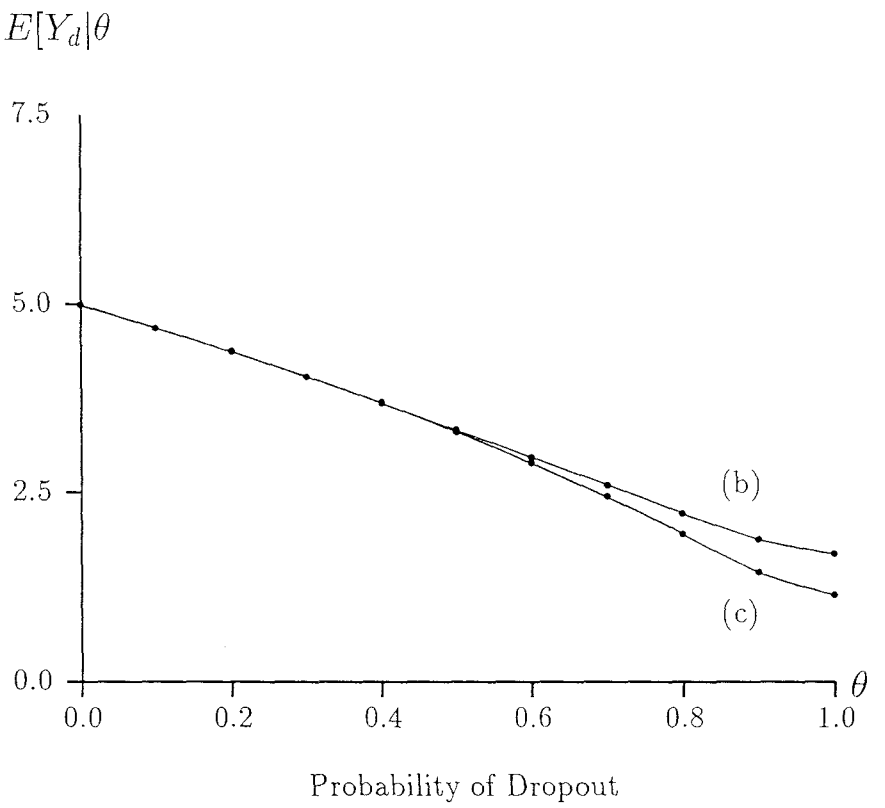


Figure 4.6: Comparisons of the graphs showing the mean of Y_d, under the A-criterion, for designs (b) and (c).

exceeds that obtained for design (b). The difference, however, is very small, never exceeding 0.01, and therefore can reasonably be considered to be negligible.

Choosing between designs (b) and (c) using the approach of Section 2.8.1 based on the A-criterion and using the evidence described above we recommend design (b) rather than design (c) for studies in which final period dropout is anticipated.

We now compare the performance of the designs using equations (2.4), (2.5), (2.6) and (2.7) and the MV-criterion. Summary measures for design (b), under the MV-criterion, have been given previously in Table 2.7. Summary measures for design (c), under the MV-criterion, are given in Table 4.6.

Table 4.6: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion. for design (c), with 16 subjects.

θ	$E[X_d\|\theta]$	$\text{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\text{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.64	0.20	4.48	0.13
0.2	6.12	0.33	4.06	0.21
0.3	5.63	0.49	3.66	0.29
0.4	5.12	0.69	3.26	0.37
0.5	4.58	0.91	2.86	0.45
0.6	3.99	1.15	2.44	0.52
0.7	3.32	1.31	1.99	0.55
0.8	2.58	1.20	1.51	0.48
0.9	1.80	0.60	1.02	0.23
1.0	1.35	0.00	0.75	0.00

Comparisons of the graphs of the mean of X_d and Y_d obtained using the MV-criterion, for $0 \leq \theta \leq 1$, for designs (b) and (c) are given in Figures 4.7 and 4.8 respectively. We observe that the mean values of X_d when $\theta < 0.4$ are marginally larger for design (c) than for design (b). This difference, however, never exceeds

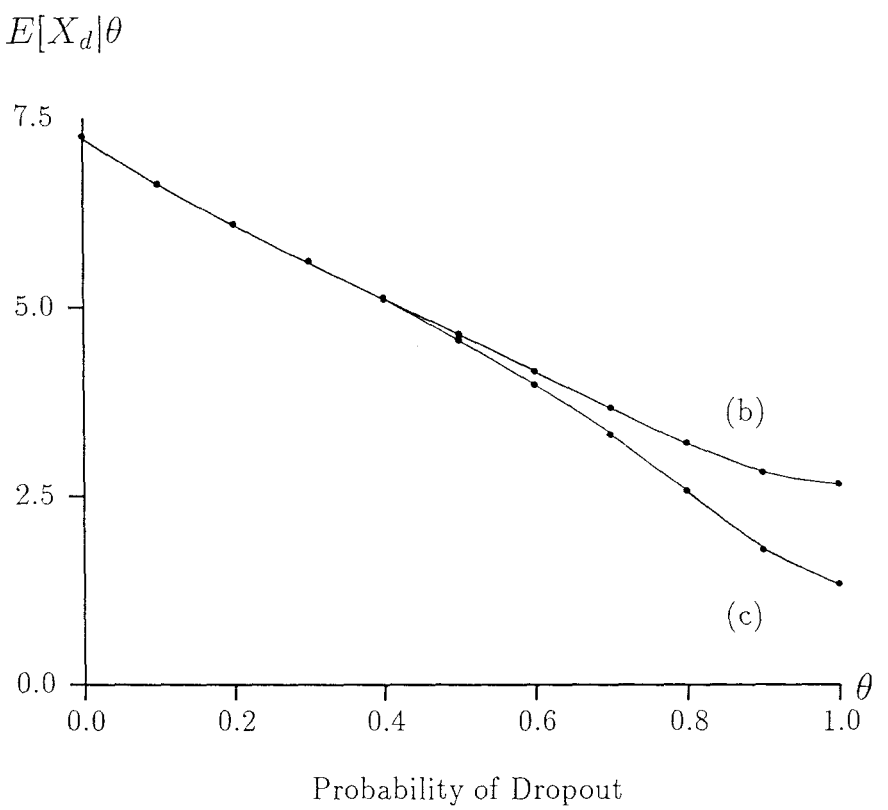


Figure 4.7: Comparisons of the graphs showing the mean of X_d, under the MV-criterion, for designs (b) and (c).

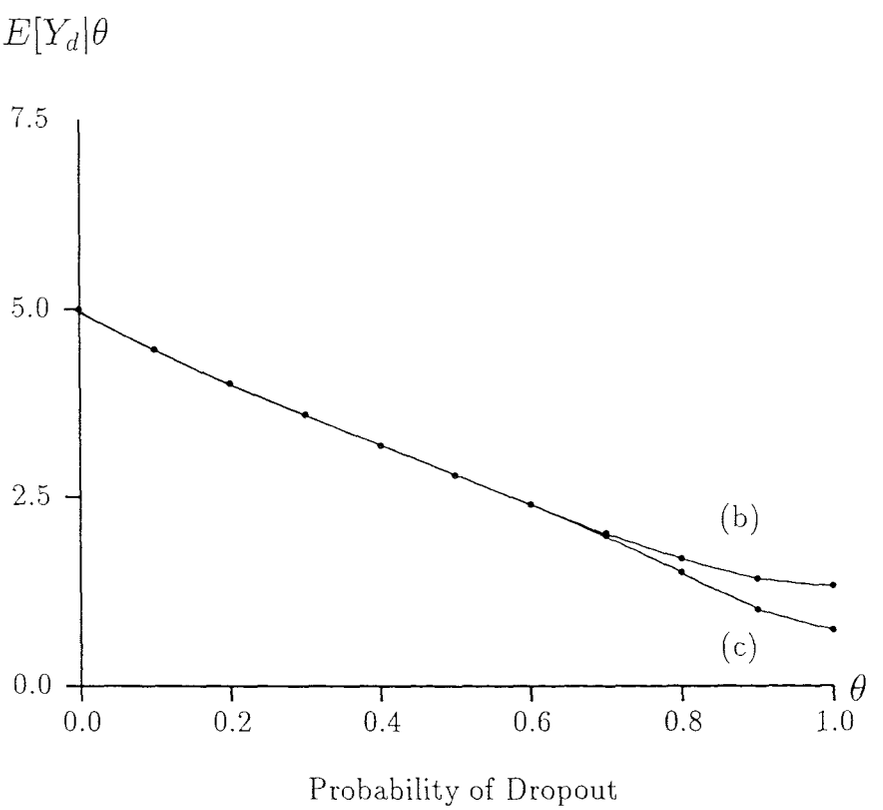


Figure 4.8: Comparisons of the graphs showing the mean of Y_d, under the MV-criterion, for designs (b) and (c).

0.01 and can therefore reasonably be considered to be negligible, particularly since the variances corresponding to these mean values are always larger for design (c). When $\theta = 0.4$ the mean values of X_d obtained for design (b) begin to exceed those obtained for design (c) with the difference between the respective mean values increasing rapidly as θ increases. The mean values of Y_d when $\theta < 0.7$ are larger for design (c). As previously, this difference is very small, never exceeding 0.07, and can therefore reasonably be considered to be negligible, particularly since the variances of the mean values obtained for design (c) always exceed those for design (b). When $\theta > 0.7$, the mean values of Y_d obtained for design (b) begin to exceed those obtained for design (c), with the magnitude of the difference between respective mean values rapidly becoming large.

When θ is small, using the design selection criteria of Section 2.8.1 with the MV-criterion, design (c) may be preferred to design (b), since the mean values of X_d and Y_d are always very slightly larger for design (c). However, given that this difference is small enough to be assumed negligible and that the corresponding variances are always considerably smaller for design (b), this slight difference for small θ is not significant enough to change the recommendation, obtained under the A-criterion, that design (b) is the preferred design.

To summarise, given an experimenter wishes to use a design having eight different treatment sequences, there should be no difficulty in employing design (b) rather than design (c), since each is formed by combining two different Williams squares. When final period dropouts may occur, the use of design (b) rather than design (c) is recommended. This is because the mean performance measures for X_d obtained using the A- criterion are always higher with correspondingly lower variances, while those obtained using the MV-criterion are only marginally smaller for small θ and higher for all other values of θ, with the variance of X_d being lower across the entire range of θ values. In addition, the values obtained for Y_d using either the A- or MV-criterion are always higher, with corresponding lower variances, when θ is large, while those obtained using either criterion are only marginally smaller for small θ,

with the variance of Y_d being lower across the entire range of possible θ values. Furthermore, our aim is to select the design which provides the greatest protection against the possibility of final period dropout. If we consider the individual performance measures X_d and Y_d obtained using either the A- or MV-criterion for each of the implementable designs associated with designs (b) and (c), we observe that the very best that can be achieved is that no subjects drop out during the final period. In this case the performance measures obtained for each design are identical. The worst situation is that every subject is lost during the final period. In this case the implemented design will be the planned design with the entire final period deleted. The performance measures obtained from the implemented design consisting of the first three periods of design (c) are considerably smaller than those obtained from the implemented design consisting of the first three periods of design (b). That is, although the best that can be achieved in terms of performance measures will be the same if either design (b) or (c) is used the worst situation is considerably poorer if design (c) is selected. Hence, better overall protection to final period dropout can be achieved using design (b) rather than design (c).

Example 4.1 demonstrates that not all eight sequence designs formed by combining two different Williams squares of side four are equally robust to final period dropout. Considering all 15 designs which can be formed, we observe that these can be separated into two categories. Those which have mean performance measures identical to those of design (b) and those which have mean performance measures identical to design (c). Each class of designs is listed in Table 4.7.

The common feature of the designs in the first category is that the two squares used to form the designs complement each other in the final two periods in the sense that neither square replicates any ordered pair of treatments appearing in the third and fourth periods of the other square. For this reason these designs are called **complementary pairs of Williams squares.**

Table 4.7: Categories of designs created from the union of the treatment sequences from two "different " Williams squares of side four.

Category I: **Complementary Pairs**

(i) and (ii)
(iii) and (iv)
(v) and (vi).

Category II: **Non-complementary Pairs**

(i) and (iii), (i) and (iv), (i) and (v), (i) and (vi)
(ii) and (iii), (ii) and (iv), (ii) and (v), (ii) and (vi)
(iii) and (v), (iii) and (vi), (iv) and (v), (iv) and (vi).

By definition, each pair of complementary squares contains eight distinct ordered pairs of treatments in the final two periods of the design, unlike the designs in the second category. In the latter category of designs there are only seven different ordered pairs of treatments in the final two periods, one of which appears in each square. Since this difference in balance is the only difference between the two categories of design, it is conjectured that the increased protection to final period dropout achieved by the complementary squares is due to the increased combinatorial balance achieved in the final two periods.

In the remainder of this section, the mean and variance of the performance measures X_d and Y_d, for a complementary pair of Williams squares are investigated under both the A- and MV-criterion. The designs considered involve up to 32 subjects. In practice, it is sometimes realistic for experiments involving more than 32 subjects to be considered. Unfortunately, despite the reductions made possible by using the results presented in Chapter 3, it would require an excessive amount of computation to obtain results for a complementary pair of Williams squares involving more than 32 subjects.

Tables 4.8.1 - 4.8.4 (given on pages 134-135) contain the mean and variance of the performance measures X_d and Y_d under the A-criterion, whilst Tables 4.9.1 -

4.9.4 (given on pages 136-137) contain the mean and variance of X_d and Y_d under the MV-criterion.

Figures 4.9 and 4.10 show how $E[X_d|\theta]$ and $E[Y_d|\theta]$ change with θ under the A-criterion for two of these designs, namely those involving 16 and 32 subjects. Similarly, Figures 4.11 and 4.12 show how the measures change with θ, under the MV-criterion for the designs involving 16 and 32 subjects. The bars shown on each of the figures represent $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$ and $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$ and have been included to give an indication of the spread of the distributions for each value of θ.

As expected, the mean performance measures increase as the number of subjects allocated to each treatment sequence is increased. The reduction in each mean performance measure appears to be fairly gradual across the entire range of values for θ. In addition, the gradient of the curves for each respective mean performance measure does not appear to change very much when the number of subjects is increased.

One important feature of a design built from a complementary pair of Williams squares is that the set of implementable designs formed by dropping one or more subjects in the final period does not contain any disconnected designs. A consequence of this is that relatively small variances are obtained for each mean performance measure. As with any planned design the variances are greater when θ is close to 0.5, and smaller when θ is either small or large.

In this section we have shown that a design to compare four treatments formed from a complementary pair of Williams squares is more robust to final period dropout than any other pair of Williams squares of side four. In addition we have conjectured that this is because only these designs have eight distinct ordered pairs of treatments in the final two periods. If this is true, increasing the number of treatment sequences in a design, in order to enable the maximum possible number of ordered pairs to appear in the final two periods, should produce a design with

Table 4.8: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for a design based on a complementary pair of Williams squares of side four and ≤ 32 subjects.

Table 4.8.1: $t = 4$, $m = 8$, $n = 1$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	3.64	0.00	2.50	0.00
0.1	3.41	0.06	2.30	0.04
0.2	3.17	0.11	2.11	0.08
0.3	2.93	0.16	1.91	0.11
0.4	2.70	0.18	1.71	0.12
0.5	2.46	0.19	1.52	0.13
0.6	2.22	0.17	1.33	0.11
0.7	2.00	0.13	1.16	0.08
0.8	1.81	0.08	1.01	0.04
0.9	1.66	0.02	0.90	0.01
1.0	1.60	0.00	0.86	0.00

Table 4.8.2: $t = 4$, $m = 8$, $n = 2$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.92	0.08	4.69	0.06
0.2	6.54	0.15	4.37	0.11
0.3	6.16	0.22	4.04	0.16
0.4	5.75	0.28	3.70	0.20
0.5	5.32	0.34	3.34	0.23
0.6	4.86	0.36	2.98	0.23
0.7	4.39	0.35	2.60	0.21
0.8	3.90	0.27	2.23	0.16
0.9	3.45	0.12	1.89	0.06
1.0	3.20	0.00	1.71	0.00

Table 4.8.3: $t = 4$, $m = 8$, $n = 3$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.91	0.00	7.50	0.00
0.1	10.41	0.10	7.07	0.07
0.2	9.89	0.19	6.63	0.14
0.3	9.35	0.29	6.16	0.21
0.4	8.77	0.37	5.67	0.26
0.5	8.16	0.44	5.16	0.31
0.6	7.51	0.50	4.63	0.33
0.7	6.82	0.51	4.07	0.32
0.8	6.07	0.45	3.50	0.26
0.9	5.30	0.25	2.93	0.13
1.0	4.80	0.00	2.57	0.00

Table 4.8.4: $t = 4$, $m = 8$, $n = 4$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	14.55	0.00	10.00	0.00
0.1	13.91	0.12	9.45	0.09
0.2	13.24	0.24	8.87	0.18
0.3	12.54	0.35	8.27	0.26
0.4	11.79	0.46	7.64	0.33
0.5	11.00	0.55	6.98	0.38
0.6	10.16	0.62	6.28	0.41
0.7	9.25	0.65	5.55	0.41
0.8	8.25	0.60	4.78	0.35
0.9	7.19	0.38	3.99	0.20
1.0	6.40	0.00	3.43	0.00

Table 4.9: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for a design based on a complementary pair of Williams squares of side four and ≤ 32 subjects.

Table 4.9.1: $t = 4$, $m = 8$, $n = 1$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	3.64	0.00	2.50	0.00
0.1	3.22	0.17	2.16	0.11
0.2	2.88	0.23	1.88	0.15
0.3	2.58	0.25	1.63	0.17
0.4	2.30	0.25	1.40	0.16
0.5	2.05	0.23	1.20	0.14
0.6	1.82	0.19	1.02	0.11
0.7	1.62	0.13	0.87	0.07
0.8	1.46	0.06	0.75	0.03
0.9	1.36	0.01	0.69	0.01
1.0	1.33	0.00	0.67	0.00

Table 4.9.2: $t = 4$, $m = 8$, $n = 2$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.27	0.00	5.00	0.00
0.1	6.63	0.20	4.47	0.14
0.2	6.11	0.30	4.02	0.21
0.3	5.62	0.37	3.60	0.26
0.4	5.14	0.43	3.19	0.29
0.5	4.66	0.46	2.80	0.29
0.6	4.17	0.46	2.41	0.27
0.7	3.68	0.42	2.03	0.22
0.8	3.21	0.29	1.69	0.14
0.9	2.82	0.10	1.43	0.04
1.0	2.67	0.00	1.33	0.00

Table 4.9.3: $t = 4$, $m = 8$, $n = 3$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.91	0.00	7.50	0.00
0.1	10.05	0.24	6.77	0.17
0.2	9.35	0.37	6.16	0.27
0.3	8.67	0.49	5.57	0.35
0.4	7.99	0.58	4.98	0.39
0.5	7.30	0.64	4.40	0.41
0.6	6.58	0.66	3.83	0.40
0.7	5.84	0.64	3.26	0.35
0.8	5.07	0.53	2.71	0.25
0.9	4.35	0.24	2.23	0.10
1.0	4.00	0.00	2.00	0.00

Table 4.9.4: $t = 4$, $m = 8$, $n = 4$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	14.55	0.00	10.00	0.00
0.1	13.47	0.28	9.09	0.21
0.2	12.59	0.46	8.30	0.34
0.3	11.72	0.61	7.53	0.44
0.4	10.84	0.73	6.77	0.50
0.5	9.93	0.81	6.01	0.52
0.6	8.99	0.84	5.26	0.51
0.7	8.01	0.83	4.50	0.46
0.8	6.98	0.73	3.75	0.35
0.9	5.93	0.40	3.05	0.17
1.0	5.33	0.00	2.67	0.00

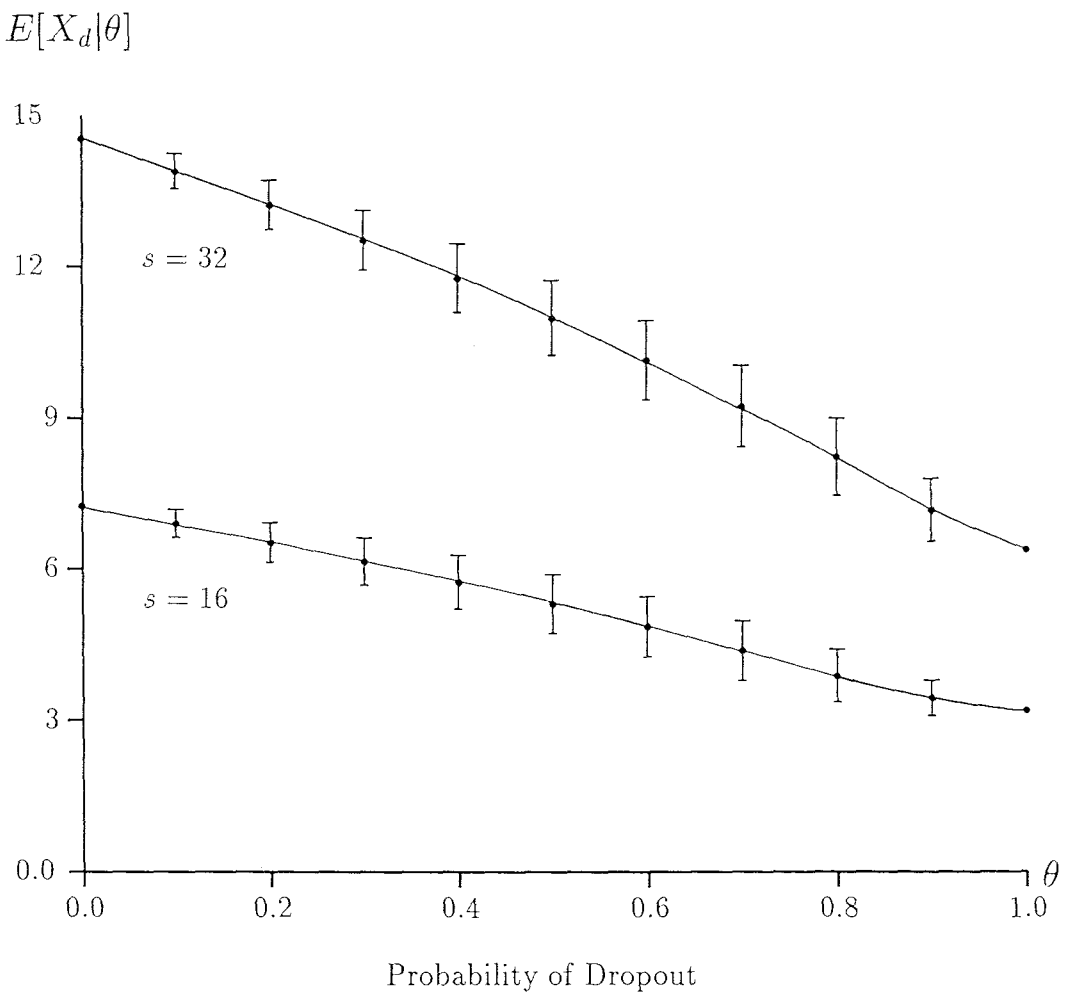


Figure 4.9: Performance for direct treatment comparisons under the A-criterion of designs based on complementary pairs of Williams squares for 16 and 32 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

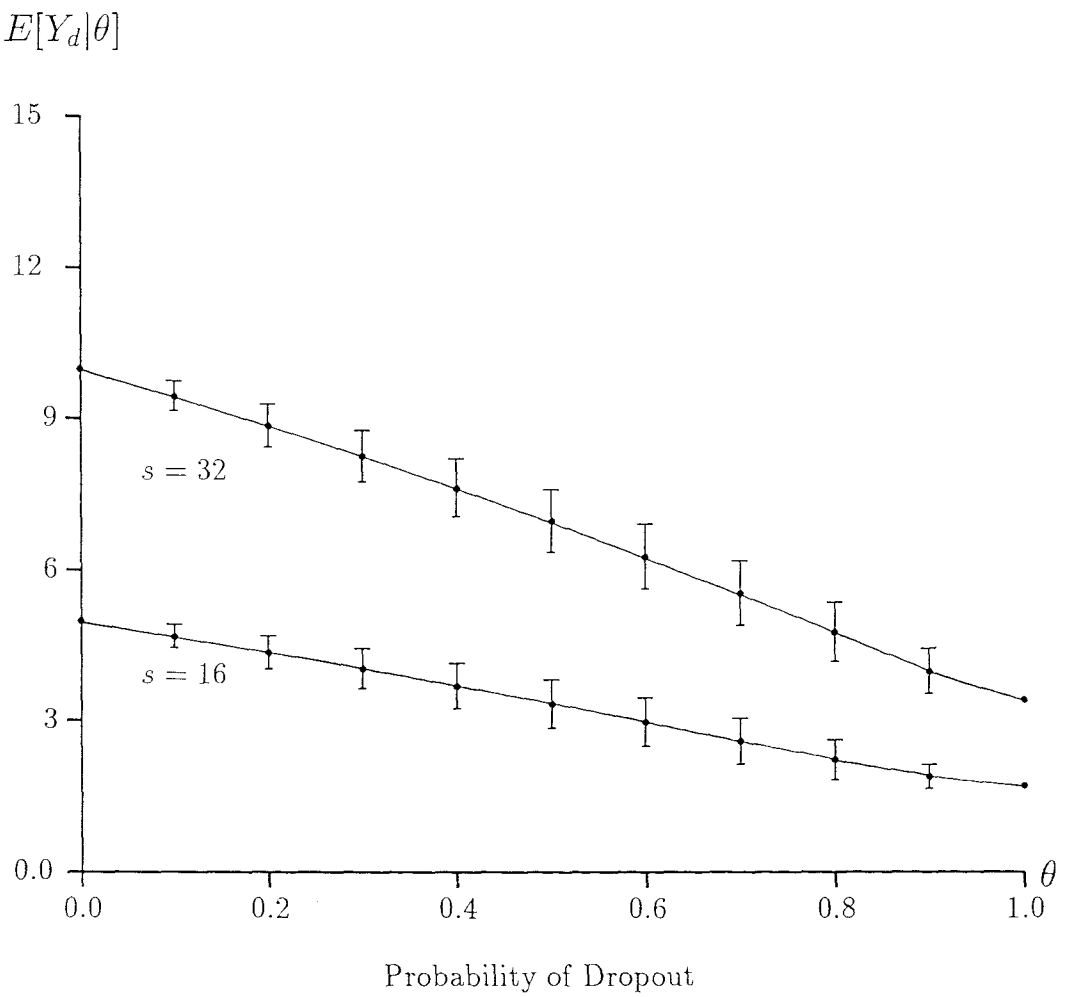


Figure 4.10: Performance for carry-over treatment comparisons under the A-criterion of designs based on complementary pairs of Williams squares for 16 and 32 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

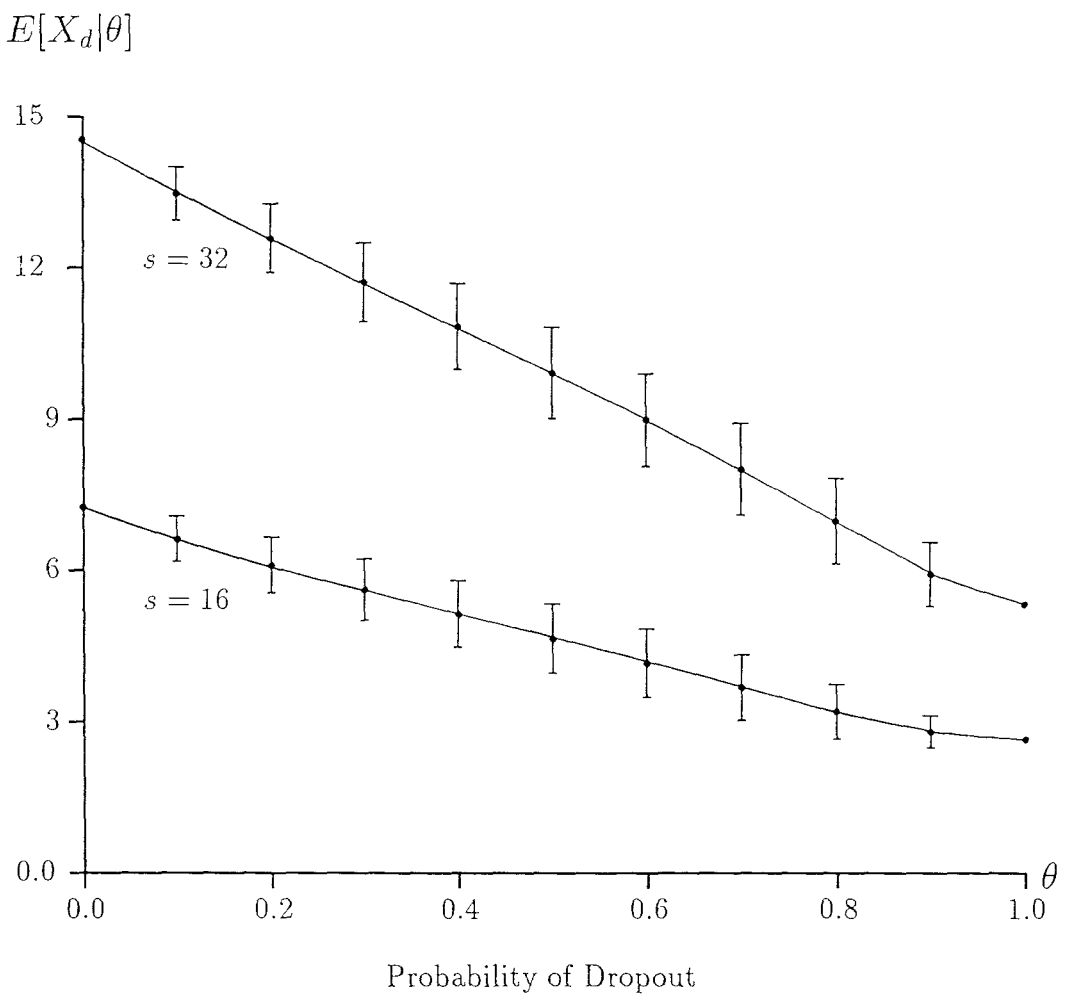


Figure 4.11: Performance for direct treatment comparisons under the MV-criterion of designs based on complementary pairs of Williams squares for 16 and 32 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

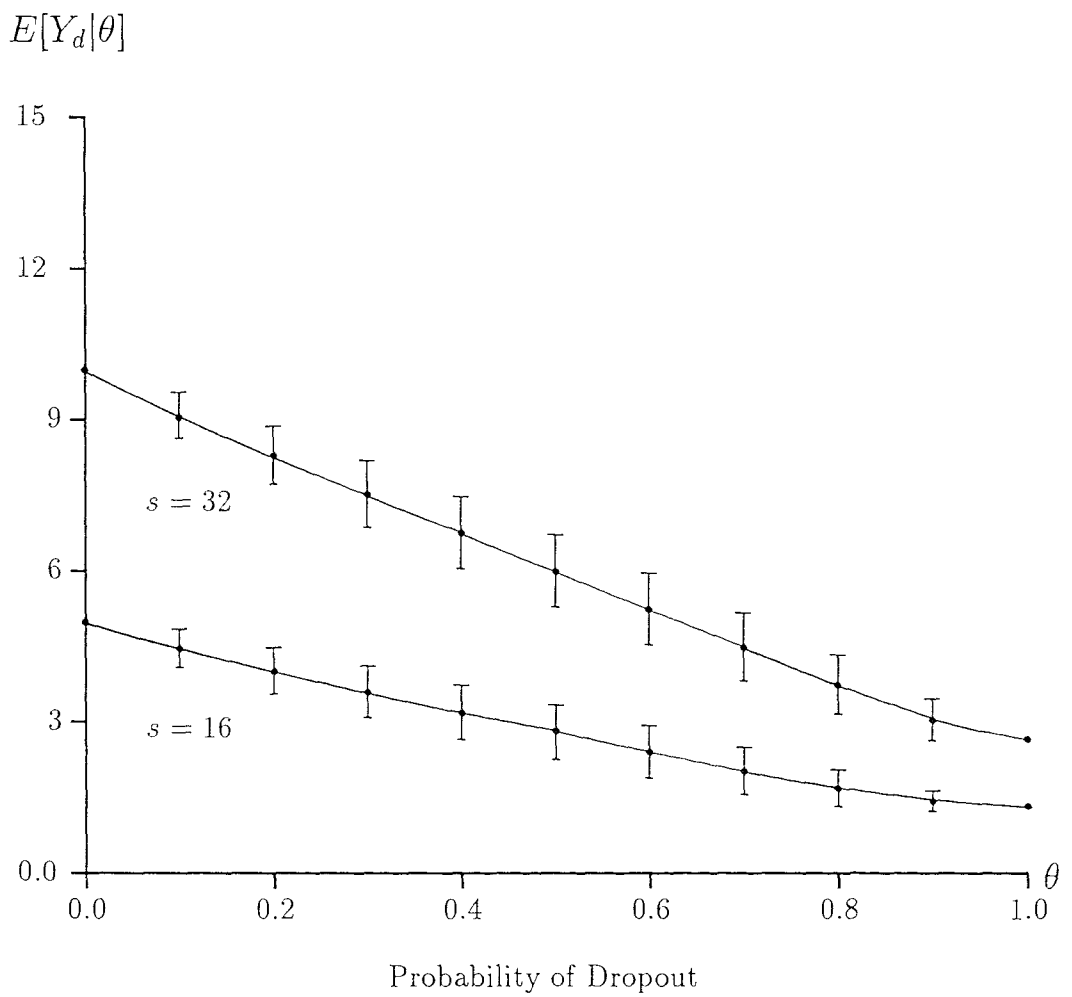


Figure 4.12: Performance for direct treatment comparisons under the MV-criterion of designs based on complementary pairs of Williams squares for 16 and 32 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

even better protection against final period dropout. Designs having this property are considered in the following section.

4.5 Mutually Orthogonal Latin Squares

For designs involving four treatments when no treatment is allowed to follow itself, there are 12 distinct ordered pairs of treatments. If we consider the Williams squares given in Table 4.4, we observe that it is not possible to combine three of these to achieve a twelve sequence design in which each of the ordered pairs of treatments occurs once in the third and fourth periods. However, this feature can be achieved by using the particular complete set of mutually orthogonal Latin squares of side four given in Table 4.10.

Table 4.10: Complete set of balanced mutually orthogonal Latin squares of side four with treatment labels 0, 1, 2, and 3.

0	1	2	3
1	0	3	2
2	3	0	1
3	2	1	0
0	3	1	2
1	2	0	3
2	1	3	0
3	0	2	1
0	2	3	1
1	3	2	0
2	0	1	3
3	1	0	2

A set of Latin squares is mutually orthogonal if every pair of squares is orthogonal, that is when superimposed on each other the treatment labels of one square

occur once with each of the labels of the other square. Furthermore, since there can only be at most $(t-1)$ mutually orthogonal Latin squares of side t, a collection of $(t-1)$ Latin squares which are mutually orthogonal is known as a complete set of mutually orthogonal Latin squares of side t.

The complete set of mutually orthogonal Latin squares given in Table 4.10 possesses the additional property of balance, that is each treatment is preceded equally often by every other treatment but never by itself. Consequently, it provides a twelve sequence design in which each of the ordered pairs of treatments occurs once in the third and fourth periods. Note that, not all complete sets of mutually orthogonal Latin squares can be combined to give designs which are balanced. For this reason, any complete set of mutually orthogonal Latin squares which can be combined to form a design which is balanced shall be referred to as **balanced mutually orthogonal Latin squares.**

In this section the mean and variance of the performance measures X_d and Y_d, under both the A- and MV-criterion, for designs formed from the sequences of the complete set of balanced mutually orthogonal Latin squares of side four given in Table 4.10, are presented. Since this is a design involving 12 different treatment sequences, the set of implementable designs requiring evaluation in order to obtain the mean performance measures is large, even when the number of subjects allocated to the sequences is small. It is possible, using the theory presented in Chapter 3, to reduce the computational burden sufficiently to obtain the mean performance measures for designs involving 12 and 24 subjects. Unfortunately, despite the enormous savings achieved by applying the results from combinatorial theory, it still requires an excessive amount of computation to obtain mean performance measures for designs formed from the sequences of a set of balanced mutually orthogonal Latin squares involving more than 24 subjects.

Tables 4.11.1 and 4.11.2 contain the mean and variance of the performance measures X_d and Y_d, under the A-criterion, whilst Tables 4.12.1 and 4.12.2 contain the mean and variance of the performance measures X_d and Y_d under the MV-criterion.

Table 4.11: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for a design based on a complete set of balanced mutually orthogonal Latin squares of side four and ≤ 24 subjects.

Table 4.11.1: $t = 4$, $m = 12$, $n = 1$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	5.45	0.00	3.75	0.00
0.1	5.17	0.07	3.51	0.04
0.2	4.89	0.17	3.28	0.07
0.3	4.61	0.24	3.05	0.09
0.4	4.36	0.23	2.84	0.09
0.5	4.13	0.17	2.64	0.09
0.6	3.91	0.13	2.44	0.08
0.7	3.69	0.09	2.26	0.06
0.8	3.49	0.05	2.09	0.04
0.9	3.32	0.02	1.95	0.01
1.0	3.24	0.00	1.89	0.00

Table 4.11.2: $t = 4$, $m = 12$, $n = 2$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.91	0.00	7.50	0.00
0.1	10.43	0.09	7.08	0.06
0.2	9.94	0.16	6.67	0.11
0.3	9.47	0.20	6.26	0.15
0.4	9.00	0.23	5.86	0.16
0.5	8.53	0.23	5.47	0.17
0.6	8.07	0.22	5.08	0.15
0.7	7.61	0.19	4.70	0.13
0.8	7.17	0.14	4.34	0.09
0.9	6.74	0.07	3.99	0.04
1.0	6.48	0.00	3.78	0.00

Table 4.12: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for a design based on a complete set of balanced mutually orthogonal Latin squares of side four and ≤ 24 subjects.

Table 4.12.1: $t = 4$, $m = 12$, $n = 1$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	5.45	0.00	3.75	0.00
0.1	4.95	0.16	3.33	0.10
0.2	4.56	0.24	3.02	0.12
0.3	4.25	0.26	2.77	0.12
0.4	4.00	0.22	2.54	0.10
0.5	3.78	0.16	2.35	0.08
0.6	3.59	0.09	2.19	0.06
0.7	3.43	0.05	2.05	0.03
0.8	3.32	0.02	1.95	0.01
0.9	3.25	2.6×10^{-3}	1.90	1.7×10^{-3}
1.0	3.24	0.00	1.89	0.00

Table 4.12.2: $t = 4$, $m = 12$, $n = 2$, $p = 4$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.91	0.00	7.50	0.00
0.1	10.09	0.19	6.82	0.13
0.2	9.49	0.25	6.31	0.18
0.3	8.96	0.29	5.86	0.20
0.4	8.47	0.29	5.44	0.20
0.5	8.00	0.27	5.05	0.19
0.6	7.57	0.23	4.68	0.16
0.7	7.17	0.17	4.35	0.11
0.8	6.81	0.09	4.05	0.06
0.9	6.56	0.02	3.84	0.01
1.0	6.48	0.00	3.78	0.00

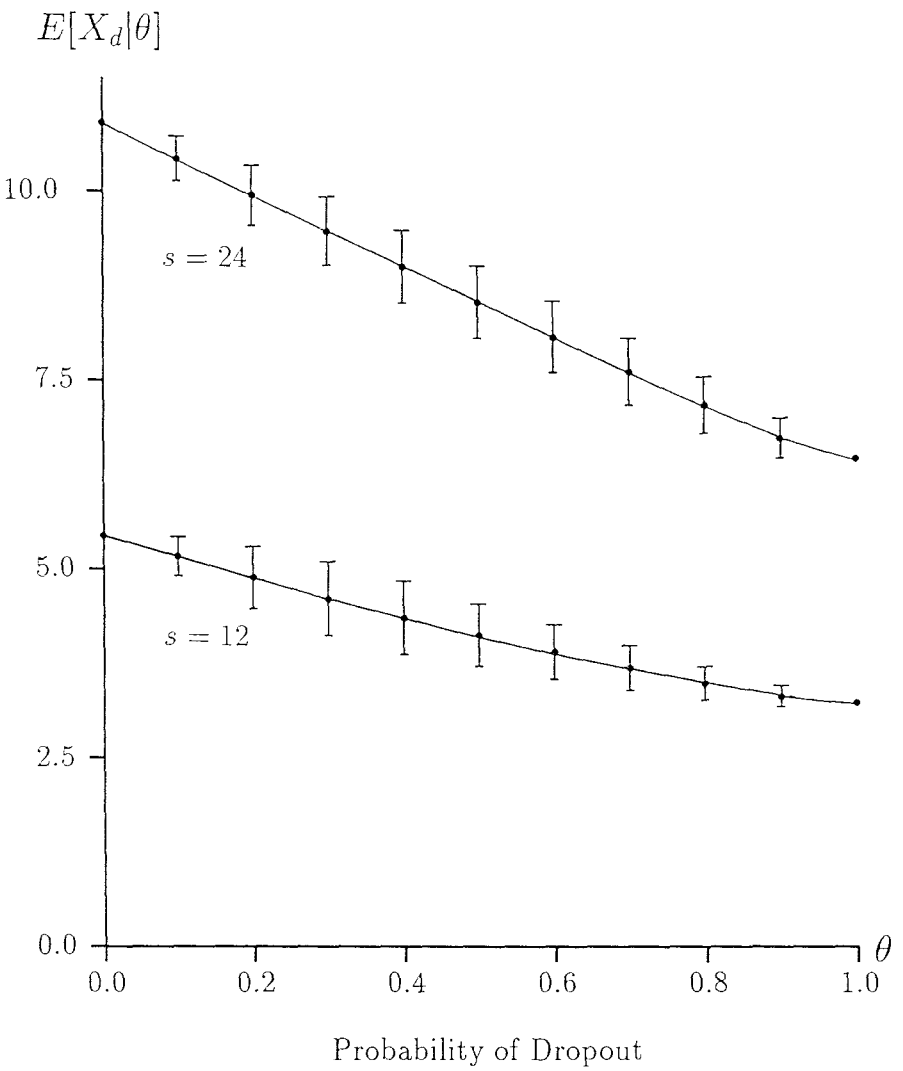


Figure 4.13: Performance for direct treatment comparisons under the A-criterion of designs based on balanced mutually orthogonal Latin squares for 12 and 24 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

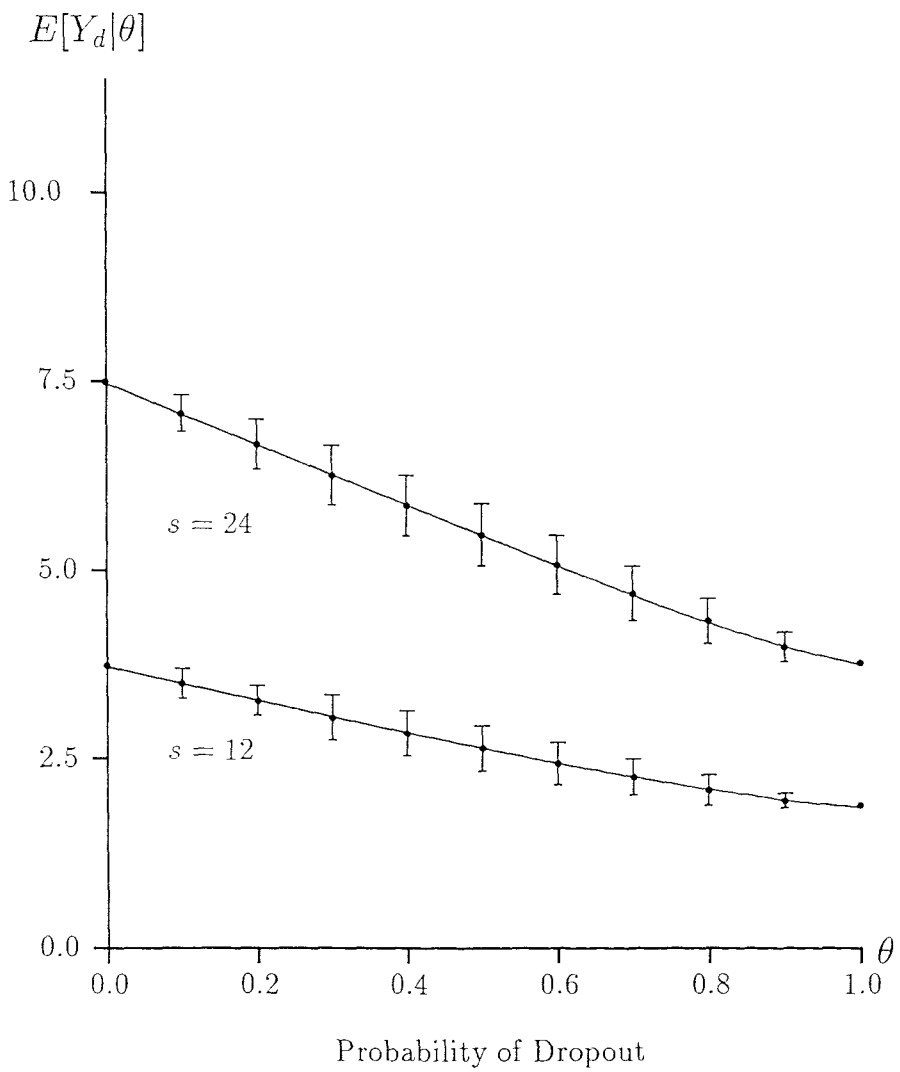


Figure 4.14: Performance for carry-over treatment comparisons under the A-criterion of designs based on balanced mutually orthogonal Latin squares for 12 and 24 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

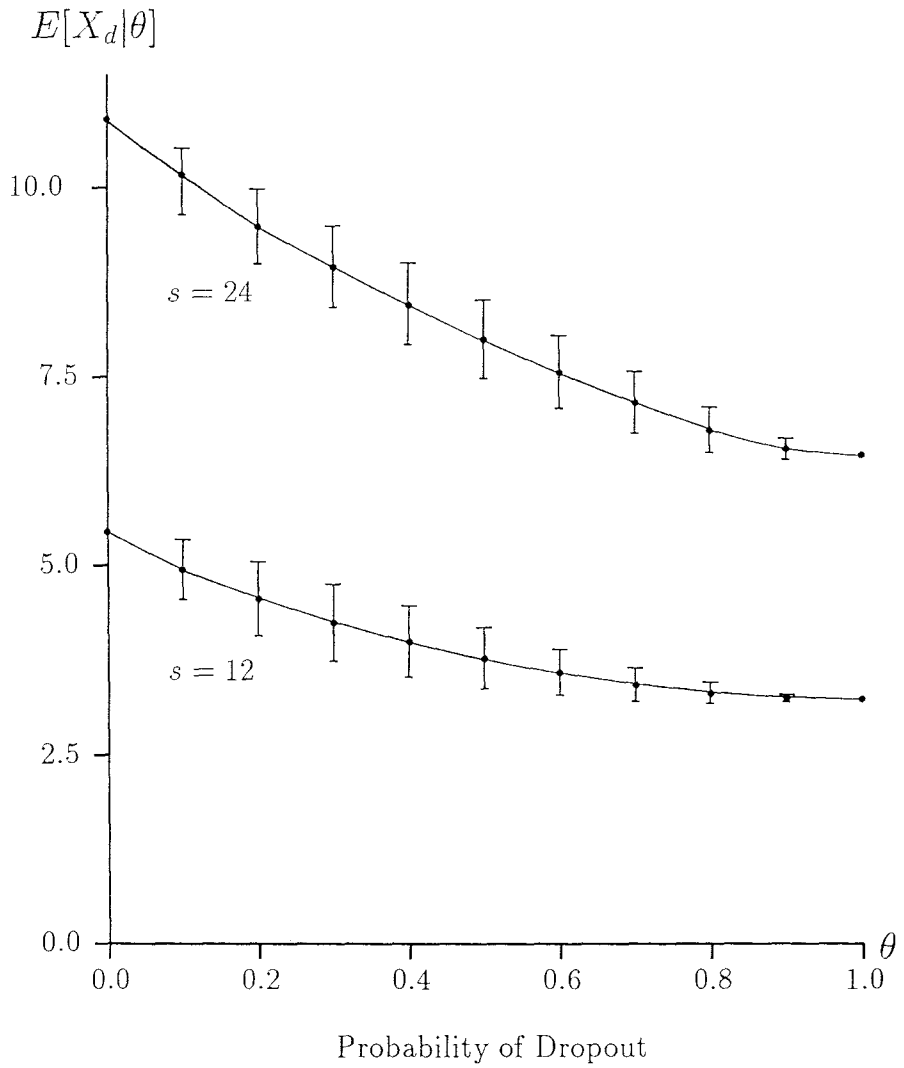


Figure 4.15: Performance for direct treatment comparisons under the MV-criterion of designs based on balanced mutually orthogonal Latin squares for 12 and 24 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

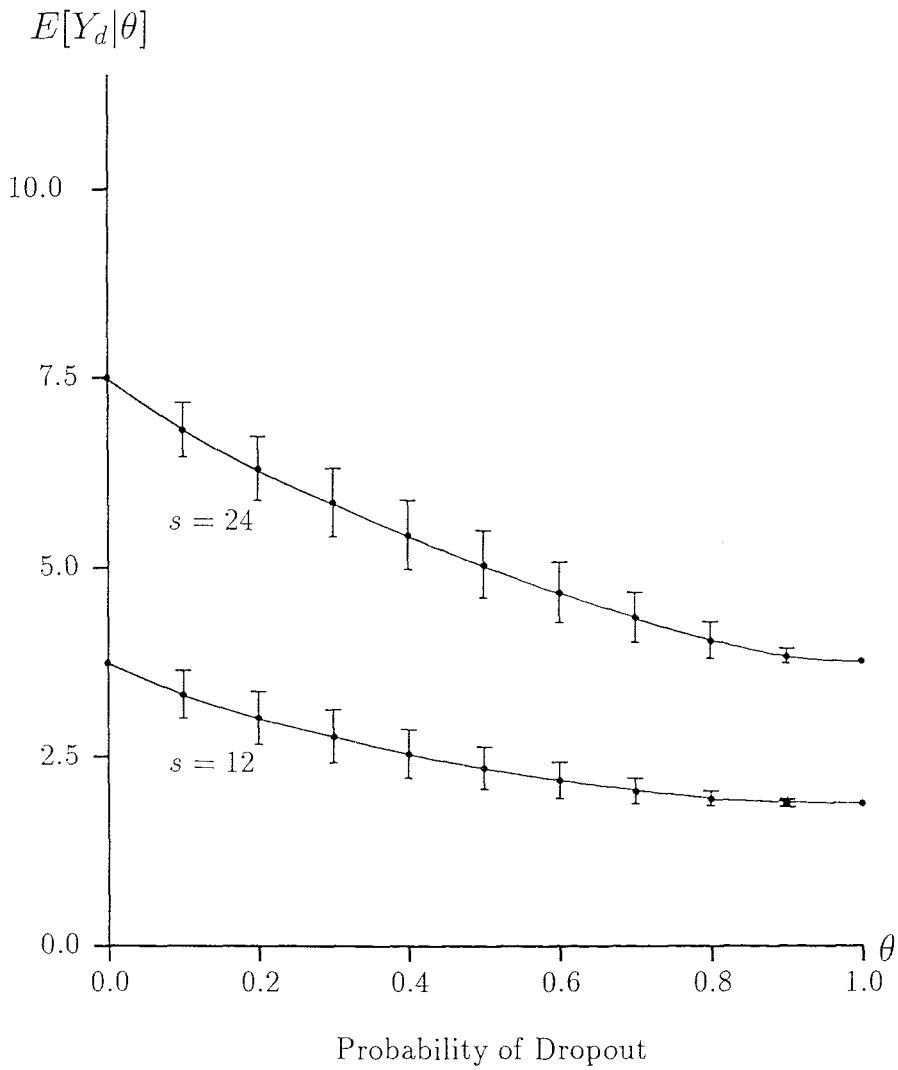


Figure 4.16: Performance for carry-over treatment comparisons under the MV-criterion of designs based on balanced mutually orthogonal Latin squares for 12 and 24 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

Figures 4.13 and 4.14 show how $E[X_d|\theta]$ and $E[Y_d|\theta]$ change with θ for each design, under the A-criterion, while Figures 4.15 and 4.16 indicate how $E[X_d|\theta]$ and $E[Y_d|\theta]$ change with θ for each design, under the MV-criterion. The bars shown on each graph represent $E[X_d|\theta] \pm \sqrt{Var[X_d|\theta]}$ and $E[Y_d|\theta] \pm \sqrt{Var[Y_d|\theta]}$ respectively. These have been included to give an indication of the spread of the distributions for the associated value of θ.

The important features of this design are that the probability of implementing a disconnected design is zero, the mean values obtained under the A- or MV-criteria for the performance measures X_d and Y_d have very small variances irrespective of the value of θ, and the respective mean performance measures obtained under each criterion do not differ very much, suggesting that the individual variances of the treatment comparisons for each potentially implementable design do not differ greatly. Each of these qualities is highly desirable in any planned design.

4.6 Comparison of different designs

One of the most important features of the design assessment procedures proposed in this thesis is their ability to compare the performance of competing designs subject to some fixed probability θ of final period dropout. These comparisons can then be used to make recommendations concerning the appropriate choice of design for different experimental situations. In the previous three sections, the robustness to final period dropout of designs involving 4, 8, and 12 treatment sequences have been investigated. In this section, we compare the performance of three of the designs so that appropriate recommendations can be made concerning their use. The designs, labelled (a), (b) and (c) are as follows:

Design (a) The single Williams square $d(4,4,6,4)$ with treatment labels 0, 1, 2 and 3 and initial treatment sequence (0 1 3 2), that is design (i) of Table 4.4 involving 24 subjects.

Design (b) The complementary pair of Williams squares $d(4, 8, 3, 4)$ with treatment labels 0, 1, 2 and 3 and initial treatment sequences (0 1 3 2) and (0 3 1 2). This is design (b) of Example 4.1 involving 24 subjects.

Design (c) A complete set of balanced mutually orthogonal Latin squares $d(4, 12, 2, 4)$ with treatment labels 0, 1, 2 and 3. This is the design formed from the sequences given in Table 4.10 involving 24 subjects.

Assume first that we wish to compare the performance of the designs in estimating all the pairwise direct and first-order carry-over treatment effects under the A-criterion. Summary measures for each design have been given in Tables 4.1.6, 4.8.3 and 4.11.2, respectively. Comparisons of the graphs of the mean of X_d and Y_d against θ, $0 \leq \theta \leq 1$, for each of these designs, are given in Figures 4.17 and 4.18 respectively. We observe that, although all three designs have identical mean performance measures when $\theta = 0.0$, the measures begin to diverge as θ increases. Further design (c) consistently gives the highest values and design (a) the lowest. Design (a) yields particularly poor mean performance measures when θ is large because the set of implementable designs produced from this particular planned design contains a number of disconnected designs. This does not happen for designs (b) or (c).

Using the design selection criteria of Section 2.8.1 leads to the choice of design (c) for any probability of final period dropout $0 \leq \theta \leq 1$. This is because it does not give rise to any disconnected implementable designs and the study shown in Figures 4.17 and 4.18 indicates that

$$\mathrm{E}[X_{d_c}|\theta] \geq \mathrm{E}[X_{d_b}|\theta] \geq \mathrm{E}[X_{d_a}|\theta] \text{ and } \mathrm{E}[Y_{d_c}|\theta] \geq \mathrm{E}[X_{d_b}|\theta] \geq \mathrm{E}[Y_{d_a}|\theta]$$

for $\theta = 0, 0.1, \ldots, 1.0$..

Similarly, if we compare the respective variances for X_d and Y_d for each design from Tables 4.1.6, 4.8.3 and 4.11.2 we observe that

$$\mathrm{Var}[X_{d_c}|\theta] \leq \mathrm{Var}[X_{d_b}|\theta] \leq \mathrm{Var}[X_{d_a}|\theta] \text{ and } \mathrm{Var}[Y_{d_c}|\theta] \leq \mathrm{Var}[Y_{d_b}|\theta] \leq \mathrm{Var}[Y_{d_a}|\theta]$$

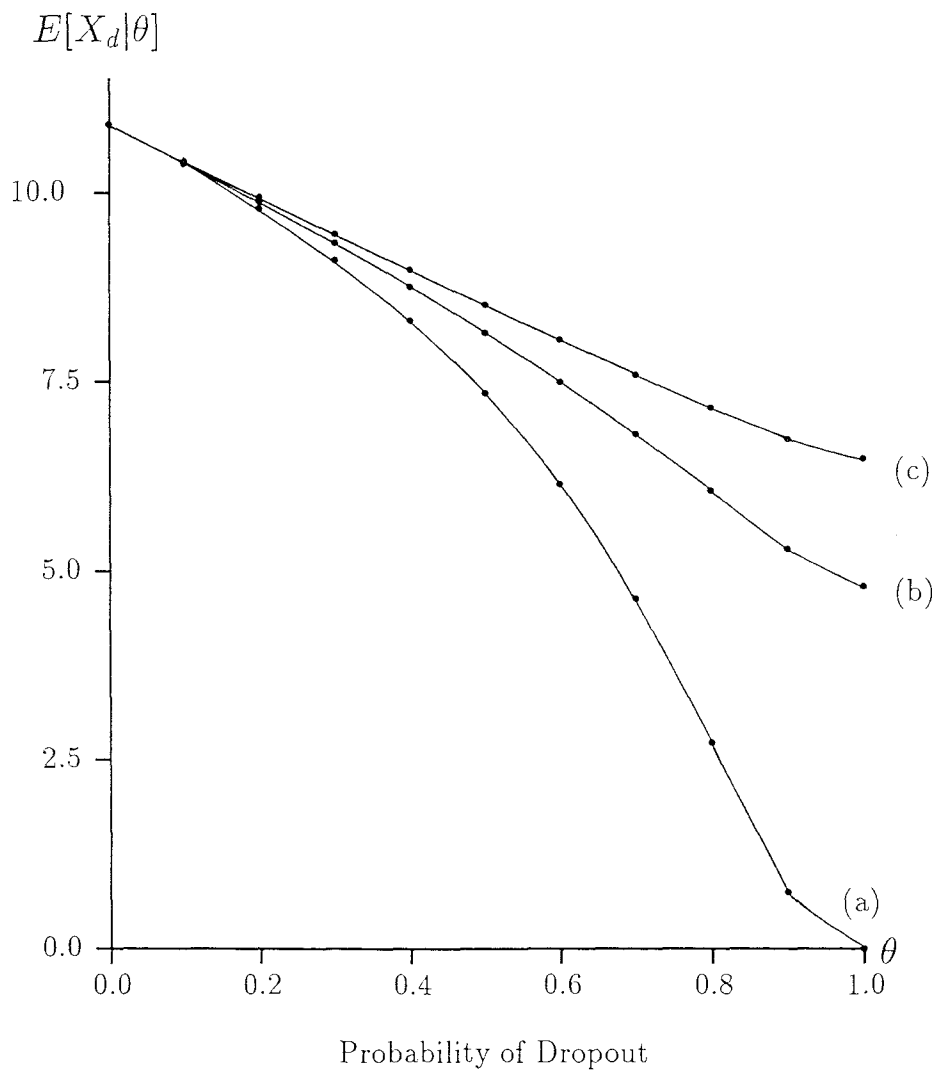


Figure 4.17: Comparisons of the graphs showing the mean of X_d, under the A-criterion, for designs (a), (b) and (c).

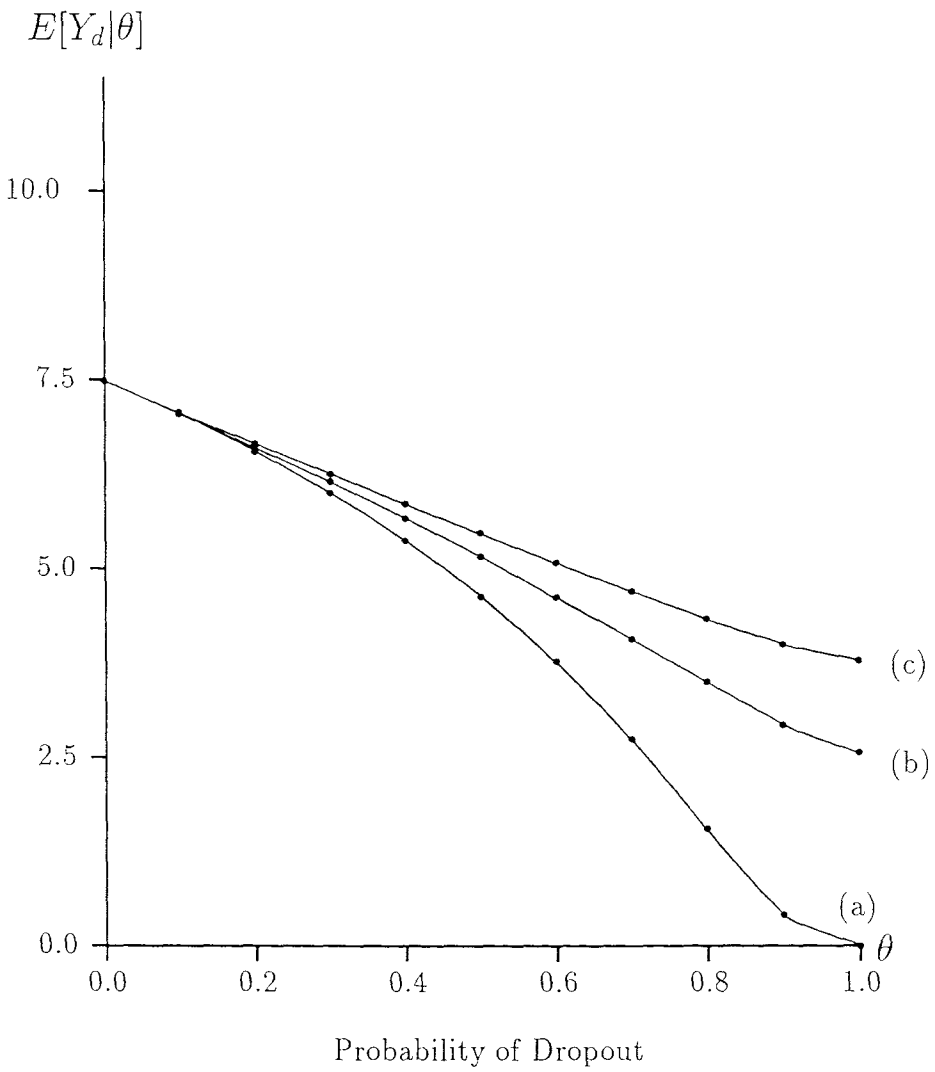


Figure 4.18: Comparisons of the graphs showing the mean of Y_d, under the A-criterion, for designs (a), (b) and (c).

for $\theta = 0, 0.1, \ldots, 1.0$.

Alternatively, we can compare all the pairwise direct and first-order carry-over treatment effects under the MV-criterion. The corresponding summary measures for each design are given in Tables 4.2.6, 4.9.3 and 4.12.2. Comparisons of the graphs of the mean of X_d and Y_d against θ, $0 \leq \theta \leq 1$, for each of these designs, are given in Figures 4.19 and 4.20.

Each comparison again clearly demonstrates the superior performance of design (c), the complete set of balanced mutually orthogonal Latin squares, across the entire range of possible θ values. This conclusion is reinforced by applying the design selection criteria of Section 2.8.1 which again leads to selection of design (c), since the mean values of X_d and Y_d are always the largest, with the correspondingly lowest variances, for any θ, $0 \leq \theta \leq 1$.

To summarise, when choosing between the three planned designs considered in this section we recommend the use of design (c), the complete set of balanced mutually orthogonal Latin squares, rather than either of the other designs, provided subject numbers permit. Design (c) has a zero probability of producing a disconnected implemented design and the mean values of X_d and Y_d are the largest with correspondingly lower variances, for all values of $\theta > 0$. The properties hold whether we use the A- or MV-criterion when calculating the performance measures. In addition, we strongly recommend that design (a), the single Williams square, is avoided because its use may give rise to a disconnected design and because it is acutely sensitive to the probability of final period dropout. These conclusions appear to be directly opposed to comments made by Jones and Kenward (1989, p 199) who state that "the loss of subjects from the complete set [of balanced mutually orthogonal Latin squares] is likely to be more damaging as its combinatorial structure is more complex." It is not entirely clear, however, whether their comments related to the problem of subjects dropping out of a study part way through the trial or to the loss of subjects for the entire trial leading to the loss of an entire treatment sequence.

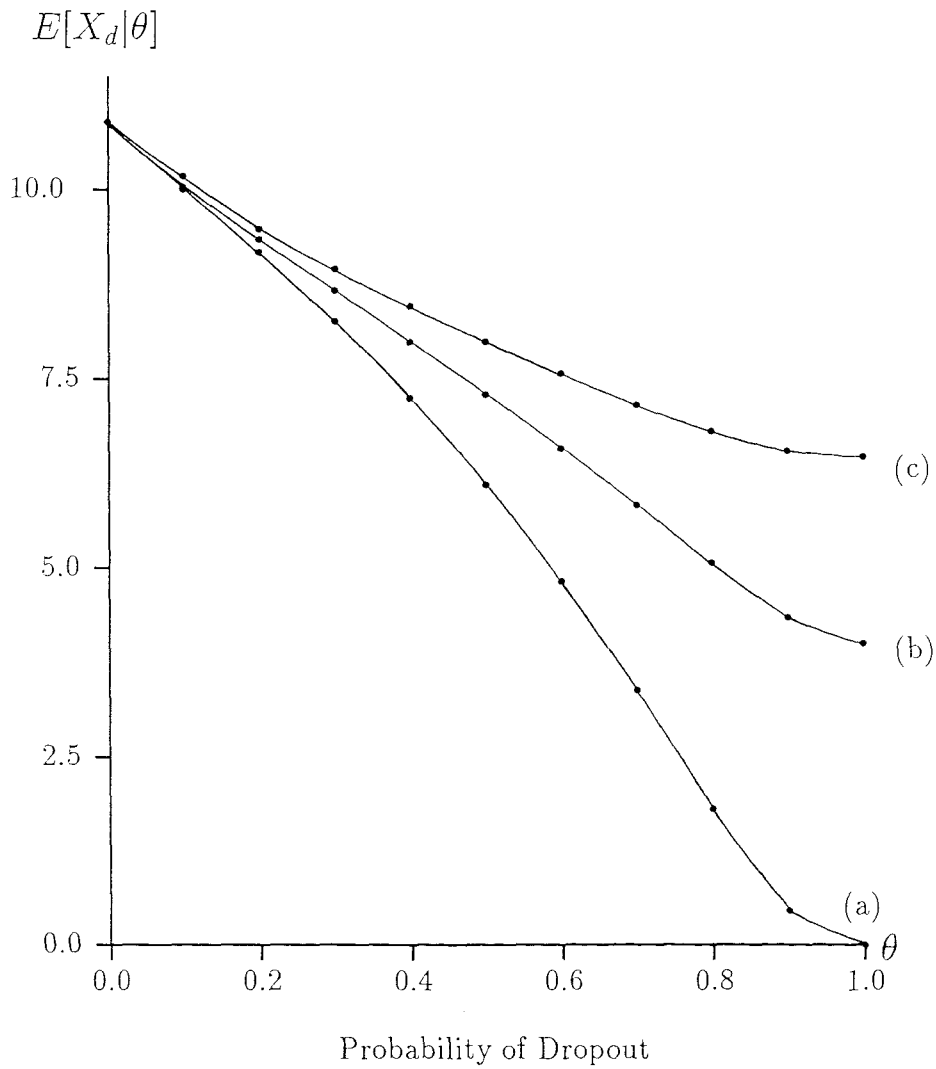


Figure 4.19: Comparisons of the graphs showing the mean of X_d, under the MV-criterion, for designs (a), (b) and (c).

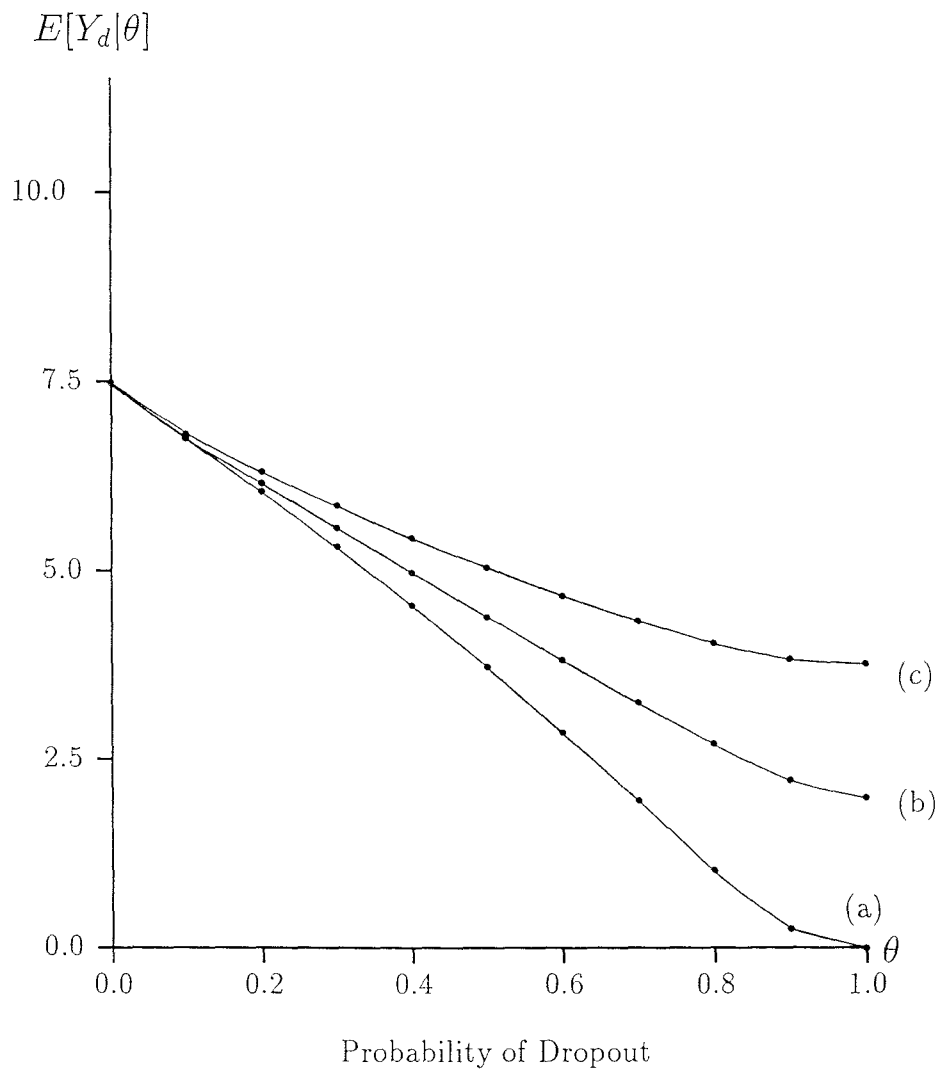


Figure 4.20: Comparisons of the graphs showing the mean of Y_d, under the MV-criterion, for designs (a), (b) and (c).

4.7 Discussion

In this chapter a study of the performance subject to final period dropout of a selection of four treatment, four period, uniform balanced designs has been presented. This has shown that although each of these designs is equally good when the probability of final period dropout is not considered, they are not equally robust to final period dropout.

We have shown that the use of a design in which equal numbers of subjects are assigned to the treatment sequences of a complementary pair of Williams squares of side four is more robust to final period dropout than the currently favoured design which employs replications of the sequences from a single square. Furthermore, a design formed from the sequences of a complete set of balanced mutually orthogonal Latin squares is more robust to final period dropout then replicates of either a single Williams square or complementary pair of Williams squares. The increased robustness is achieved at no extra cost in subject numbers. The only penalty incurred is that the number of distinct treatment sequences is increased from four to either eight or twelve. We should not forget that when θ is large the mean values obtained for each performance measure are considerably smaller than those of the planned design. An experiment should only proceed if the mean performance measures for the anticipated level of dropout are acceptably large. If not, an alternative design involving more subjects or fewer treatment periods should be considered.

We recommend that designs formed from replicates of a single Williams square are avoided whenever possible because there is always a chance, albeit reasonably small when the number of subjects to be assessed is large, that the implemented experiment is disconnected. In addition, the mean performance measures are not as high as can be achieved using alternative designs. If the maximum number of treatment sequences that can be employed is eight, we recommend that a complementary pair of Williams squares is used since this is more robust to final period dropout than any other combination of Williams squares. Finally, if it is possible

to employ 12 different treatment sequences we recommend that a design based on a complete set of balanced mutually orthogonal Latin squares is used because, for the sizes of experiment evaluated thus far, we have found that it is the most robust to final period dropout. Furthermore, combinatorial considerations suggest that this will continue to be the case for experiments involving larger numbers of subjects. The only disadvantage in using this design is that there is an increased chance that a sequence is incorrectly administered due to human error. This is the main objection to the use of designs involving a large number of different treatment sequences.

It has not been possible to present studies of designs involving more than 12 treatment sequences, such as designs formed by combining four, five or even all six Williams squares of side four. This is because of the excessive amount of computation which would be required, despite the application of the results presented in Chapter 3. It is not clear, however, that increasing the number of treatment sequences beyond 12 will lead to designs which will be more robust to final period dropout than a complete set of balanced mutually orthogonal Latin squares since this design already contains all 12 distinct ordered pairs of treatments in the final two periods. Any design having greater than 12 treatment sequences will have to contain some of the ordered pairs of treatments in the final two periods more than once. Therefore, unless this replication is in equal numbers, such as every ordered pair twice, the resultant design may not be as robust to final period dropout as a complete set of balanced mutually orthogonal Latin squares of corresponding size. This is, however, only speculation and we need to demonstrate whether or not this is in fact true by means of an actual example. Even if this proves to be the case designs formed from the treatment sequences of balanced mutually orthogonal Latin squares can only be applied in experiments in which the number of subjects available is some multiple of 12. When this is not the case a design involving 16 treatment sequences may be preferable to an eight sequence design such as a complementary pair of Williams squares. This needs to be investigated further.

Another interesting fact worthy of further study is that the 24 sequence design

formed by combining all six different Williams squares of side four consists of the same 24 treatment sequences as two sets of balanced mutually orthogonal Latin squares which are isomorphic under a permutation of the treatment labels. It would be interesting to examine whether or not such a design is any more robust to final period dropout than a single set of balanced mutually orthogonal Latin squares with two subjects allocated to each treatment sequence.

There is scope, when the computational difficulties can be overcome, for further investigation of four treatment, four period designs. In the next chapter, however we turn our attention to examining the performance of three treatment, three period designs subject to final period dropout.

Chapter 5

Three Treatment, Three Period Designs

5.1 Introduction

In Chapter 4 the important practical case of four treatment, four period cross-over experiments was studied. In this chapter designs for another common multi-period experiment, namely three treatments and three periods are investigated, when the experiments are subject to final period dropout.

We begin by examining the performance of uniform balanced designs based on a pair of Williams squares of side three. In Section 5.3 results from the literature on the optimality of uniform balanced designs, introduced in Section 4.2, are briefly reviewed for experiments in which dropouts do not occur. Some of the limitations of these designs are identified in order to determine additional designs whose robustness to final period dropout is then examined. In Section 5.4, a comparison is made of the different designs and recommendations are given on design selection. Finally, a numerical study is made to determine if altering the treatments administered in the final period of a pair of Williams squares produces an increase in robustness to final period dropout.

5.2 Uniform Balanced Designs

Williams (1949) established that, when the number of treatments is odd, a uniform balanced design can be constructed by using two particular Latin squares, usually referred to as a pair of Williams squares. For example, the pair of Williams squares, for $t = 3$ and treatment labels 0, 1 and 2, with initial treatment sequences (0 1 2) and (0 2 1) is given in Figure 5.1. Note that, when $t = 3$, the pair of Williams squares is a complete set of balanced mutually orthogonal Latin squares of side three.

Figure 5.1: Pair of Williams squares design for $t = 3$.

0	1	2
1	2	0
2	0	1
0	2	1
1	0	2
2	1	0

For experiments in which dropouts do not occur, Hedayat and Afsarinejad (1975) considered the construction of cross-over designs in which the property of balance is achieved using the minimum possible number of subjects in the experiment. They showed that, when $t = 3$, it is not possible to construct a balanced design using only three subjects. When using six subjects, however, they showed that balance could be achieved using the six treatment sequences from the pair of Williams squares shown in Figure 5.1. Since this design is a uniform balanced design, it is universally optimal over the class of all uniform designs in which $p = t = 3$ for the estimation of the direct and first-order carry-over treatment effects, when dropouts are not considered. For these reasons, and the fact that the design has no obvious competitors, it is a design frequently employed in cross-over studies.

The purpose of this section is to describe the results of an investigation into the robustness to dropouts of designs in which an equal number of subjects is al-

located to each of the treatment sequences of a pair of Williams squares of side three. As in Chapter 4, the designs are assessed by considering the variance of the estimated pairwise comparisons for the direct and first-order carry-over treatment effects when final period dropout may occur. The investigation was carried out for designs involving a maximum of 36 subjects.

There are six different sizes of design to consider, namely the designs in which n, the number of subjects allocated to each treatment sequence takes each of the values $1, \ldots, 6$.

An important feature of planned designs formed from replicates of the sequences in Figure 5.1 is that, in each case, the set of implementable designs requiring evaluation does not contain any disconnected designs. The probability of implementing a disconnected design in each case is, therefore, zero. Examining the pair of Williams squares in Figure 5.1 we observe that, in common with the complete set of balanced mutually orthogonal Latin squares of side four, it contains each of the distinct ordered pairs of treatments in the final two periods. In this case, there are six distinct ordered pairs of treatments which can be formed when no treatment is allowed to follow itself.

For each design, Tables 5.1.1 - 5.1.6 (given on pages 163-165) contain the mean and variance of the performance measures X_d and Y_d, under the A-criterion, whilst Tables 5.2.1 - 5.2.6 (given on pages 166-168) contain the mean and variance of X_d and Y_d, under the MV-criterion.

Figures 5.2 and 5.3 show how $\mathrm{E}[X_d|\theta]$ and $\mathrm{E}[Y_d|\theta]$ change with θ, under the A-criterion, for two of these designs, namely those involving 24 and 36 subjects. Similarly, Figures 5.4 and 5.5 show how $\mathrm{E}[X_d|\theta]$ and $\mathrm{E}[Y_d|\theta]$ change with θ, under the MV-criterion, for these two designs.

Examining the trends of the mean performance measures given in Tables 5.1.1 - 5.1.6 and Tables 5.2.1 - 5.2.6 and illustrated in Figures 5.2 - 5.5 we observe that, as expected, for any particular value of θ, the mean performance measures increase as the number of subjects is increased. Also, we note there is a gradual reduction

Table 5.1: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for designs based on a pair of Williams squares of side three and ≤ 36 subjects.

Table 5.1.1: $t = 3$, $m = 6$, $n = 1$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	2.40	0.00	1.33	0.00
0.1	2.07	0.17	1.12	0.07
0.2	1.76	0.29	0.92	0.11
0.3	1.47	0.34	0.73	0.12
0.4	1.20	0.34	0.57	0.12
0.5	0.96	0.28	0.43	0.09
0.6	0.75	0.20	0.32	0.06
0.7	0.59	0.11	0.23	0.03
0.8	0.47	0.45	0.17	0.01
0.9	0.40	0.01	0.14	2.1×10^{-3}
1.0	0.38	0.00	0.13	0.00

Table 5.1.2: $t = 3$, $m = 6$, $n = 2$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	4.80	0.00	2.67	0.00
0.1	4.36	0.16	2.36	0.07
0.2	3.91	0.31	2.06	0.14
0.3	3.46	0.45	1.76	0.18
0.4	3.00	0.57	1.46	0.21
0.5	2.52	0.64	1.18	0.21
0.6	2.05	0.62	0.91	0.18
0.7	1.60	0.49	0.66	0.13
0.8	1.19	0.28	0.45	0.07
0.9	0.88	0.08	0.31	0.02
1.0	0.75	0.00	0.25	0.00

Table 5.1.3: $t = 3$, $m = 6$, $n = 3$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.20	0.00	4.00	0.00
0.1	6.61	0.18	3.58	0.09
0.2	6.01	0.35	3.17	0.16
0.3	5.40	0.51	2.76	0.22
0.4	4.77	0.66	2.35	0.26
0.5	4.11	0.78	1.94	0.28
0.6	3.44	0.86	1.54	0.27
0.7	2.74	0.82	1.15	0.22
0.8	2.05	0.59	0.80	0.14
0.9	1.43	0.21	0.51	0.04
1.0	1.13	0.00	0.38	0.00

Table 5.1.4: $t = 3$, $m = 6$, $n = 4$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	9.60	0.00	5.33	0.00
0.1	8.85	0.22	4.80	0.11
0.2	8.09	0.42	4.27	0.20
0.3	7.31	0.59	3.75	0.26
0.4	6.51	0.75	3.22	0.31
0.5	5.69	0.90	2.69	0.33
0.6	4.82	1.01	2.17	0.33
0.7	3.91	1.05	1.66	0.29
0.8	2.96	0.88	1.16	0.21
0.9	2.04	0.39	0.73	0.08
1.0	1.50	0.00	0.50	0.00

Table 5.1.5: $t = 3$, $m = 6$, $n = 5$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.00	0.00	6.67	0.00
0.1	11.09	0.26	6.02	0.13
0.2	10.17	0.48	5.38	0.23
0.3	9.22	0.68	4.73	0.31
0.4	8.25	0.86	4.09	0.36
0.5	7.24	1.02	3.44	0.39
0.6	6.19	1.15	2.80	0.39
0.7	5.07	1.22	2.16	0.35
0.8	3.89	1.10	1.54	0.26
0.9	2.67	0.58	0.96	0.11
1.0	1.88	0.00	0.63	0.00

Table 5.1.6: $t = 3$, $m = 6$, $n = 6$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	14.40	0.00	8.00	0.00
0.1	13.33	0.30	7.24	0.15
0.2	12.24	0.56	6.48	0.27
0.3	11.13	0.78	5.72	0.36
0.4	9.98	0.98	4.95	0.42
0.5	8.80	1.15	4.19	0.45
0.6	7.56	1.29	3.43	0.45
0.7	6.24	1.37	2.67	0.41
0.8	4.82	1.29	1.92	0.31
0.9	3.32	0.76	1.21	0.15
1.0	2.25	0.00	0.75	0.00

Table 5.2: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for designs based on a pair of Williams squares of side three and ≤ 36 subjects.

Table 5.2.1: $t = 3$, $m = 6$, $n = 1$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	2.40	0.00	1.33	0.00
0.1	1.93	0.28	1.03	0.11
0.2	1.56	0.35	0.80	0.13
0.3	1.26	0.33	0.62	0.12
0.4	1.01	0.28	0.47	0.10
0.5	0.81	0.22	0.35	0.07
0.6	0.64	0.14	0.26	0.04
0.7	0.52	0.08	0.19	0.02
0.8	0.43	0.03	0.15	0.01
0.9	0.39	4.3×10^{-3}	0.13	1.1×10^{-3}
1.0	0.38	0.00	0.13	0.00

Table 5.2.2: $t = 3$, $m = 6$, $n = 2$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	4.80	0.00	2.67	0.00
0.1	4.14	0.28	2.22	0.12
0.2	3.62	0.44	1.87	0.18
0.3	3.12	0.56	1.56	0.21
0.4	2.65	0.61	1.27	0.21
0.5	2.20	0.60	1.00	0.19
0.6	1.77	0.53	0.76	0.15
0.7	1.38	0.39	0.55	0.10
0.8	1.05	0.21	0.38	0.05
0.9	0.83	0.05	0.28	0.01
1.0	0.75	0.00	0.25	0.00

Table 5.2.3: $t = 3$, $m = 6$, $n = 3$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	7.20	0.00	4.00	0.00
0.1	6.34	0.31	3.41	0.14
0.2	5.65	0.51	2.94	0.22
0.3	4.98	0.66	2.50	0.26
0.4	4.33	0.78	2.09	0.29
0.5	3.68	0.85	1.70	0.28
0.6	3.03	0.84	1.32	0.25
0.7	2.40	0.73	0.98	0.19
0.8	1.80	0.48	0.67	0.11
0.9	1.32	0.14	0.45	0.03
1.0	1.13	0.00	0.38	0.00

Table 5.2.4: $t = 3$, $m = 6$, $n = 4$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	9.60	0.00	5.33	0.00
0.1	8.54	0.36	4.60	0.17
0.2	7.67	0.59	4.00	0.26
0.3	6.83	0.77	3.45	0.32
0.4	6.00	0.92	2.92	0.35
0.5	5.17	1.03	2.40	0.36
0.6	4.33	1.07	1.90	0.33
0.7	3.47	1.01	1.43	0.27
0.8	2.62	0.76	1.00	0.18
0.9	1.86	0.29	0.64	0.05
1.0	1.50	0.00	0.50	0.00

Table 5.2.5: $t = 3$, $m = 6$, $n = 5$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.00	0.00	6.67	0.00
0.1	10.75	0.41	5.79	0.19
0.2	9.70	0.68	5.07	0.30
0.3	8.69	0.89	4.40	0.38
0.4	7.68	1.06	3.74	0.42
0.5	6.66	1.19	3.11	0.43
0.6	5.63	1.27	2.49	0.40
0.7	4.56	1.24	1.89	0.34
0.8	3.47	1.02	1.33	0.24
0.9	2.43	0.46	0.85	0.09
1.0	1.88	0.00	0.63	0.00

Table 5.2.6: $t = 3$, $m = 6$, $n = 6$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	14.40	0.00	8.00	0.00
0.1	12.96	0.46	6.98	0.22
0.2	11.74	0.76	6.15	0.35
0.3	10.55	1.01	5.35	0.44
0.4	9.36	1.21	4.57	0.49
0.5	8.16	1.36	3.82	0.50
0.6	6.93	1.45	3.08	0.47
0.7	5.65	1.45	2.36	0.41
0.8	4.33	1.25	1.68	0.29
0.9	3.02	0.63	1.06	0.12
1.0	2.25	0.00	0.75	0.00

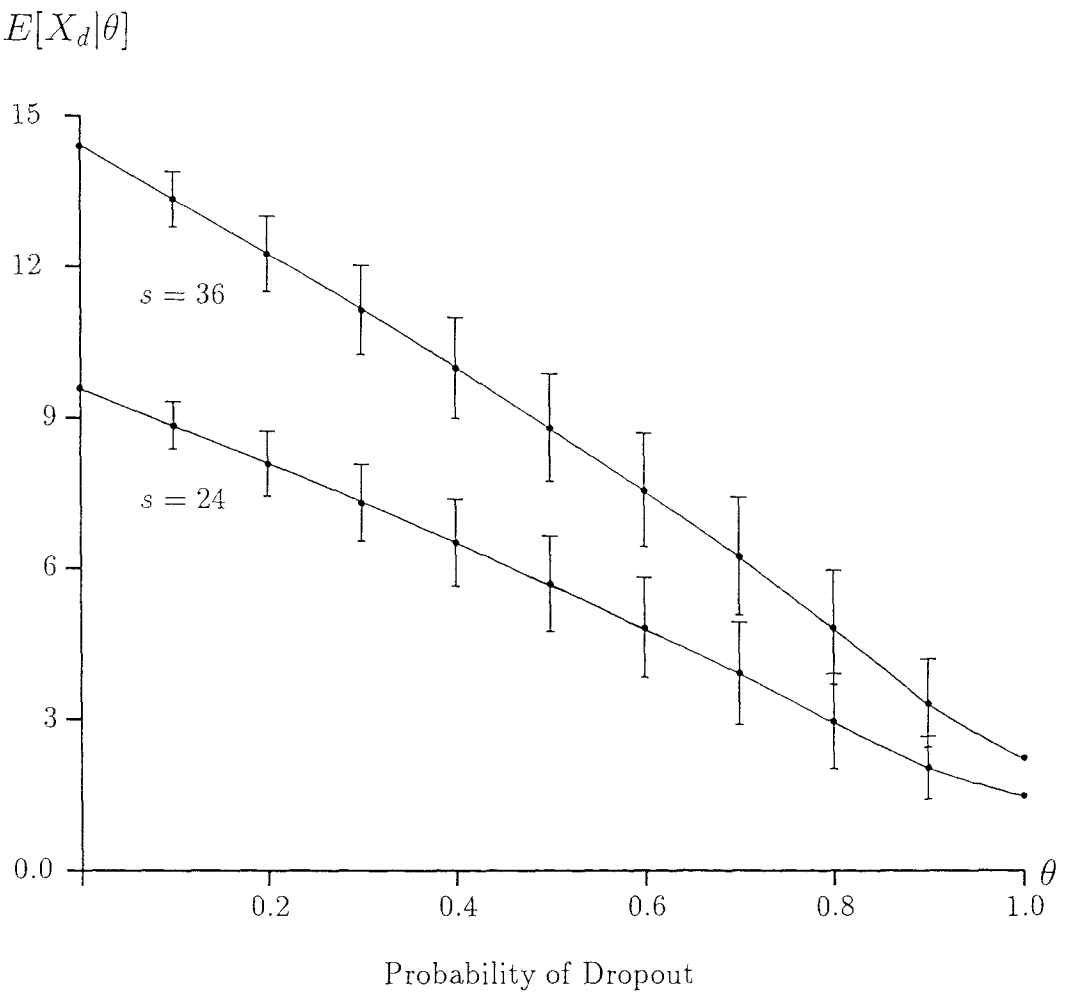


Figure 5.2: Performance for direct treatment comparisons under the A-criterion of designs based on Figure 5.1 for 24 and 36 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

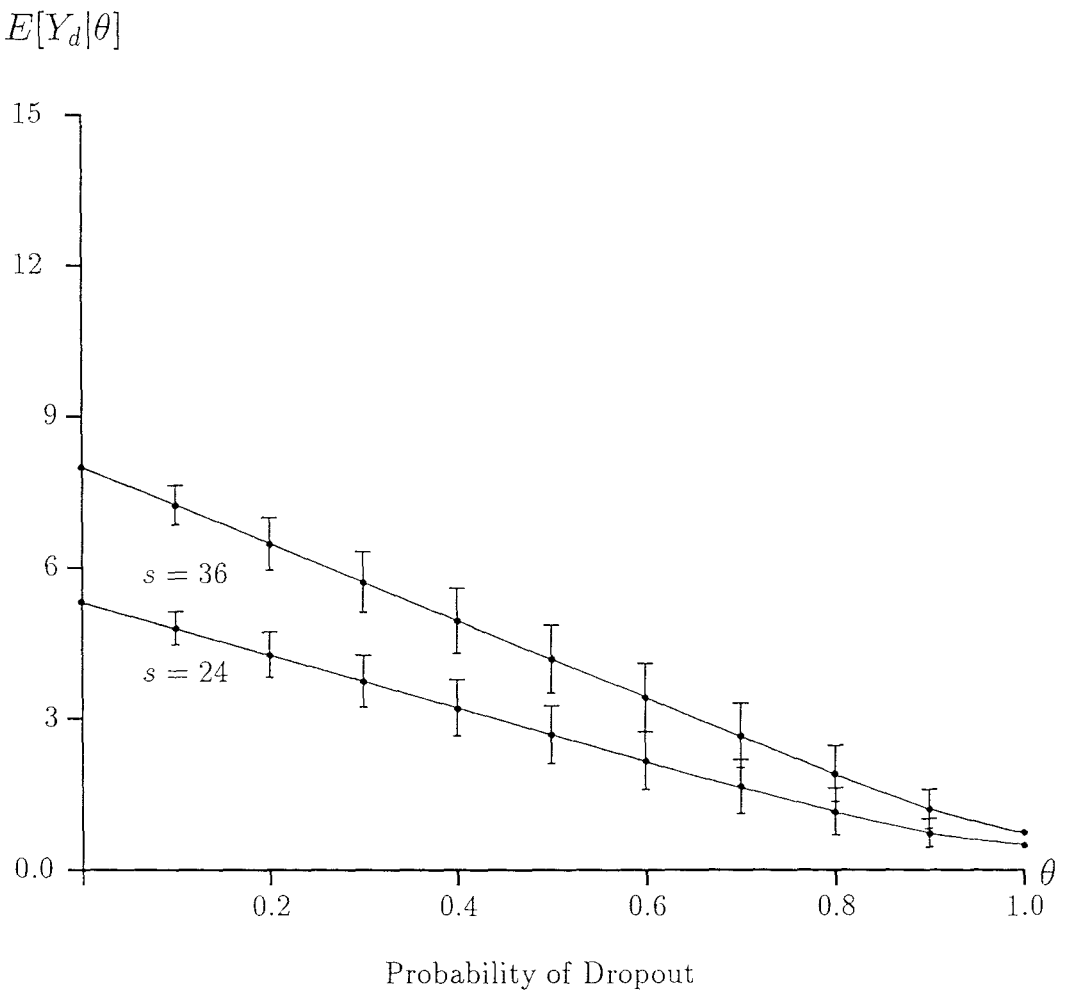


Figure 5.3: Performance for direct treatment comparisons under the A-criterion of designs based on Figure 5.1 for 24 and 36 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

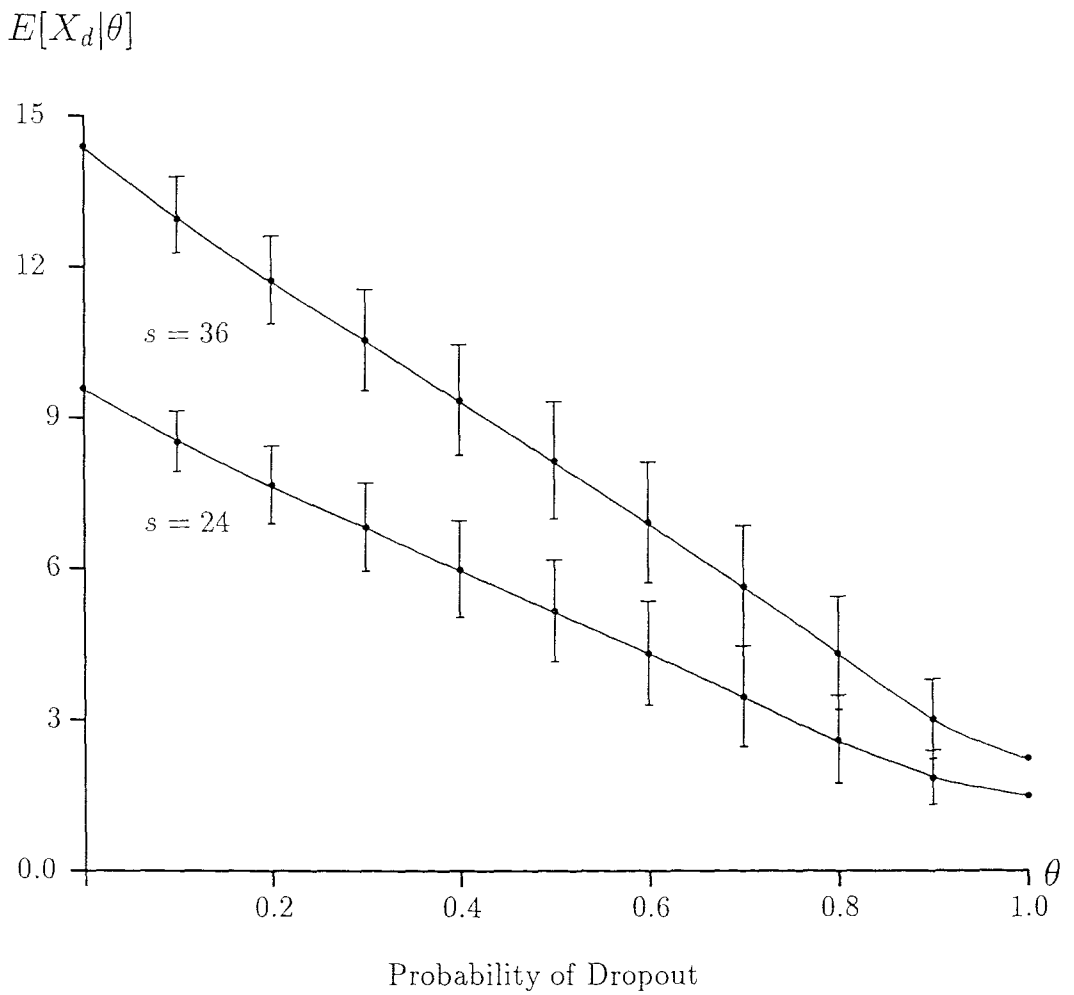


Figure 5.4: Performance for direct treatment comparisons under the MV-criterion of designs based on Figure 5.1 for 24 and 36 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

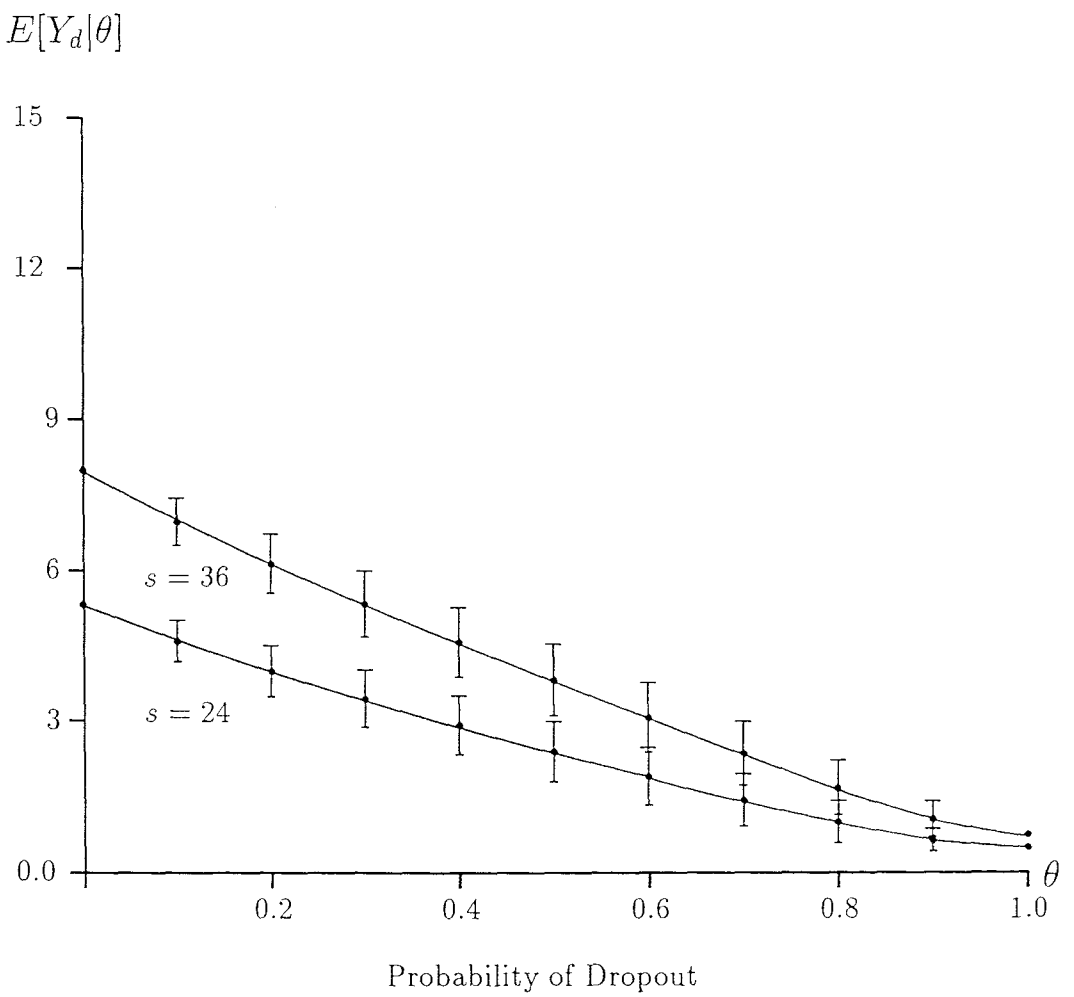


Figure 5.5: Performance for direct treatment comparisons under the MV-criterion of designs based on Figure 5.1 for 24 and 36 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

in the value of each mean performance measure as θ increases. Examining Tables 5.1.1 - 5.1.6 and Tables 5.2.1 - 5.2.6 we observe that the variance of the performance measures is reasonably small and the difference between the respective mean performance measures obtained under either criterion is not large, suggesting that the spread in the variances of the pairwise treatment comparisons for each of the designs can never be large. Equality or near equality amongst the variances of the pairwise treatment comparisons is a highly desirable feature of any design used in experiments which aim to compare all the different treatments with equal precision.

The results obtained in this section demonstrate that designs based on a pair of Williams squares of side three are reasonably robust to final period dropout, provided that the value of θ is small. However, in common with the designs of Chapter 4 for the four treatment case, the results obtained also show that the information available in the implemented experiment may be considerably less than that of the original planned design, even when the anticipated number of final period dropouts is small. It is important, therefore, that the problem of dropouts is considered very carefully during the planning stage.

Ideally, in order to decide whether or not to proceed with a particular design its performance measures should be compared with those obtained from other competing, designs. In the next section we examine the robustness to final period dropout of two alternative designs and, in Section 5.4, all three designs are compared. The comparisons lead to recommendations on the choice of a three treatment, three period design for a given number of subjects and a given value of θ.

5.3 Non-uniform, Unbalanced Designs

Every design considered so far has been a uniform balanced designs. As discussed previously, when $t = p = 3$ the only uniform balanced design in common use in clinical cross-over studies is the pair of Williams squares investigated in the previous section.

Ideally, we should like to compare the performance of this design, in the presence of final period dropout, with other designs which employ the same number of subjects. In order to do this it is necessary to establish which of the many non-uniform, three treatment, three period designs available are worth investigating. In this section a short review of the results concerning the optimality of uniform and non-uniform designs is presented together with some discussion concerning their use in cross-over experiments in which dropouts are anticipated.

One of the main arguments put forward for using a uniform design is that, under the simple carry-over model of equation (1.1), it ensures that the least squares estimators of the direct treatment effects are orthogonal to the estimators of both the periods and the subjects. This property leads to simplification in the interpretation of the results and, less importantly, in the analysis. However, if subjects drop out of a study during the final period, the realised design will no longer be uniform. Hence, irrespective of the choice of planned design, orthogonality of the parameter estimators is no longer possible. Therefore, when final period dropouts are anticipated, there is only a very small probability that the gain from having a uniform implemented design will be realised.

The relationship between the orthogonality of various design parameters and the universal optimality of designs was considered by Kunert (1983) who established the optimality of several unbalanced designs. The approach he used was to compare the information matrices of a design under two different linear models, referred to as the finer and simpler models. The finer model contains all the parameters of the simpler model in addition to some extra nuisance parameter(s). Kunert used an ordering property, due to Magda (1980), defined on the information matrices obtained for each model in order to establish an orthogonality condition. When the condition is satisfied, it ensures that the information matrix for estimating an effect is the same under each of the models. It follows that if a design can be shown to be universally optimal for the estimation of either the direct or the first-order carry-over treatment

effects under one of the models and the orthogonality condition is satisfied, then the design is also optimal for estimating the same effects under the other model. For example, a design which satisfies the orthogonality condition for the estimation of the direct treatment effects for two linear models, one with a first-order carry-over treatment effect term and the other without, will be universally optimal for the estimation of the direct treatment effects under both models if it can be shown to be universally optimal under just one of these models.

Kunert gives examples of unbalanced designs which fulfil the orthogonality condition and proves their universal optimality over the class of all possible cross-over designs. Many of the designs require a large number of treatment periods which is an undesirable feature for a design intended for studies in which dropout is anticipated. In many of the designs it is necessary to incorporate an additional initial period, or pre-period, in the study in order to satisfy the orthogonality condition. The purpose of this period is to ensure that the observations made during the next period include a carry-over effect and no use is normally made of the observation on the subject at the end of a pre-period. The tendency of subjects to drop out of a study when there are too many treatment periods is an argument against employing a design which requires a pre-period.

Another reason for the popularity of uniform balanced designs is that they are known to be universally optimal for the estimation of direct and first-order carry-over treatment comparisons within the class of uniform designs for which $t = p$. Cheng and Wu (1980) doubted whether a uniform balanced design would be universally optimal over the class of all cross-over designs in which $t = p$. They argued that removing the restriction of uniformity makes intractable the maximisation of the trace of the information matrices (see equations 1.11 and 1.15), a necessary step in establishing the universal optimality of a design. Nevertheless, they did establish some important results concerning the optimality of uniform balanced and strongly balanced designs.

Definition 5.1 A design is **strongly balanced** if each treatment is preceded equally often by every other treatment including itself.

In particular they proved that a strongly balanced design formed by adding an extra period to a uniform balanced design is universally optimal for the estimation of direct and first-order carry-over treatment effects within the class of all designs involving kt subjects, for k integer, in which $p = t + 1$.

Clearly, if it is believed that subjects may drop out during the final period of a three period study, it is not desirable to consider forming a strongly balanced design by extending the number of periods. If, however, a strongly balanced design, or a design with almost equal occurrence of all ordered pairs of treatment labels, can be found which uses only three periods then such a design is worth investigating further as an alternative to the currently favoured designs based on pairs of Williams squares of side three.

Kunert (1984) investigated the optimality of uniform balanced designs within the class of all possible designs. He found that, when the restriction of uniformity is removed, then the optimal designs for estimating the direct and first-order carry-over effects are not usually the same. In addition, even when $t = p$, uniform balanced designs will not generally be optimal when designs involving adjacent pairs of identical treatments are allowed. In particular Kunert established the following:

1. When $t = s = p \neq 2$ and a balanced Latin square exists, it is universally optimal for the estimation of the direct treatment effects over the class of all designs in which $t = s = p$.

2. When $t = p \geq 6$ and $s = 2t$, a uniform balanced design is universally optimal for the estimation of the direct treatment effects over the class of all designs in which $t = p$ and $s = 2t$.

3. When $t = p$ and $s = t(t-1)$ an **orthogonal residual effects design**, defined below, is universally optimal for the estimation of the first-order carry-over

treatment effects over the class of all designs in which $t = p$ and $s = t(t-1)$.

Definition 5.2 Suppose there is a uniform balanced design, d, involving p periods, t treatments and $m = t(t-1)$ distinct treatment sequences such that each ordered pair of treatments appears once and only once between the last and second to last periods of the design. The design formed by taking the first $(p-1)$ periods of d and forming a final period by repeating the $(p-1)$th period of design d is called an **orthogonal residual effects design.**

An example of an orthogonal residual effects design for $t = 3$ and $s = 6$ is given in Figure 5.6. Note that the property of universal optimality for estimating the first-order carry-over effects only holds for this design and does not extend to designs involving $6n$ subjects, for integer $n > 1$, formed by taking n copies of the six treatment sequences.

In addition to the results above, Kunert (1984) also established that, when the number of subjects is large, uniform balanced designs are not universally optimal for the estimation of the direct treatment effects since designs giving universally better estimates of the treatment comparisons can be found. Of particular interest is the following result.

Proposition 5.1 [Kunert (1984) Proposition 2.4] Let $t = p > 2$ and $s = \lambda t$, where λ is an integer such that

$$\lambda > t(t-1)^2/2.$$

Assume there is a design g involving t treatments, p periods and s subjects with the following properties.

(i) The first $t(t-1)$ subjects of g form an orthogonal residual effects design f, and

(ii) the remaining $s - t(t-1)$ subjects of g form a uniform balanced design.

Then g is universally better for the estimation of the direct treatment effects than any uniform balanced design for t treatments t periods and s subjects.

Note that, when $t = 3$, the minimum number of distinct treatment sequences required to obtain a design which satisfies Proposition 5.1 is 18. An examination of the robustness to final period dropout of an eighteen sequence design requires a prohibitive amount of computation. The concept of formulating designs by combining a uniform balanced design with an orthogonal residual effects design, however, is worth further investigation and is the motivation behind the design investigated in Section 5.3.2.

Kunert (1984) has demonstrated that, when dropouts are not considered, certain non-uniform and unbalanced designs give universally better estimates of the treatment comparisons than uniform balanced designs. His work raises the question of whether or not the designs he presents continue to out-perform uniform balanced designs in the presence of final period dropouts and provides further justification for not restricting the choice of design to the class of uniform designs.

In the remainder of this section, the robustness to final period dropout of two particular non-uniform designs is examined. In Section 5.3.1 the performance of an orthogonal residual effects design for $t = p = 3$ is investigated. In Section 5.3.2 the performance of a design which is a compromise between the pair of Williams squares and the orthogonal residual effects design is investigated.

5.3.1 Examination of Orthogonal Residual Effects Designs

Let $d(3, 6, n, 3)$ be the planned design with treatment labels 0, 1, and 2 formed by taking n copies of the design given in Figure 5.6.

In order to investigate the robustness to final period dropout of designs built from the orthogonal residual effects design of Figure 5.6, for studies involving up to a maximum of 36 subjects, there are six different sizes of design to consider. In each case the appropriate design is formed by allocating n subjects to each treatment sequence, where n takes each of the values $1, \ldots, 6$.

One important feature of each of the planned designs, in common with designs

Figure 5.6: Orthogonal residual effects design for $t = 3$ and treatment labels 0, 1, and 2.

0	1	1
1	2	2
2	0	0
0	2	2
1	0	0
2	1	1

based on a pair of Williams squares of side three, is that, in each case, the set of implementable designs arising from the planned design does not contain any disconnected designs.

For each design, the mean and variance of the performance measures X_d and Y_d, under the A-criterion, are given in Tables 5.3.1 - 5.3.6 (given on pages 180-182) whilst Tables 5.4.1 - 5.4.6 (given on pages 183-185) contain the mean and variance of the performance measures X_d and Y_d under the MV-criterion.

Figures 5.7 and 5.8 show how $\mathrm{E}[X_d|\theta]$ and $\mathrm{E}[Y_d|\theta]$ change with θ, under the A-criterion, for two of these designs, namely those involving 24 and 36 subjects. Similarly, Figures 5.9 and 5.10 show how $\mathrm{E}[X_d|\theta]$ and $\mathrm{E}[Y_d|\theta]$ change with θ, for each design under the MV-criterion.

The observations made concerning the general trend of the mean of X_d and Y_d as θ increases for the designs based on a pair of Williams squares of side three under the A- and MV-criteria also apply to these designs. Note, however, that although the amount of information in the direct treatment effects is, as expected, consistently greater than that in the first-order carry-over effects for each design, the magnitude of this difference is not as great as for the corresponding values obtained for the designs based on a pair of Williams squares.

Table 5.3: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for designs based on the orthogonal residual effects design of Figure 5.6 and ≤ 36 subjects.

Table 5.3.1: $t = 3$, $m = 6$, $n = 1$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	2.00	0.00	1.67	0.00
0.1	1.89	0.03	1.47	0.06
0.2	1.75	0.09	1.27	0.13
0.3	1.57	0.17	1.05	0.19
0.4	1.36	0.24	0.84	0.21
0.5	1.12	0.27	0.63	0.19
0.6	0.89	0.24	0.45	0.15
0.7	0.68	0.17	0.30	0.08
0.8	0.51	0.08	0.20	0.03
0.9	0.41	0.02	0.14	0.01
1.0	0.38	0.00	0.13	0.00

Table 5.3.2: $t = 3$, $m = 6$, $n = 2$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	4.00	0.00	3.33	0.00
0.1	3.86	0.02	3.04	0.08
0.2	3.69	0.06	2.73	0.15
0.3	3.47	0.13	2.40	0.23
0.4	3.20	0.25	2.04	0.31
0.5	2.85	0.42	1.67	0.36
0.6	2.42	0.59	1.29	0.37
0.7	1.91	0.63	0.91	0.30
0.8	1.38	0.45	0.58	0.16
0.9	0.93	0.14	0.34	0.04
1.0	0.75	0.00	0.25	0.00

Table 5.3.3: $t = 3$, $m = 6$, $n = 3$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	6.00	0.00	5.00	0.00
0.1	5.81	0.02	4.59	0.09
0.2	5.58	0.06	4.16	0.18
0.3	5.31	0.12	3.70	0.28
0.4	4.97	0.23	3.22	0.37
0.5	4.53	0.42	2.70	0.45
0.6	3.98	0.68	2.16	0.49
0.7	3.27	0.94	1.60	0.47
0.8	2.42	0.91	1.04	0.32
0.9	1.57	0.39	0.58	0.10
1.0	1.13	0.00	0.38	0.00

Table 5.3.4: $t = 3$, $m = 6$, $n = 4$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	8.00	0.00	6.67	0.00
0.1	7.76	0.03	6.14	0.11
0.2	7.47	0.07	5.58	0.22
0.3	7.13	0.13	5.00	0.33
0.4	6.71	0.24	4.38	0.44
0.5	6.18	0.43	3.72	0.53
0.6	5.52	0.72	3.02	0.59
0.7	4.65	1.09	2.29	0.59
0.8	3.54	1.28	1.54	0.46
0.9	2.27	0.72	0.85	0.18
1.0	1.50	0.00	0.50	0.00

Table 5.3.5: $t = 3$, $m = 6$, $n = 5$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.00	0.00	8.33	0.00
0.1	9.70	0.03	7.69	0.13
0.2	9.36	0.08	7.01	0.27
0.3	8.94	0.15	6.29	0.40
0.4	8.44	0.27	5.53	0.51
0.5	7.82	0.46	4.73	0.62
0.6	7.03	0.76	3.88	0.69
0.7	6.01	1.19	2.99	0.70
0.8	4.67	1.56	2.05	0.58
0.9	3.03	1.08	1.15	0.27
1.0	1.88	0.00	0.63	0.00

Table 5.3.6: $t = 3$, $m = 6$, $n = 6$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.00	0.00	10.00	0.00
0.1	11.65	0.04	9.24	0.16
0.2	11.24	0.09	8.43	0.31
0.3	10.76	0.17	7.58	0.46
0.4	10.17	0.30	6.69	0.59
0.5	9.45	0.50	5.74	0.71
0.6	8.54	0.81	4.74	0.78
0.7	7.36	1.28	3.68	0.79
0.8	5.80	1.77	2.57	0.69
0.9	3.81	1.42	1.45	0.35
1.0	2.25	0.00	0.75	0.00

Table 5.4: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for designs based on the orthogonal residual effects design of Figure 5.6 and ≤ 36 subjects

Table 5.4.1: $t = 3$, $m = 6$, $n = 1$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	2.00	0.00	1.67	0.00
0.1	1.83	0.05	1.39	0.11
0.2	1.65	0.13	1.14	0.17
0.3	1.44	0.21	0.91	0.19
0.4	1.21	0.27	0.71	0.19
0.5	0.98	0.28	0.52	0.16
0.6	0.76	0.22	0.36	0.12
0.7	0.58	0.13	0.25	0.06
0.8	0.45	0.05	0.17	0.02
0.9	0.39	0.01	0.13	3.1×10^{-3}
1.0	0.38	0.00	0.13	0.00

Table 5.4.2: $t = 3$, $m = 6$, $n = 2$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	4.00	0.00	3.33	0.00
0.1	3.78	0.04	2.90	0.13
0.2	3.56	0.10	2.53	0.20
0.3	3.29	0.19	2.16	0.28
0.4	2.98	0.34	1.80	0.34
0.5	2.59	0.52	1.44	0.36
0.6	2.13	0.65	1.08	0.33
0.7	1.64	0.62	0.74	0.25
0.8	1.17	0.37	0.47	0.12
0.9	0.84	0.08	0.29	0.02
1.0	0.75	0.00	0.25	0.00

Table 5.4.3: $t = 3$, $m = 6$, $n = 3$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	6.00	0.00	5.00	0.00
0.1	5.71	0.04	4.41	0.15
0.2	5.43	0.10	3.90	0.26
0.3	5.10	0.19	3.41	0.35
0.4	4.70	0.34	2.90	0.43
0.5	4.22	0.56	2.39	0.49
0.6	3.61	0.84	1.86	0.49
0.7	2.88	1.03	1.34	0.43
0.8	2.07	0.85	0.85	0.26
0.9	1.37	0.26	0.49	0.07
1.0	1.13	0.00	0.38	0.00

Table 5.4.4: $t = 3$, $m = 6$, $n = 4$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	8.00	0.00	6.67	0.00
0.1	7.65	0.05	5.93	0.18
0.2	7.30	0.11	5.29	0.31
0.3	6.89	0.21	4.65	0.43
0.4	6.41	0.36	4.00	0.53
0.5	5.82	0.60	3.34	0.60
0.6	5.09	0.93	2.66	0.62
0.7	4.17	1.28	1.96	0.57
0.8	3.07	1.31	1.28	0.41
0.9	1.97	0.54	0.71	0.13
1.0	1.50	0.00	0.50	0.00

Table 5.4.5: $t = 3$, $m = 6$, $n = 5$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.00	0.00	8.33	0.00
0.1	9.58	0.05	7.45	0.21
0.2	9.17	0.13	6.68	0.37
0.3	8.69	0.23	5.90	0.51
0.4	8.12	0.40	5.11	0.63
0.5	7.42	0.64	4.30	0.71
0.6	6.56	1.00	3.46	0.74
0.7	5.47	1.45	2.60	0.70
0.8	4.11	1.68	1.73	0.54
0.9	2.62	0.88	0.96	0.20
1.0	1.88	0.00	0.63	0.00

Table 5.4.6: $t = 3$, $m = 6$, $n = 6$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.00	0.00	10.00	0.00
0.1	11.52	0.06	8.97	0.24
0.2	11.04	0.14	8.07	0.42
0.3	10.48	0.26	7.15	0.59
0.4	9.82	0.43	6.22	0.73
0.5	9.02	0.70	5.25	0.83
0.6	8.02	1.09	4.26	0.86
0.7	6.77	1.60	3.23	0.82
0.8	5.17	1.98	2.19	0.65
0.9	3.30	1.24	1.22	0.28
1.0	2.25	0.00	0.75	0.00

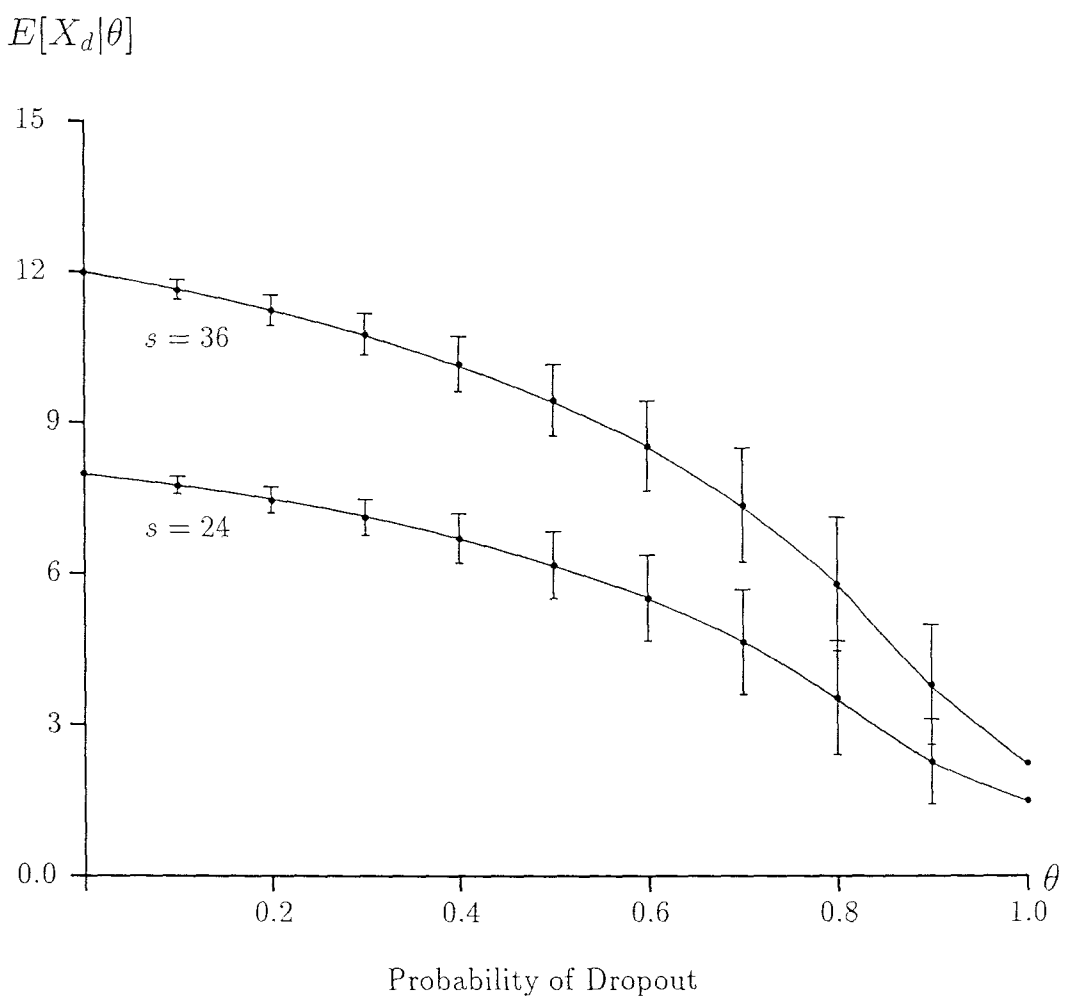


Figure 5.7: Performance for direct treatment comparisons under the A-criterion of designs based on Figure 5.6 for 24 and 36 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

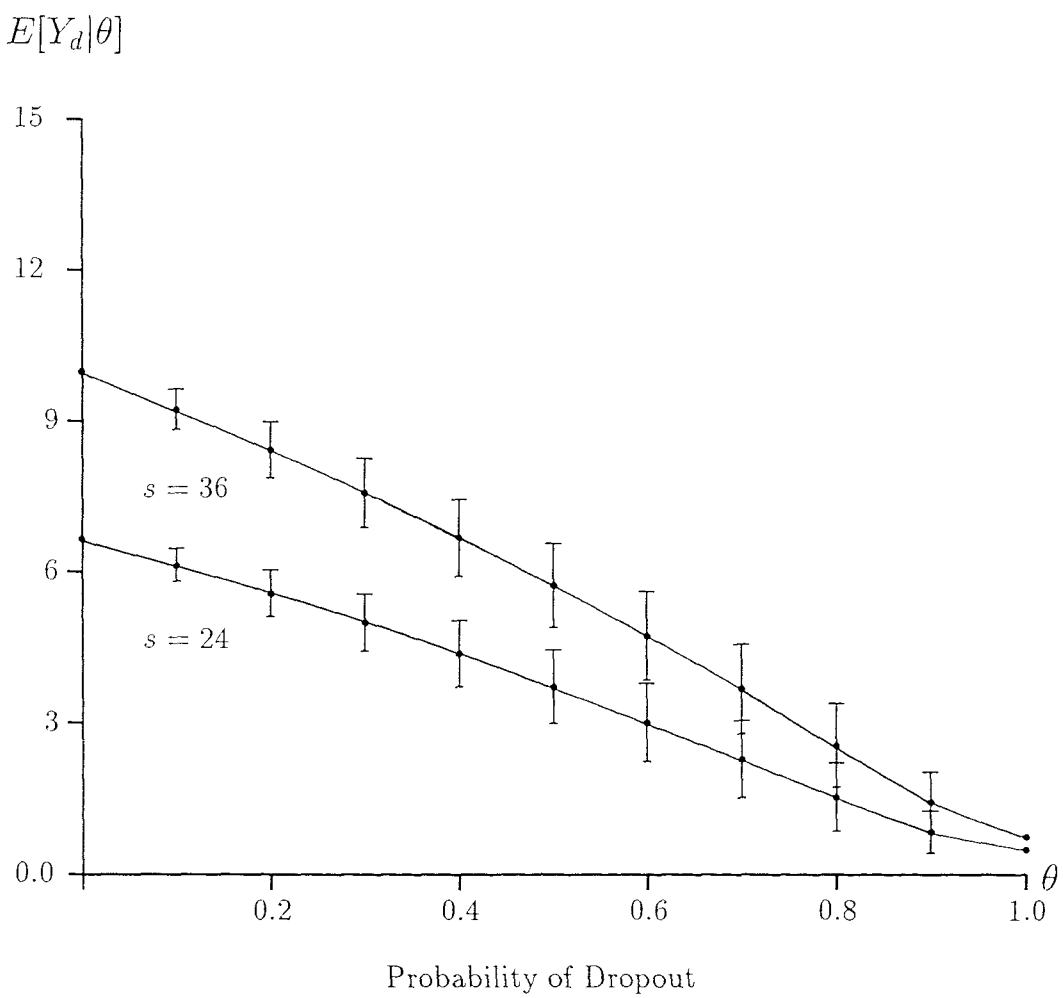


Figure 5.8: Performance for direct treatment comparisons under the A-criterion of designs based on Figure 5.6 for 24 and 36 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

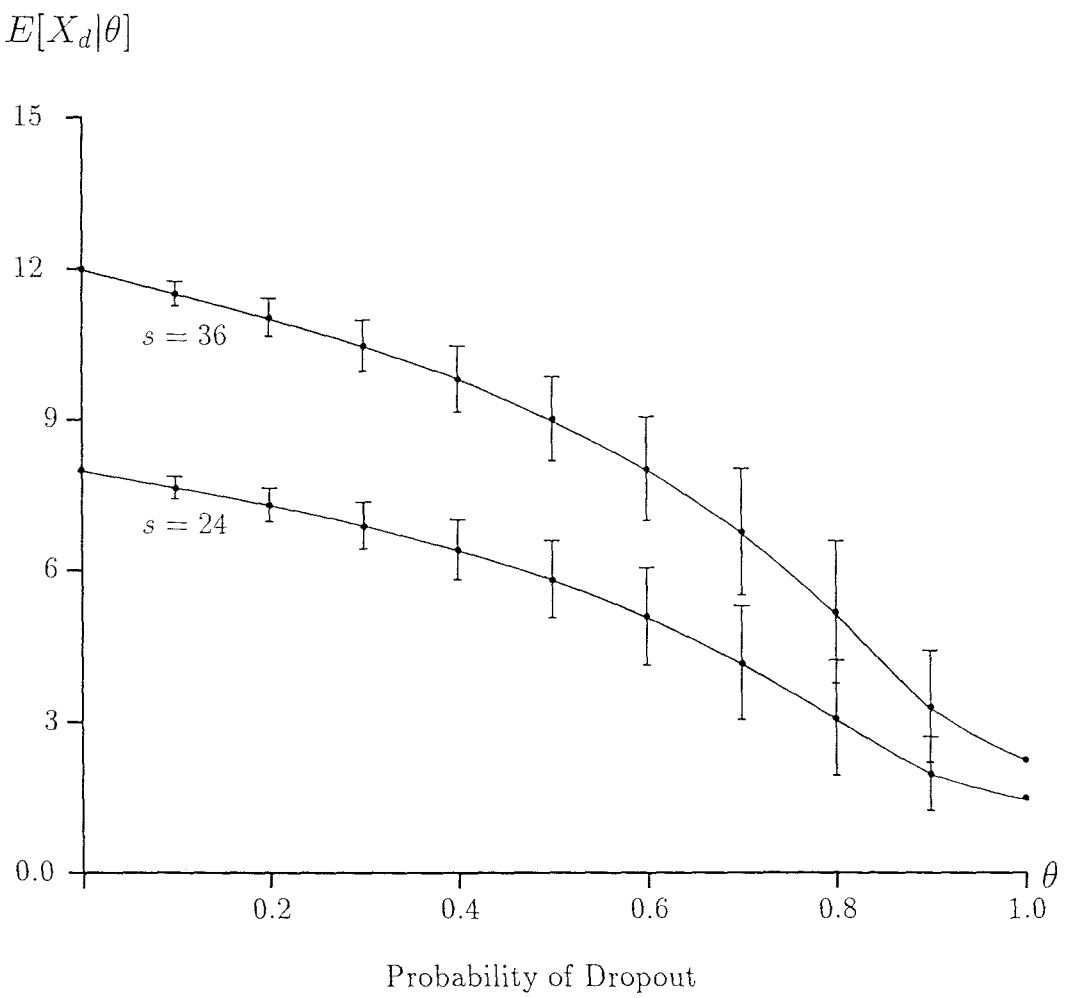


Figure 5.9: Performance for direct treatment comparisons under the MV-criterion of designs based on Figure 5.6 for 24 and 36 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

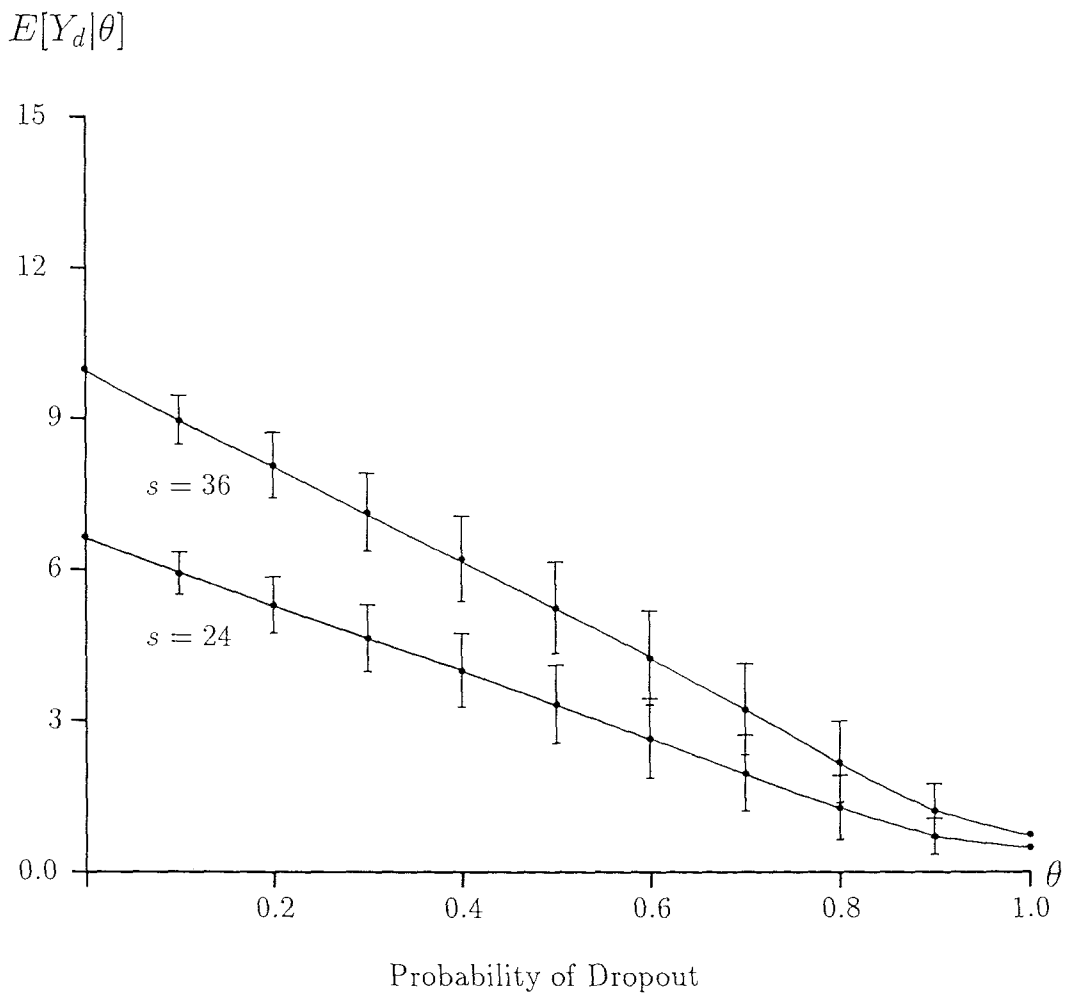


Figure 5.10: Performance for direct treatment comparisons under the MV-criterion of designs based on Figure 5.6 for 24 and 36 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

5.3.2 Examination of Compromise Designs

Let $d(3, 6, n, 3)$ be the planned design with treatment labels 0, 1, and 2 formed by taking n copies of the design given in Figure 5.11. The design given in Figure 5.11 is a design which is a compromise between the pair of Williams squares of side three (Figure 5.1) and the orthogonal residual effects design (Figure 5.6), since it consists of the first three sequences of the Williams design and the final three sequences of the orthogonal residual effects design.

Figure 5.11: Three treatment, three period design with treatment labels 0, 1 and 2.

0	1	2
1	2	0
2	0	1
0	2	2
1	0	0
2	1	1

As for the designs investigated in Section 5.3.1, there are six distinct treatment sequences and hence, in order to investigate the robustness to final period dropout of designs built form the design of Figure 5.11, involving up to 36 subjects, there are six different sizes of design to consider.

In common with designs built from pairs of Williams squares of side three or replicates of the orthogonal residual effects designs given in Figure 5.6, designs built from replicates of the design given in Figure 5.11 do not give rise to any disconnected implementable designs. For each design, Tables 5.5.1 - 5.5.6 (given on pages 191-193) contain the mean and variance of the performance measures X_d and Y_d under the A-criterion, whilst Tables 5.6.1 - 5.6.6 (given on pages 194-196) contain the mean and variance of the performance measures X_d and Y_d under the MV-criterion.

Figures 5.12 and 5.13 show how $\mathrm{E}[X_d|\theta]$ and $\mathrm{E}[Y_d|\theta]$ change with θ, for each design under the A-criterion, for two of these designs, namely those involving 24 and 36 subjects. Similarly, Figures 5.14 and 5.15 show how $\mathrm{E}[X_d|\theta]$ and $\mathrm{E}[Y_d|\theta]$

Table 5.5: Mean and variance of the performance measures X_d and Y_d, under the A-criterion, for designs based on the pair of squares of Figure 5.11 and ≤ 36 subjects.

Table 5.5.1: $t = 3$, $m = 6$, $n = 1$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	2.15	0.00	1.43	0.00
0.1	1.95	0.08	1.24	0.07
0.2	1.74	0.16	1.05	0.12
0.3	1.51	0.24	0.86	0.14
0.4	1.28	0.28	0.68	0.14
0.5	1.04	0.28	0.52	0.12
0.6	0.83	0.22	0.38	0.09
0.7	0.64	0.14	0.26	0.05
0.8	0.49	0.06	0.18	0.02
0.9	0.40	0.01	0.14	3.2×10^{-3}
1.0	0.38	0.00	0.13	0.00

Table 5.5.2: $t = 3$, $m = 6$, $n = 2$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	4.30	0.00	2.87	0.00
0.1	4.03	0.07	2.59	0.07
0.2	3.74	0.15	2.30	0.14
0.3	3.42	0.26	2.00	0.20
0.4	3.07	0.39	1.70	0.24
0.5	2.67	0.51	1.38	0.26
0.6	2.23	0.59	1.07	0.25
0.7	1.75	0.56	0.77	0.19
0.8	1.29	0.36	0.51	0.10
0.9	0.91	0.11	0.32	0.02
1.0	0.75	0.00	0.25	0.00

Table 5.5.3: $t = 3$, $m = 6$, $n = 3$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	6.45	0.00	4.30	0.00
0.1	6.08	0.08	3.91	0.09
0.2	5.69	0.17	3.52	0.17
0.3	5.27	0.28	3.11	0.24
0.4	4.81	0.42	2.69	0.30
0.5	4.28	0.59	2.25	0.34
0.6	3.69	0.77	1.81	0.35
0.7	3.00	0.87	1.35	0.31
0.8	2.24	0.74	0.91	0.21
0.9	1.50	0.29	0.54	0.06
1.0	1.13	0.00	0.38	0.00

Table 5.5.4: $t = 3$, $m = 6$, $n = 4$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	8.60	0.00	5.73	0.00
0.1	8.13	0.10	5.24	0.11
0.2	7.64	0.20	4.73	0.21
0.3	7.11	0.32	4.21	0.30
0.4	6.52	0.47	3.67	0.36
0.5	5.87	0.65	3.11	0.41
0.6	5.13	0.87	2.53	0.43
0.7	4.26	1.07	1.94	0.41
0.8	3.24	1.06	1.34	0.30
0.9	2.16	0.54	0.79	0.12
1.0	1.50	0.00	0.50	0.00

Table 5.5.5: $t = 3$, $m = 6$, $n = 5$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.75	0.00	7.17	0.00
0.1	10.18	0.11	6.56	0.13
0.2	9.58	0.24	5.94	0.25
0.3	8.94	0.37	5.30	0.35
0.4	8.23	0.53	4.65	0.43
0.5	7.45	0.73	3.97	0.48
0.6	6.56	0.96	3.26	0.51
0.7	5.51	1.22	2.53	0.49
0.8	4.27	1.32	1.77	0.39
0.9	2.85	0.80	1.05	0.17
1.0	1.88	0.00	0.63	0.00

Table 5.5.6: $t = 3$, $m = 6$, $n = 6$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.90	0.00	8.60	0.00
0.1	12.23	0.13	7.89	0.16
0.2	11.53	0.27	7.15	0.29
0.3	10.77	0.42	6.40	0.41
0.4	9.94	0.60	5.62	0.50
0.5	9.02	0.81	4.82	0.56
0.6	7.98	1.06	3.98	0.58
0.7	6.76	1.35	3.11	0.56
0.8	5.29	1.53	2.21	0.46
0.9	3.57	1.06	1.32	0.23
1.0	2.25	0.00	0.75	0.00

Table 5.6: Mean and variance of the performance measures X_d and Y_d, under the MV-criterion, for designs based on the pair of squares of Figure 5.11 and $\leq$ 36 subjects.

Table 5.6.1: $t = 3$, $m = 6$, $n = 1$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	2.15	0.00	1.43	0.00
0.1	1.85	0.14	1.16	0.11
0.2	1.59	0.23	0.93	0.15
0.3	1.34	0.29	0.73	0.16
0.4	1.10	0.31	0.56	0.14
0.5	0.88	0.28	0.41	0.11
0.6	0.69	0.20	0.30	0.07
0.7	0.54	0.12	0.21	0.04
0.8	0.44	0.04	0.15	0.01
0.9	0.39	0.01	0.13	1.6×10^{-3}
1.0	0.38	0.00	0.13	0.00

Table 5.6.2: $t = 3$, $m = 6$, $n = 2$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	4.30	0.00	2.87	0.00
0.1	3.89	0.14	2.45	0.13
0.2	3.52	0.25	2.10	0.19
0.3	3.16	0.37	1.78	0.24
0.4	2.78	0.51	1.47	0.27
0.5	2.36	0.62	1.17	0.27
0.6	1.93	0.66	0.88	0.23
0.7	1.48	0.56	0.62	0.16
0.8	1.08	0.31	0.41	0.07
0.9	0.82	0.07	0.28	0.01
1.0	0.75	0.00	0.25	0.00

Table 5.6.3: $t = 3$, $m = 6$, $n = 3$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	6.45	0.00	4.30	0.00
0.1	5.91	0.15	3.74	0.16
0.2	5.43	0.28	3.27	0.24
0.3	4.95	0.43	2.83	0.31
0.4	4.44	0.60	2.40	0.35
0.5	3.89	0.78	1.97	0.37
0.6	3.27	0.94	1.54	0.36
0.7	2.60	0.97	1.12	0.29
0.8	1.89	0.71	0.73	0.17
0.9	1.32	0.21	0.46	0.04
1.0	1.13	0.00	0.38	0.00

Table 5.6.4: $t = 3$, $m = 6$, $n = 4$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	8.60	0.00	5.73	0.00
0.1	7.93	0.17	5.03	0.18
0.2	7.35	0.32	4.45	0.29
0.3	6.74	0.49	3.89	0.38
0.4	6.11	0.69	3.33	0.44
0.5	5.41	0.91	2.77	0.47
0.6	4.64	1.13	2.21	0.46
0.7	3.76	1.28	1.64	0.40
0.8	2.78	1.12	1.10	0.27
0.9	1.87	0.42	0.66	0.08
1.0	1.50	0.00	0.50	0.00

Table 5.6.5: $t = 3$, $m = 6$, $n = 5$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	10.75	0.00	7.17	0.00
0.1	9.96	0.19	6.33	0.21
0.2	9.26	0.37	5.63	0.35
0.3	8.54	0.56	4.94	0.45
0.4	7.77	0.78	4.26	0.52
0.5	6.93	1.03	3.57	0.56
0.6	6.00	1.30	2.88	0.55
0.7	4.93	1.52	2.18	0.50
0.8	3.71	1.47	1.49	0.36
0.9	2.46	0.68	0.88	0.13
1.0	1.88	0.00	0.63	0.00

Table 5.6.6: $t = 3$, $m = 6$, $n = 6$, $p = 3$.

θ	$E[X_d\|\theta]$	$\mathrm{Var}[X_d\|\theta]$	$E[Y_d\|\theta]$	$\mathrm{Var}[Y_d\|\theta]$
0.0	12.90	0.00	8.60	0.00
0.1	11.99	0.22	7.63	0.24
0.2	11.17	0.41	6.80	0.40
0.3	10.33	0.63	6.00	0.52
0.4	9.43	0.88	5.19	0.61
0.5	8.54	1.15	4.38	0.65
0.6	7.36	1.45	3.55	0.65
0.7	6.11	1.73	2.72	0.59
0.8	4.66	1.78	1.88	0.45
0.9	3.08	0.97	1.11	0.18
1.0	2.25	0.00	0.75	0.00

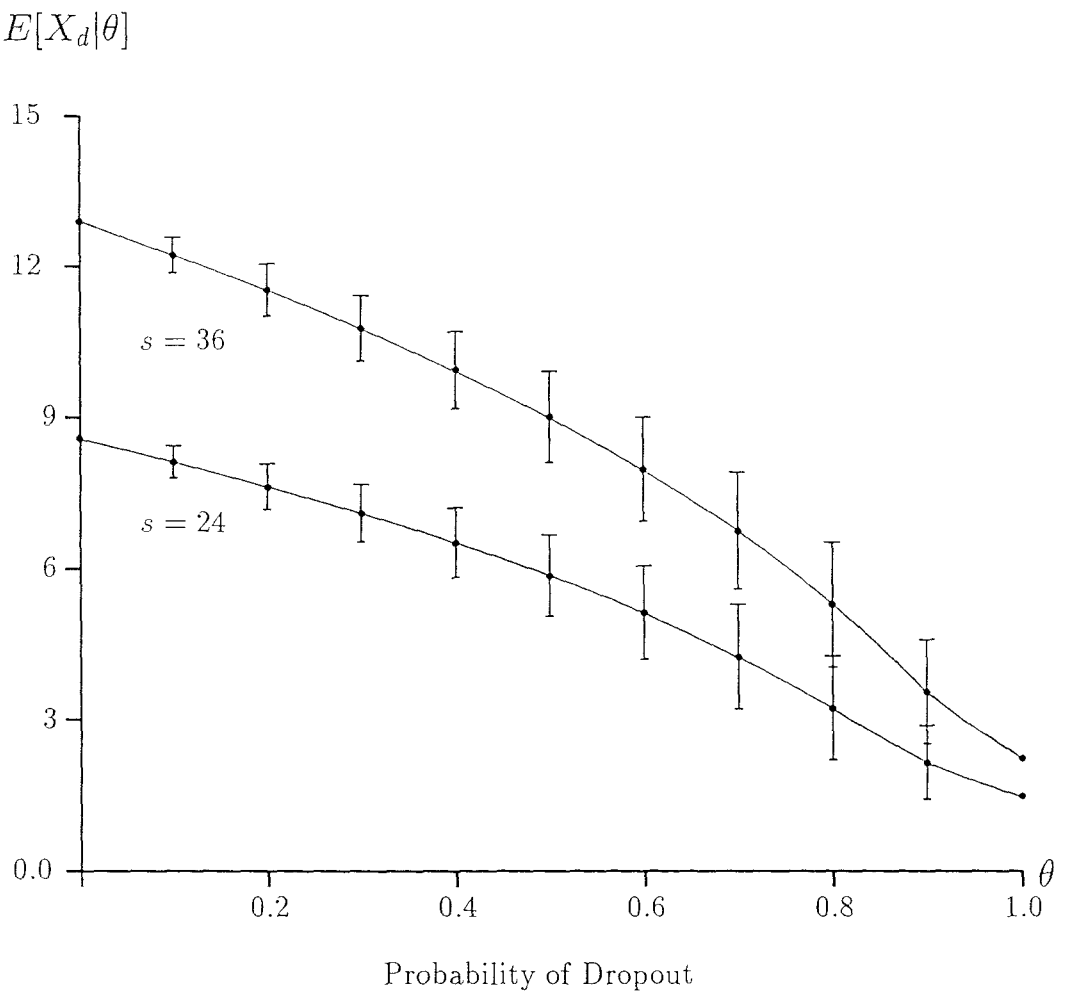


Figure 5.12: Performance for direct treatment comparisons under the A-criterion of designs based on Figure 5.11 for 24 and 36 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

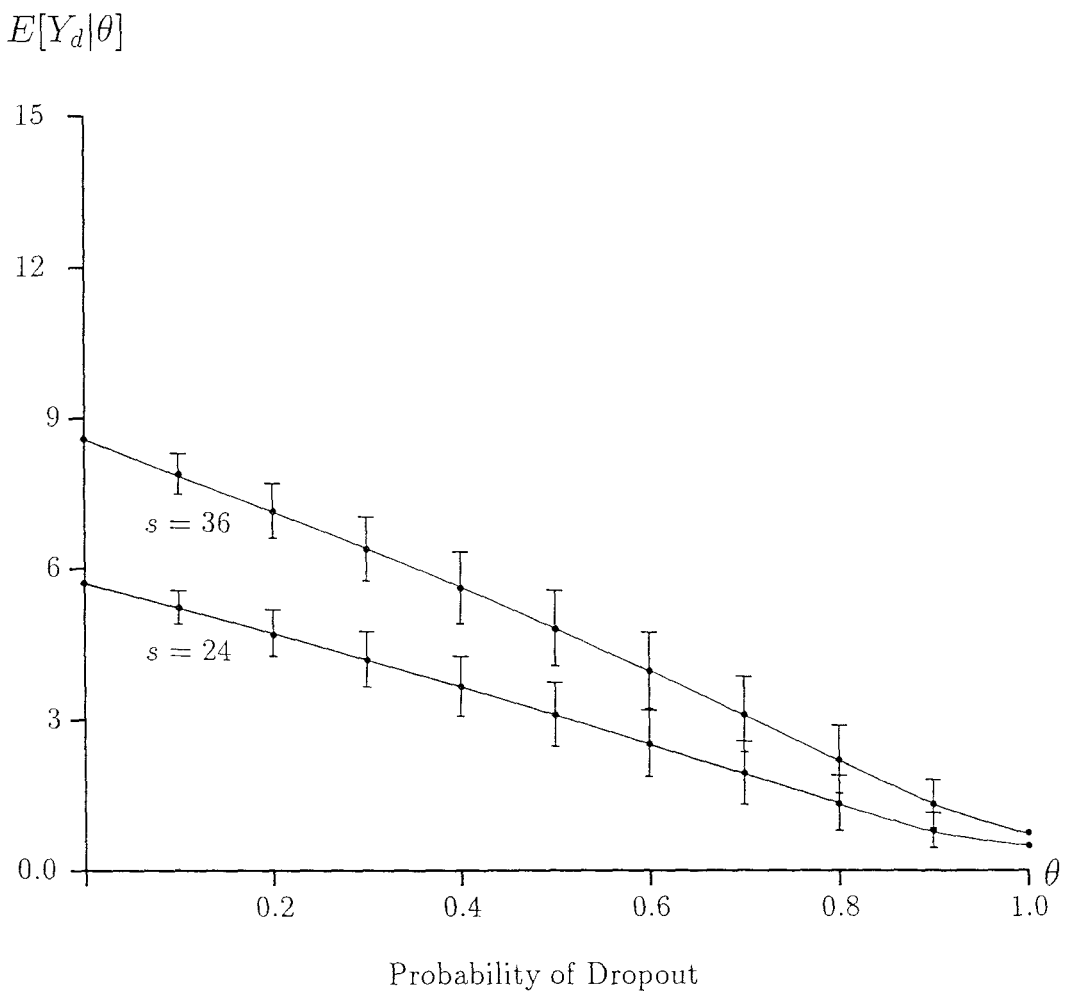


Figure 5.13: Performance for direct treatment comparisons under the A-criterion of designs based on Figure 5.11 for 24 and 36 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

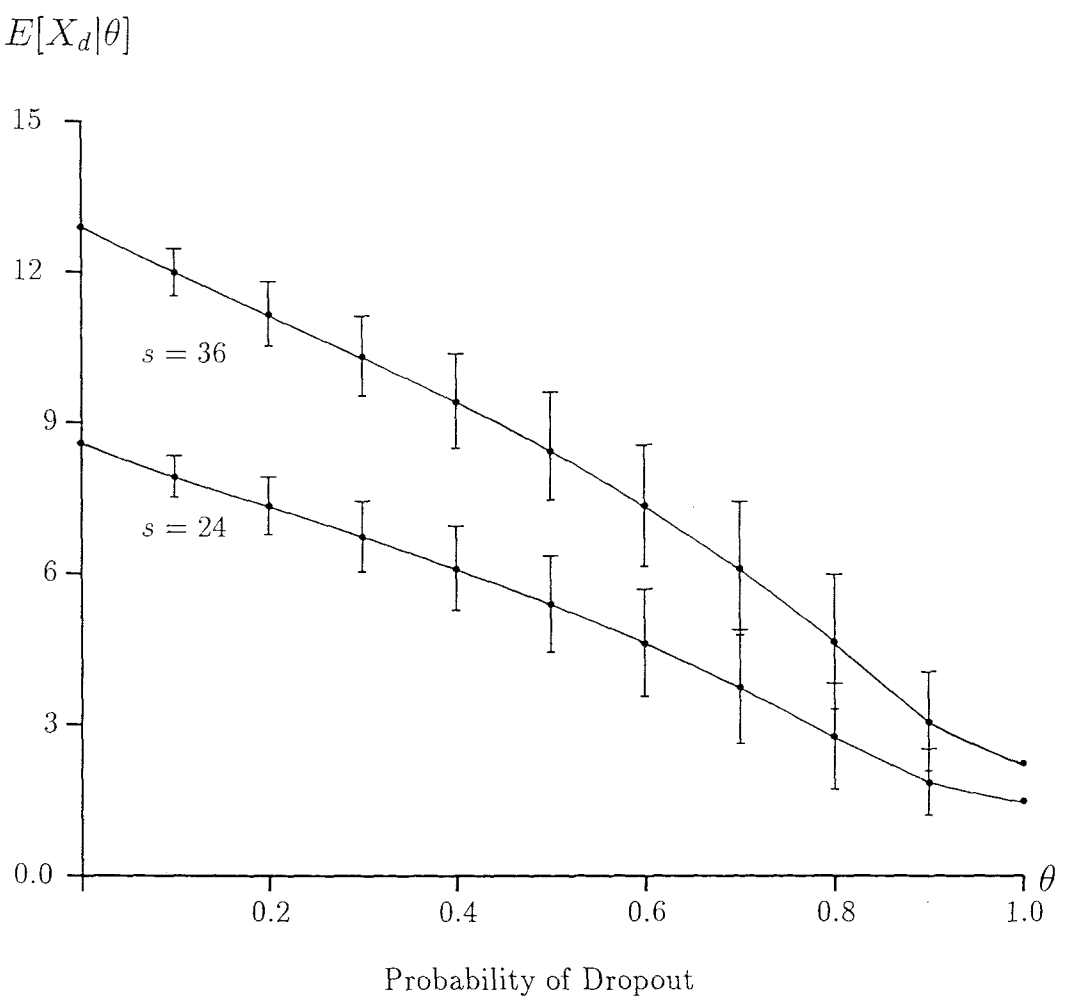


Figure 5.14: Performance for direct treatment comparisons under the MV-criterion of designs based on Figure 5.11 for 24 and 36 subjects, where the bars denote $E[X_d|\theta] \pm \sqrt{\text{Var}[X_d|\theta]}$.

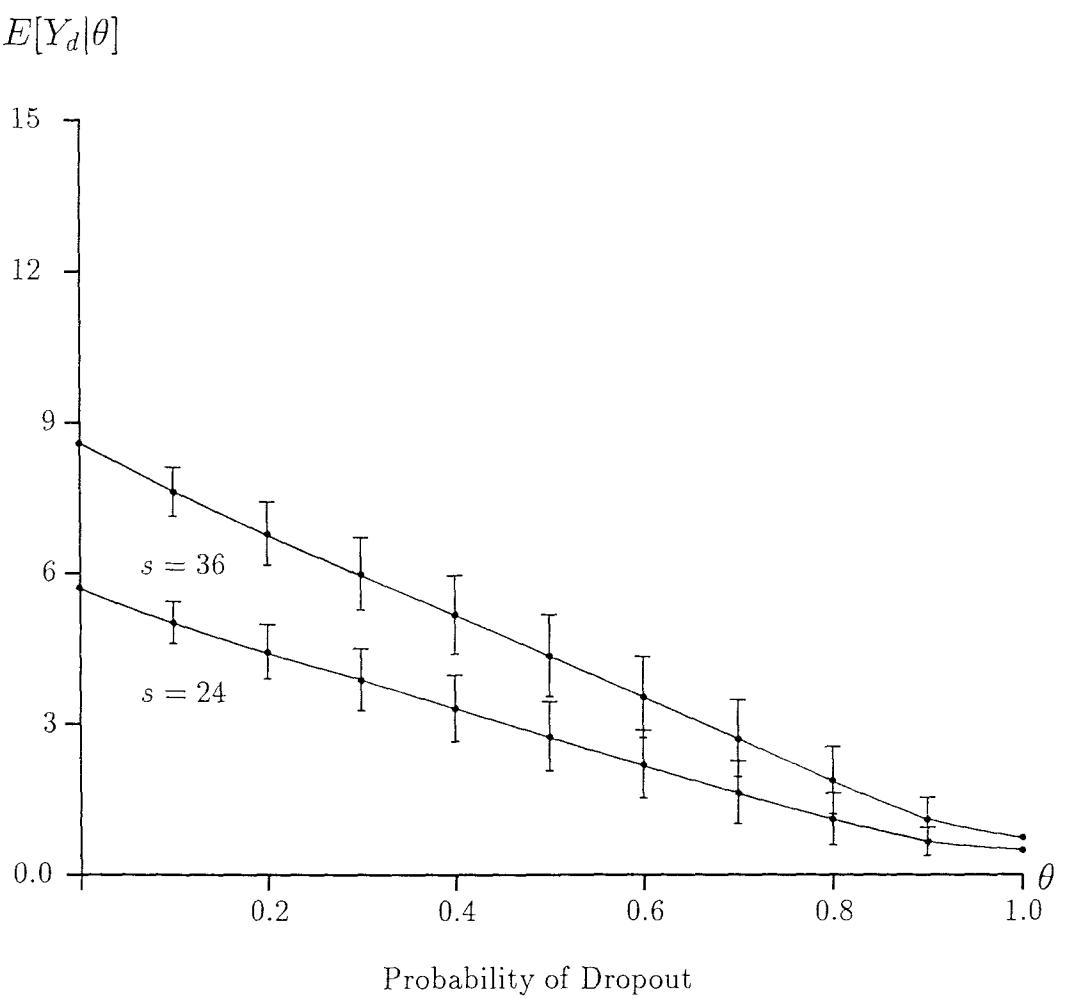


Figure 5.15: Performance for direct treatment comparisons under the MV-criterion of designs based on Figure 5.11 for 24 and 36 subjects, where the bars denote $E[Y_d|\theta] \pm \sqrt{\text{Var}[Y_d|\theta]}$.

change with θ, for each design under the MV-criterion.

Examining the trends of the mean performance measures given in Tables 5.5.1 - 5.5.6 and 5.6.1 - 5.6.6 and illustrated in Figures 5.12 - 5.15 we observe that for any particular value of θ, each of the respective mean performance measures increases as the number of subjects is increased. Also, we note that there is a gradual reduction in the value of each mean performance measure as θ increases. In addition we observe that the variance of the mean performance measures is small and the difference between the mean performance measures obtained under either criterion is not large which suggests that the spread of the variances amongst the pairwise treatment comparisons for each design will never be large.

Thus far we have investigated separately the performance, subject to final period dropout, of the three classes of designs built from replicating each of the designs given in Figures 5.1, 5.6 and 5.11. In the next sections the performance of the designs are compared and recommendations on the choice of design for different experimental situations are given. The recommendations are dependent on the assumption that the simple model of equation (1.1) describes the observations. Both the orthogonal residual effects design (Figure 5.6) and the design of Figure 5.11 involve treatment sequences in which a treatment is followed by a second application of the same treatment. Caution in interpreting the comparisons is needed if the appropriateness of the model is in doubt, see Section 1.3.2.

5.4 Comparison of Designs: $\theta = 0.0, 0.1, \ldots, 1.0$

In this section, the performance subject to final period dropout, of three of the designs investigated in the earlier sections of this chapter are compared so that appropriate recommendations can be made concerning their use. Throughout this section a subset of all possible values of θ, namely $\theta = 0.0, 0.1, \ldots, 1.0$, has been investigated. All the conclusions concerning design selection are therefore limited to the values of θ considered. In Section 5.5 the investigation is extended to examine

the relative performance of each design for $0 \leq \theta \leq 1$.

The designs to be compared, labelled I, II and III are as follows:

I A pair of Williams squares $d(3, 6, 1, 3)$. This is the design given in Figure 5.1 with six subjects in total.

II An orthogonal residual effects design $d(3, 6, 1, 3)$. This is the design given in Figure 5.6 with six subjects.

III The design $d(3, 6, 1, 3)$ given in Figure 5.11 with six subjects.

Assume that we wish to compare all the pairwise direct and first-order carry-over treatment effects. Using equations (2.4), (2.5), (2.6) and (2.7) with the A-criterion we can obtain a summary of the average variance of the direct and carry-over treatment effects for each design. These summary measures have been given in Tables 5.1.1, 5.3.1 and 5.5.1 respectively.

Comparisons of the graphs of the means of X_d and Y_d against θ, for designs I, II, and III, are given in Figures 5.16 and 5.17, respectively. We observe that the range of values for the mean of X_d and Y_d from $\theta = 0.0$ - $\theta = 1.0$ is different for each design. For design I the ranges are 2.40-0.38 and 1.33-0.13 respectively; for design II, 2.00-0.38 and 1.67-0.13 respectively; for design III, 2.15-0.38 and 1.43-0.13 respectively. An examination of the ranges shows that the orthogonal residual effects design, design II, has the smallest range of values for the estimation of the direct treatment effects but has the largest range for the estimation of the first-order carry-over treatment effects. In contrast, the pair of Williams squares, design I, has the smallest range of values for the estimation of the first-order carry-over treatment effects and the largest range for the estimation of the direct treatment effects. We conclude that design II is the least sensitive to the choice of θ for the estimation of the direct treatment effects and design I is the least sensitive to the choice of θ for the estimation of the carry-over treatment effects. In contrast to the comparisons of designs given in Sections 2.8 and 4.6, the designs which demonstrate the least

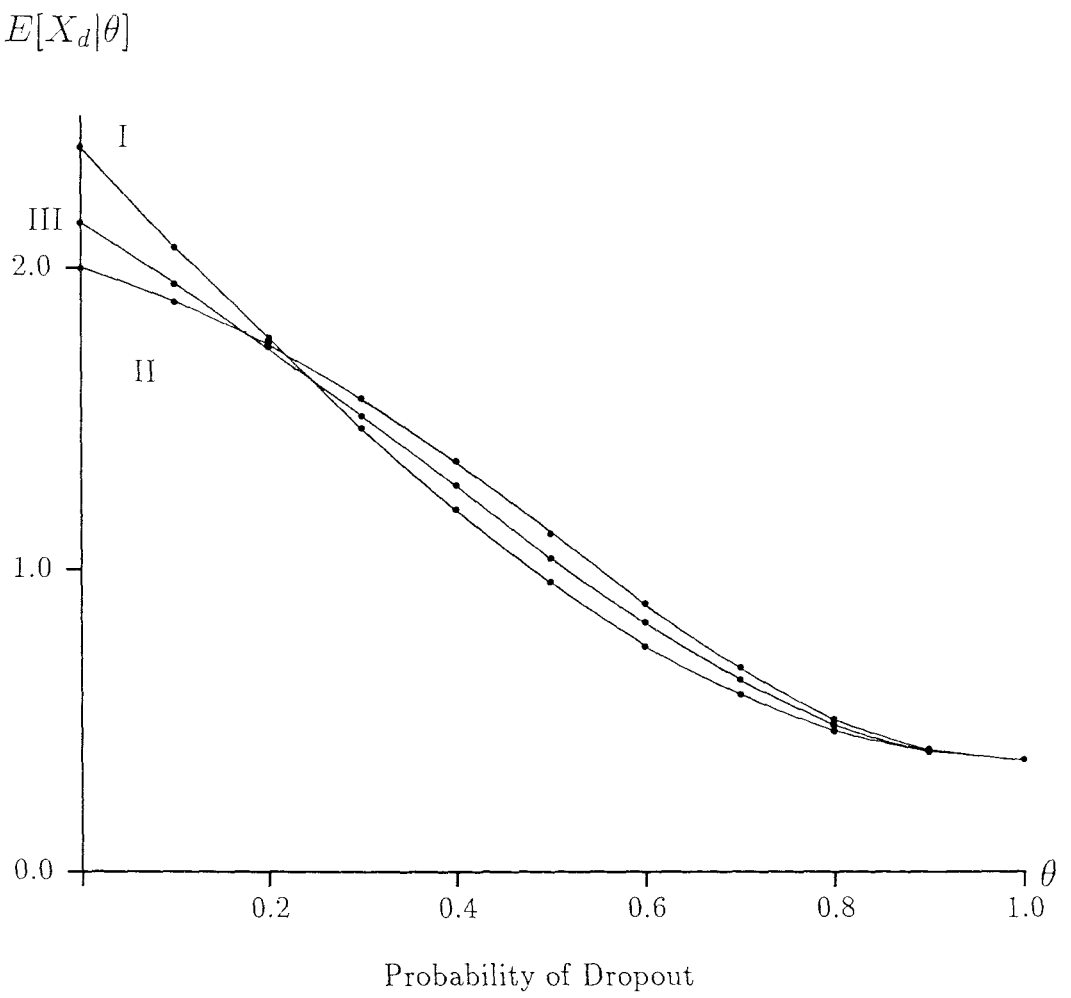


Figure 5.16: Comparison of the graphs for showing the mean of X_d, under the A-criterion, for designs I, II, and III.

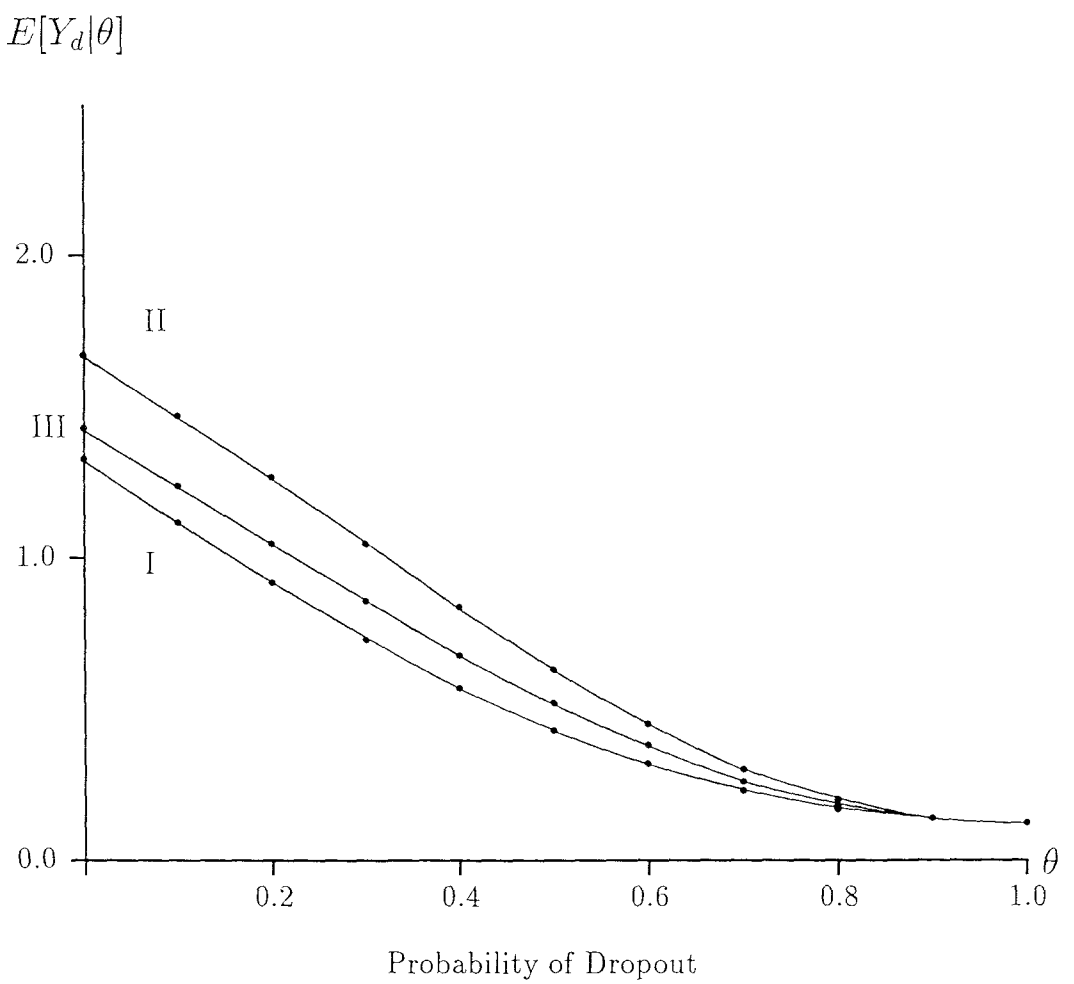


Figure 5.17: Comparison of the graphs for showing the mean of Y_d, under the A-criterion, for designs I, II, and III.

sensitivity to the choice of θ do not necessarily represent the "best" choice of design. This is because the designs being compared are identical in the first two periods only differing in the final period. Consequently, it is when $\theta = 1.0$ rather than when $\theta = 0.0$ that the values of the mean performance measure $E[X_d|\theta]$ are equal for each design. Similarly, when $\theta = 1.0$, the values of the mean performance measure $E[Y_d|\theta]$ are equal for each design. In the comparisons given in Section 2.8 and 4.6, although the designs compared are different, they are all uniform balanced designs and hence have identical performance measures when no dropouts occur. In these circumstances designs with larger ranges of values for the mean performance measures represent designs which experience a greater loss of information as increasing numbers of subjects are lost in the final period. Consequently those designs most sensitive to the choice of θ are also the poorer designs. For the designs compared in this section, this is no longer true. The design whose mean performance changes most from $\theta = 0.0$- $\theta = 1.0$ is the design which has the largest mean performance measure when complete and when only a few subjects are lost in the final period. Thus a design with less sensitivity to θ does not necessarily have greater robustness to final period dropout.

In practice it will be unusual to proceed with a three period study if it is anticipated that large numbers of subjects will dropout during the third period. Consequently designs which can be shown to have higher mean values for X_d and/or Y_d when the value of θ is small are of practical worth.

From Tables 5.1.1, 5.3.1, and 5.5.1 we observe that when $\theta = 0.0$ design I has the largest mean value for X_d, while design II has the largest mean value for Y_d. The result concerning the first-order carry-over effects is as expected since Kunert (1984) proved that, over the class of all cross-over designs in which $t = p = 3$ involving a total of six subjects, the universally optimal design for the estimation of the first-order carry-over treatment effects is an orthogonal residual effects design.

By considering the mean values of X_d and Y_d over the entire range of possible values for θ, it is possible to investigate whether the design which gives the largest

mean performance measure when no dropouts occur also performs better than competing designs when final period dropouts are anticipated. In particular, whether the orthogonal residual effects design which is known to be optimal for the estimation of the first-order carry-over treatment effects when complete, performs better than competing designs when dropouts may occur. We do this by considering the design selection criteria of Section 2.8.1.

The application of the criteria is not as straight forward as in previous examples. Unlike the designs compared in Chapter 4, one design does not emerge as being "better" than the others in the sense of having larger values for $\mathrm{E}[X_d|\theta]$ and $\mathrm{E}[Y_d|\theta]$ with correspondingly smaller values for $\mathrm{Var}[X_d|\theta]$ and $\mathrm{Var}[Y_d|\theta]$ across the entire range of possible values of θ. Consequently, if the probability of final period dropout is changed, the design recommended under the proposed criteria may also change.

From the tables of summary measures for each design we observe that a clear ordering of the designs does exist for the estimation of first-order carry-over treatment effects alone. The ordering is:

$$E[Y_{d_{II}}|\theta] \geq E[Y_{d_{III}}|\theta] \geq E[Y_{d_I}|\theta]$$

and

$$Var[Y_{d_{II}}|\theta] \leq Var[Y_{d_{III}}|\theta] \leq Var[Y_{d_I}|\theta].$$

and holds when both the A- and MV-criteria are employed.

Hence, if the estimation of the first-order carry-over treatment effects is the primary concern, we recommend the use of design II, that is the orthogonal residual effects design, for all possible values of θ.

In practice, however, it is much more common for the estimation of the direct treatment effects to be the major concern of an experimenter. In these circumstances it may be appropriate to either disregard the information obtained concerning the estimation of first-order carry-over effects or seek a design which gives a higher value for $\mathrm{E}[X_d|\theta]$ and, if possible, reduced values for $\mathrm{Var}[X_d|\theta]$ by allowing a decrease in $EY_d|\theta]$ and/or an increase in $\mathrm{Var}[Y_d|\theta]$.

We now establish which design to recommend for each of the values of θ considered.

5.4.1 Design Selection Based on the A-criterion

A summary of the results obtained using the design selection criteria of Section 2.8.1, for the estimation of the direct treatment effects with the A-criterion, can be obtained by examining Tables 5.1.1, 5.3.1 and 5.5.1. The summary is given in Table 5.7 in which the designs are ordered according to the magnitude of $E[X_d|\theta]$ and $\text{Var}[X_d|\theta]$ for $\theta = 0.0, 0.0, \ldots, 1.0$. The comparison of the designs summarised in Table 5.7 shows that no one single design out-performs both of the other designs for all values of θ investigated. However, apart from when $\theta = 0.2$ when design III has the smallest mean performance measure, design III consistently out-performs one of the alternative designs. When $\theta \leq 0.1$ design III out-performs design II and when $\theta \geq 0.3$ design III out-performs design I. Design III is a compromise between the other designs, having a final period which makes the first square of the design identical to that of design I and the second square identical to that of design II. It follows that design III has several implementable designs in common with the other designs. It is therefore unsurprising that, for most of the values of θ for which design I is out-performed by design II, it is also out performed by design III. Similarly, it is unsurprising that, for most of the values of θ for which design II is out-performed by design I, it is also out-performed by design III.

When $0.3 \leq \theta \leq 0.5$ we recommend the use of design II since it has the largest mean value of X_d together with the correspondingly smallest variance. When $\theta \leq 0.2$ or when $\theta \geq 0.6$, the design with the largest mean value of X_d is also the design with the largest variance. In these circumstances, provided the variance is not very different from that obtained for each of the other designs, the preferred design is the one with the largest mean performance measure. When $\theta \leq 0.2$ this is design I and when $\theta \geq 0.6$ this is design II.

Table 5.7: Comparison of the mean and variance of X_d, under the A-criterion, for designs I, II and III. The designs are given in decreasing order of the mean and increasing order of the variance.

θ	Order of $E[X_d\|\theta]$
0.0, 0.1	I, III, II
0.2	I, II, III
$0.3, \ldots, 1.0$	II, III, I
θ	Order of $\text{Var}[X_d\|\theta]$
$0.0, \ldots, 0.5$	II, III, I
$0.6, \ldots, 1.0$	I, III, II

5.4.2 Design Selection Based on the MV-criterion

Alternatively, we may wish to compare the relative performance of the designs I-III using the summary measures provided by equations (2.4), (2.5), (2.6) and (2.7) and the MV-criterion, see Tables 5.2.1, 5.4.1 and 5.6.1. Comparisons of the graphs of the mean of X_d and Y_d against θ, for $\theta = 0.0, 0.1, \ldots 1.0$, are given in Figures 5.18 and 5.19 respectively.

When the estimation of the first-order carry-over treatment effects is most important, the use of the design selection criteria of Section 2.8.1 with the MV-criterion again leads to the recommendation of design II. This is because, as before, the mean values of Y_d are always the largest with the correspondingly smallest variances for any probability of final period dropout for the values of θ considered.

For experiments in which the estimation of the direct treatment effects is at least as important as the estimation of the first-order carry-over treatment effects, then from Figures 5.18 and 5.19 the recommended design depends upon the particular value of θ considered. If θ changes then so may the recommended design.

A summary of the results obtained using the design selection criteria of Section 2.8.1, for the estimation of the direct treatment effects with the MV-criterion can

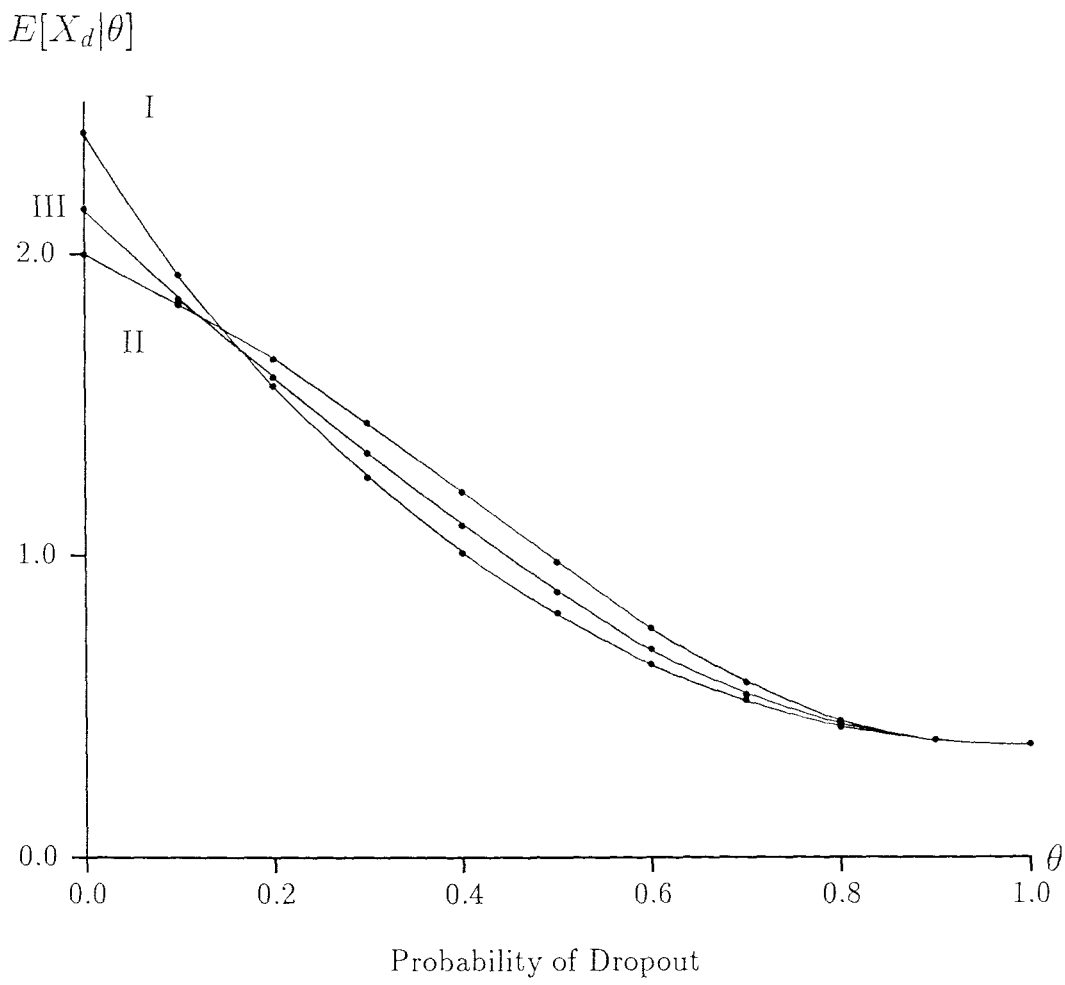


Figure 5.18: Comparison of the graphs for showing the mean of X_d, under the MV-criterion, for designs I, II, and III.

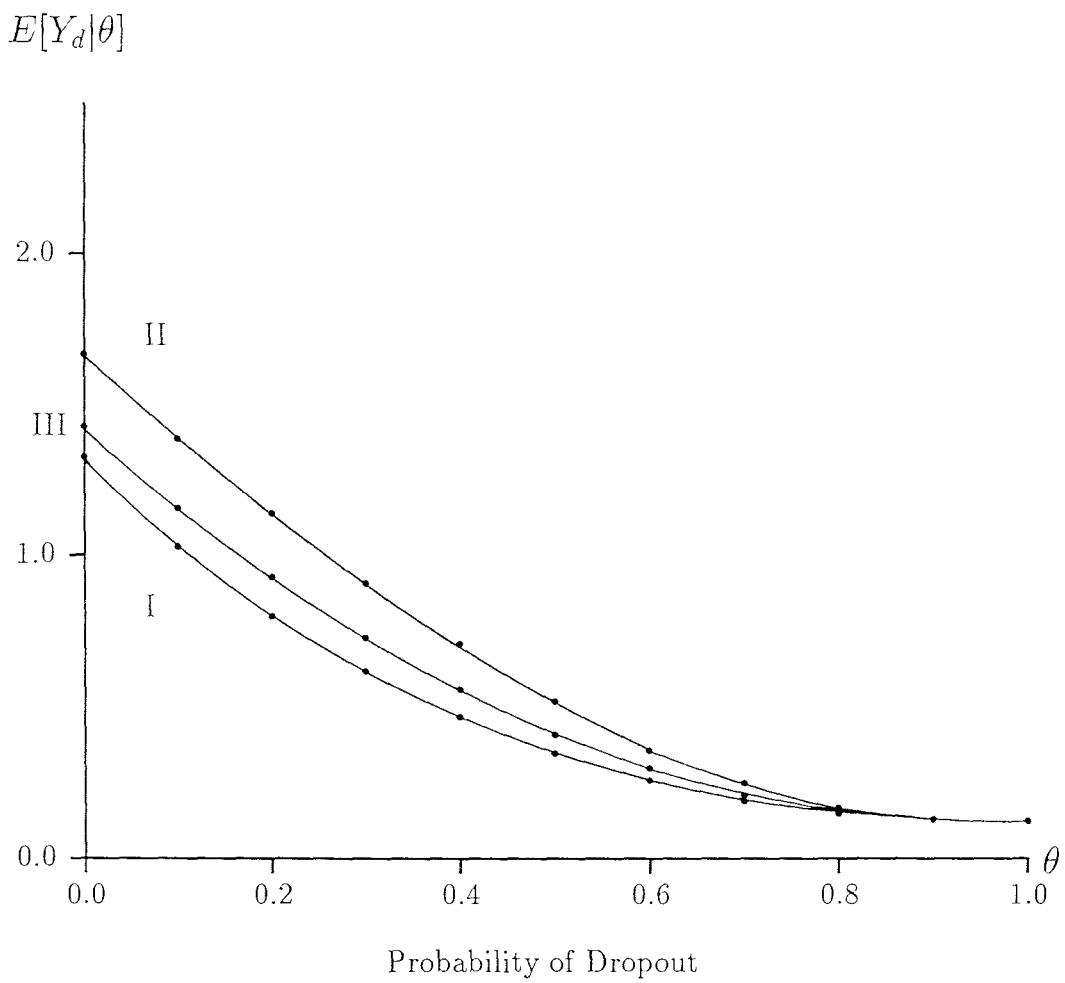


Figure 5.19: Comparison of the graphs for showing the mean of Y_d, under the MV-criterion, for designs I, II, and III.

be obtained by examining Tables 5.2.1, 5.4.1 and 5.6.1. The summary is given in Table 5.8 in which the designs are ordered according to the magnitude of $E[X_d|\theta]$ and $\text{Var}[X_d|\theta]$ for $\theta = 0.0, 0.1, \ldots, 1.0$. The comparison of the designs summarised in Table 5.8 shows that for each value of θ considered the design with the largest mean performance measure never has the smallest variance. Again in these circumstances provided that the variance is not large the preferred design will be the design with the largest mean performance measure. When $\theta \leq 0.1$ this is design I and when $\theta \geq 0.2$ this is design II.

Table 5.8: Comparison of the mean and variance of X_d, under the MV-criterion, for designs I, II and III. The designs are given in decreasing order of the mean and increasing order of the variance.

θ	Order of $E[X_d\|\theta]$
0.0, 0.1	I, III, II
$0.2, \ldots, 0.8$	II, III, I
0.9, 1.0	II, I, III
θ	Order of $\text{Var}[X_d\|\theta]$
0.0, 0.1	II, III, I
$0.2, \ldots, 1.0$	I, III, II

5.5 Comparison of Designs: $0 \leq \theta \leq 1$

Ideally, in order to determine the exact value(s) of θ for which a particular design is recommended, it is necessary to determine the behaviour of the mean performance measures for each competing design across all possible values of θ. In the following subsections, we show how this can be achieved by considering the mean performance measures, $E[X_d|\theta]$ and $E[Y_d|\theta]$, obtained using firstly the A- and secondly the MV-criterion, as functions of θ.

5.5.1 Mean Performance Based on the A-criterion

Let designs I, II and III be the three planned designs $d(3,6,1,3)$ considered in Section 5.4 and given in Figures 5.1, 5.6 and 5.11 respectively.

Using equations (2.4) and (2.6) with the A-criterion we can obtain the mean of the performance measures X_d and Y_d as polynomials in terms of θ.

Let $f_j(\theta)$ denote $E[X_d|\theta]$ for design j where j =I, II or III, then

$$\begin{aligned} f_I(\theta) &= 2.4(1-\theta)^6 + 11.077\theta(1-\theta)^5 + 19.815\theta^2(1-\theta)^4 + \\ &\quad 17.277\theta^3(1-\theta)^3 + 8.045\theta^4(1-\theta)^2 + 2.25\theta^5(1-\theta) + 0.375\theta^6 \quad (5.1) \\ f_{II}(\theta) &= 2.0(1-\theta)^6 + 11.105\theta(1-\theta)^5 + 23.995\theta^2(1-\theta)^4 + \\ &\quad 23.462\theta^3(1-\theta)^3 + 8.762\theta^4(1-\theta)^2 + 2.25\theta^5(1-\theta) + 0.375\theta^6 \quad (5.2) \\ f_{III}(\theta) &= 2.15(1-\theta)^6 + 10.968\theta(1-\theta)^5 + 21.987\theta^2(1-\theta)^4 + \\ &\quad 20.649\theta^3(1-\theta)^3 + 8.473\theta^4(1-\theta)^2 + 2.25\theta^5(1-\theta) + 0.375\theta^6 \quad (5.3) \end{aligned}$$

Similarly, let g_j denote $E[Y_d|\theta]$ for design j where j =I, II and III, then

$$\begin{aligned} g_I(\theta) &= 1.333(1-\theta)^6 + 5.76\theta(1-\theta)^5 + 9.527\theta^2(1-\theta)^4 + \\ &\quad 7.371\theta^3(1-\theta)^3 + 2.869\theta^4(1-\theta)^2 + 0.75\theta^5(1-\theta) + 0.125\theta^6 \quad (5.4) \\ g_{II}(\theta) &= 1.667(1-\theta)^6 + 8.158\theta(1-\theta)^5 + 14.98\theta^2(1-\theta)^4 + \\ &\quad 11.746\theta^3(1-\theta)^3 + 3.162\theta^4(1-\theta)^2 + 0.75\theta^5(1-\theta) + 0.125\theta^6 \quad (5.5) \\ g_{III}(\theta) &= 1.433(1-\theta)^6 + 6.706\theta(1-\theta)^5 + 11.819\theta^2(1-\theta)^4 + \\ &\quad 9.34\theta^3(1-\theta)^3 + 3.052\theta^4(1-\theta)^2 + 0.75\theta^5(1-\theta) + 0.125\theta^6 \quad (5.6) \end{aligned}$$

From the design selection criteria of Section 2.8.1, a recommended design must have, for some anticipated value of θ, the largest value of $E[X_d|\theta]$ over the class of competing designs. If the design satisfying this condition changes, according to the value of θ selected, then one or more of the curves describing the behaviour of

$E[X_d|\theta]$, for each of the competing designs must intersect at some value of θ in the range $0 \leq \theta \leq 1$.

Since the value of θ chosen for design purposes is only an estimate of the anticipated level of final period dropout, it is important to know whether slight changes to this value have any effect upon the design recommended. If the curves describing the behaviour of $E[X_d|\theta]$ intersect at some value of θ close to the anticipated value, then a change in the recommended design could result. Therefore, information on the values of θ at which the curves intersect needs to be taken into account during the planning stage. However, if the difference between the respective performance measures is small, for all values of θ close to the anticipated value, the fact that the curves intersect is less likely to alter the choice of design than when the difference is large.

In order to obtain a full comparison of the performance of designs I, II and III for estimating the average variance of all the pairwise direct treatment effects, it is necessary to establish the points of intersection of f_I, f_{II} and f_{III} given in equations (5.1), (5.2) and (5.3) respectively. It will then be possible to determine the value(s) of θ at which one design begins to out-perform another; that is the value(s) of θ for which the recommended design changes.

In order to find the points of intersection it is necessary to obtain the roots of the following equations:

$$f_I - f_{II} = 0 \tag{5.7}$$

$$f_I - f_{III} = 0 \tag{5.8}$$

$$f_{II} - f_{III} = 0 \tag{5.9}$$

Designs I and II have identical mean performance measures, $E[X_d|\theta]$, at the values of θ which satisfy equation (5.7). Similarly, designs I and III have identical mean performance measures, $E[X_d|\theta]$, at the values of θ which satisfy equation (5.8) and designs II and III have identical mean performance measures, $E[X_d|\theta]$, at the

values of θ which satisfy equation (5.9).

The only roots of interest for each of equations (5.7), (5.8) and (5.9) are real roots such that $0 \leq \theta \leq 1$. From Section 5.4, since all three designs are identical in the first two periods, they have equal values for the mean of the performance measures X_d and Y_d when $\theta = 1.0$ obtained using either the A- or the MV-criterion. Hence, each of the equations (5.7), (5.8) and (5.9) has a root $\theta = 1.0$.

Further, from Section 5.4, a second root of equation (5.7) lies in the range $0.2 \leq \theta \leq 0.3$, since design I has a larger value for $E[X_d|\theta]$ than designs II when $\theta = 0.2$, while design II has a larger value that design I when $\theta = 0.3$. Similarly, a second root of equations (5.8) also lies in the range $0.2 \leq \theta \leq 0.3$ while a second root of equation (5.9) occurs in the range $0.1 \leq \theta \leq 0.2$.

The roots of equations (5.7), (5.8) and (5.9) can be obtained by using the **solve** or **fsolve** commands in the computer algebra package **MAPLE**. The command **solve** seeks to find the exact solutions to an equation in closed-form while the command **fsolve** carries out numerical analysis procedures such as Newton's method of iteration to find the roots of an equation. The following roots were obtained:

Equation (5.7): $\theta = 0.2060$ and $\theta = 1.0$;
Equation (5.8): $\theta = 0.2287$ and $\theta = 1.0$;
Equation (5.9): $\theta = 0.1766$ and $\theta = 1.0$.

These results confirm the observations made earlier.

Examining these points of intersection together with the individual polynomials representing the mean performance measures $E[X_d|\theta]$ for each design, the relative performance of each design across all possible values of θ in the range $0 \leq \theta \leq 1$ can be determined: see Table 5.9 in which, for ranges of θ values the recommended design is given together with the ordering of the designs with respect to the mean performance measure $E[X_d|\theta]$.

From these results we conclude that, if the aim of an experiment is to compare all the direct treatment effects as accurately as possible, when the probability of final

Table 5.9: Comparison of the mean of X_d, under the A-criterion, for designs I, II, and III. The designs are given in decreasing order of the mean.

θ	Order of $E[X_d\|\theta]$
$0.0000 \leq \theta \leq 0.1766$	I, III, II
$0.1766 \leq \theta \leq 0.2060$	I, II, III
$0.2060 \leq \theta \leq 0.2287$	II, I, III
$0.2287 \leq \theta \leq 1.0000$	II, III, I

period dropout is anticipated to be in the range $0.0 \leq \theta \leq 0.2060$ the recommended design is the pair of Williams squares, design I. If the probability of final period dropout is believed to be in the range $0.2060 \leq \theta \leq 1.0$ the recommended design is the orthogonal residual effects design, design II. However, since there is a change in the recommended design at $\theta = 0.2060$ experiments in which the value of θ is believed to be close to 0.2 may need careful consideration with other factors being taken into account before making the final design selection. These findings are similar to those of Section 5.4 but cover all possible values of θ which could arise rather than a subset of specific values.

Similar procedures can be carried out to obtain a full comparison of the performance of designs I, II and III for estimating the average variance of all the pairwise first-order carry-over treatment effects by investigating the points of intersection between equations (5.4), (5.5) and (5.6). On doing this we discover that, in each case, the only point of intersection occurs when $\theta = 1.0$. Furthermore,

$$E[X_{d_{II}}|\theta] \geq [E[X_{d_{III}}|\theta] \geq [E[X_{d_I}|\theta] \quad \text{for} \quad 0 \leq \theta \leq 1.$$

Hence, regardless of the anticipated value of θ, for experiments which aim to compare all the pairwise first-order carry-over treatment effects as accurately as possible, the recommended design is always design II. It is unlikely, however, that an experiment would only aim to compare first-order carry-over treatment effects.

This information would, therefore, be used in conjunction with that obtained for the estimation of the direct treatment effects and could be particularly influential if the value of θ was anticipated to be around 0.2.

5.5.2 Mean Performance Based on the MV-criterion

In Section 5.5.1, the effect of final period dropout on the estimation of the direct and first-order carry-over treatment comparisons has been examined. If the effect of final period dropout on the maximum variance of the treatment comparisons is sought then similar comparisons can be obtained using performance measures based on the MV-criterion, as described below.

Let designs I, II and III be the three planned designs $d(3,6,1,3)$ considered in Section 5.5.1 and given in Figures 5.1, 5.6, and 5.11 respectively. Using equations (2.4) and (2.6) with the MV-criterion we can obtain the mean of the performance measures X_d and Y_d as polynomials in terms of θ.

Let $f_j(\theta)$ denote $E[X_d|\theta]$ for design j where j =I, II or III then,

$$\begin{aligned} f_I(\theta) &= 2.4(1-\theta)^6 + 9.0\theta(1-\theta)^5 + 16.728\theta^2(1-\theta)^4 + \\ &\quad 14.247\theta^3(1-\theta)^3 + 6.576\theta^4(1-\theta)^2 + 2.25\theta^5(1-\theta) + 0.375\theta^6 \quad (5.10) \\ f_{II}(\theta) &= 2.0(1-\theta)^6 + 10.333\theta(1-\theta)^5 + 22.025\theta^2(1-\theta)^4 + \\ &\quad 19.078\theta^3(1-\theta)^3 + 6.34\theta^4(1-\theta)^2 + 2.25\theta^5(1-\theta) + 0.375\theta^6 \quad (5.11) \\ f_{III}(\theta) &= 2.15(1-\theta)^6 + 9.736\theta(1-\theta)^5 + 18.923\theta^2(1-\theta)^4 + \\ &\quad 16.794\theta^3(1-\theta)^3 + 6.020\theta^4(1-\theta)^2 + 2.25\theta^5(1-\theta) + 0.375\theta^6 \quad (5.12) \end{aligned}$$

Similarly, let $g_j(\theta)$ denote $E[Y_d|\theta]$ for design j where j =I, II or III, then

$$\begin{aligned} g_I(\theta) &= 1.333(1-\theta)^6 + 4.571\theta(1-\theta)^5 + 7.682\theta^2(1-\theta)^4 + \\ &\quad 5.919\theta^3(1-\theta)^3 + 2.155\theta^4(1-\theta)^2 + 0.75\theta^5(1-\theta) + 0.125\theta^6 \quad (5.13) \\ g_{II}(\theta) &= 1.667(1-\theta)^6 + 6.966\theta(1-\theta)^5 + 12.685\theta^2(1-\theta)^4 + \\ &\quad 8.971\theta^3(1-\theta)^3 + 2.155\theta^4(1-\theta)^2 + 0.75\theta^5(1-\theta) + 0.125\theta^6 \quad (5.14) \end{aligned}$$

$$g_{III}(\theta) = 1.433(1-\theta)^6 + 5.679\theta(1-\theta)^5 + 9.5\theta^2(1-\theta)^4 + 6.976\theta^3(1-\theta)^3 + 2.019\theta^4(1-\theta)^2 + 0.75\theta^5(1-\theta) + 0.125\theta^6 \quad (5.15)$$

As in Section 5.5.1, the values of θ at which the recommended design may change are obtained by investigating the roots of the appropriate difference equations. On examining these values together with the polynomials representing $E[X_d|\theta]$ and $E[Y_d|\theta]$ we can determine the relative performance of each design over the entire range of possible θ values, $0 \leq \theta \leq 1$. This is summarised in Table 5.10 in which, for ranges of θ values, the recommended design is given together with the ordering of the designs with respect to the magnitude of the mean performance measures $E[X_d|\theta]$ and $E[Y_d|\theta]$.

Table 5.10: Comparison of the mean of X_d and Y_d, under the MV-criterion, for designs I, II and III. The designs are given in decreasing order of the mean.

θ	Order of $E[X_d\|\theta]$
$0.0000 \leq \theta \leq 0.1224$	I, III, II
$0.1224 \leq \theta \leq 0.1446$	I, II, III
$0.1446 \leq \theta \leq 0.1650$	II, I, III
$0.1650 \leq \theta \leq 0.8429$	II, III, I
$0.8429 \leq \theta \leq 0.9556$	II, I, III
$0.9556 \leq \theta \leq 1.0000$	I, II, III
θ	Order of $E[Y_d\|\theta]$
$0.0000 \leq \theta \leq 0.9028$	II, I, III
$0.9028 \leq \theta \leq 1.0000$	II, III, I

From Table 5.10 we observe that when $0.1446 \leq \theta \leq 0.9556$ the recommended design for estimating both the direct and first-order carry-over effects is design II. For all other values of θ the preferred design for estimating the direct treatment effects is design I. The preferred design for the estimation of the first-order carry-over effects is design II when $0 \leq \theta \leq 1$.

5.6 Comparisons for Larger Numbers of Subjects

In Section 5.5 the robustness to dropout of designs I, II and III was investigated. The designs involved only six subjects, that is one subject was allocated at random to each of the distinct treatment sequences of Figures 5.1, 5.6, and 5.11. In this section, the relative performances are investigated of designs formed by allocating larger numbers of subjects ($n \leq 4$) to each of the distinct treatment sequences in the figures. In order to simplify a comparison of the designs with each other and those of the previous sections, any design formed by allocating equal subject numbers to the sequences of Figures 5.1, 5.6 and 5.11 will be labelled designs I-III respectively.

The approach used is identical to that of Sections 5.5.1 and 5.5.2. The polynomials derived for the mean performance measures for the direct and first-order carry-over treatment effects (analogous to (5.1)-(5.6) and (5.10)-(5.15)) include terms of order 24 in θ. Due to their complexity the polynomials are not explicitly given here, but they are used to derive the results in Tables 5.11 and 5.12, for the A-criterion, and in Tables 5.13 and 5.14, for the MV-criterion.

In the tables the designs are ordered according to the size of the mean performance measure, where the first design listed has the largest mean performance measure. For any particular experiment, the experimenters can use this information to guide their choice of design for the anticipated value of θ as illustrated in Examples 5.1 and 5.2.

Example 5.1 Suppose an experiment is proposed to compare direct treatment effects using 18 subjects, and the probability of dropout is thought to be in the region of 0.1. From Table 5.11 we observe that the recommended design is design I, when $E[X_d|\theta]$ is obtained using the A-criterion. Similarly, from Table 5.13 we observe that the recommended design is also design I when the performance measures are obtained using the MV-criterion.

Example 5.2 Suppose an experiment is proposed to compare direct treatment effects using 24 subjects, and the anticipated value of θ is in the region of 0.3. From Tables 5.11 and 5.13 we observe that the recommended design is design I when the performance measures are obtained using the A-criterion, but design II when the performance measures are obtained using the MV-criterion.

For the experiment described in Example 5.1, the experimenters should have no difficulty in selecting design I since this is the recommended design for performance measures obtained using both the A- and MV-criteria. Design selection is less straight forward for the experiment described in Example 5.2. In this case the recommended design is not the same for performance measures obtained using different optimality criteria. In these circumstances the priorities of the experimenters must guide the choice of design. For example, if minimising the expected average variance of the pairwise treatment comparisons is a priority, then the most appropriate design to use is design I. If however minimising the expected maximum variance of the pairwise treatment comparisons is essential, then design II will be the preferred design.

From Table 5.11, we observe that when θ is very small, the preferred design for the estimation of direct treatment effects is design I, for each size of study investigated. However, in each case, there reaches a point when design II begins to out-perform design I and its superiority continues for all subsequent values of θ. Let θ_0 be the value of θ at which design II begins to out-perform design I. From Table 5.11 we see that the value of θ_0 increases with the number of subjects allocated to each treatment sequence. For example, when the number of subjects in the study is six , $\theta_0 = 0.2060$. When the number of subjects on the study is increased to 24 $\theta_0 = 0.3469$. Note that, right across the ordering of the designs, the larger the value of n the greater the value of θ at which the change in the ordering of the designs occurs. It is conjectured that for studies involving larger numbers of subjects θ_0 will continue to increase in a similar manner. If this is true, then for $\theta \leq 0.3$ and $n \geq 3$,

Table 5.11: Comparisons of the mean of X_d, under the A-criterion, for designs I, II and III and ≤ 24 subjects. The designs are given in decreasing order of the mean.

	θ	Order of $E[X_d\|\theta]$
$n = 1,\ s = 6$	$0.0000 \leq \theta \leq 0.1766$	I, III, II
	$0.1766 \leq \theta \leq 0.2060$	I, II, III
	$0.2060 \leq \theta \leq 0.2287$	II, I, III
	$0.2287 \leq \theta \leq 1.0000$	II, III, I
$n = 2,\ s = 12$	$0.0000 \leq \theta \leq 0.2466$	As previous
	$0.2466 \leq \theta \leq 0.2942$	ordering.
	$0.2942 \leq \theta \leq 0.3397$	
	$0.3397 \leq \theta \leq 1.0000$	
$n = 3,\ s = 18$	$0.0000 \leq \theta \leq 0.2748$	As previous
	$0.2748 \leq \theta \leq 0.3290$	ordering.
	$0.3290 \leq \theta \leq 0.3738$	
	$0.3738 \leq \theta \leq 1.0000$	
$n = 4,\ s = 24$	$0.0000 \leq \theta \leq 0.2894$	As previous
	$0.2894 \leq \theta \leq 0.3469$	ordering.
	$0.3469 \leq \theta \leq 0.3394$	
	$0.3394 \leq \theta \leq 1.0000$	

Table 5.12: Comparisons of the mean of Y_d, under the A-criterion, for designs I, II and III and ≤ 24 subjects. The designs are given in decreasing order of the mean.

	θ	Order of $E[Y_d\|\theta]$
$n = 1,\ s = 6$	$0.0 \leq \theta \leq 1.0$	II, III, I
$n = 2,\ s = 12$	$0.0 \leq \theta \leq 1.0$	As previous ordering.
$n = 3,\ s = 18$	$0.0 \leq \theta \leq 1.0$	As previous ordering.
$n = 4,\ s = 24$	$0.0 \leq \theta \leq 1.0$	As previous ordering.

Table 5.13: Comparisons of the mean of X_d, under the MV-criterion, for designs I, II and III and ≤ 24 subjects. The designs are given in decreasing order of the mean.

	θ	Order of $E[X_d\|\theta]$
$n = 1,\ s = 6$	$0.0000 \leq \theta \leq 0.1224$	I, III, II
	$0.1224 \leq \theta \leq 0.1446$	I, II, III
	$0.1446 \leq \theta \leq 0.1650$	II, I, III
	$0.1650 \leq \theta \leq 0.8429$	II, III, I
	$0.8429 \leq \theta \leq 0.9556$	II, I, III
	$0.9556 \leq \theta \leq 1.0000$	I, II, III
$n = 2,\ s = 12$	$0.0000 \leq \theta \leq 0.1723$	As previous
	$0.1723 \leq \theta \leq 0.2243$	ordering.
	$0.2243 \leq \theta \leq 0.2648$	
	$0.2648 \leq \theta \leq 0.8740$	
	$0.8740 \leq \theta \leq 0.9613$	
	$0.9613 \leq \theta \leq 1.0000$	
$n = 3,\ s = 18$	$0.0000 \leq \theta \leq 0.2038$	As previous
	$0.2038 \leq \theta \leq 0.2621$	ordering.
	$0.2621 \leq \theta \leq 0.3167$	
	$0.3167 \leq \theta \leq 0.8995$	
	$0.8995 \leq \theta \leq 0.9661$	
	$0.9661 \leq \theta \leq 1.0000$	
$n = 4,\ s = 24$	$0.0000 \leq \theta \leq 0.2038$	As previous
	$0.2229 \leq \theta \leq 0.2726$	ordering.
	$0.2726 \leq \theta \leq 0.3437$	
	$0.3437 \leq \theta \leq 0.9055$	
	$0.9055 \leq \theta \leq 0.9751$	
	$0.9751 \leq \theta \leq 1.0000$	

Table 5.14: Comparisons of the mean of Y_d, under the MV-criterion, for designs I, II and III and ≤ 24 subjects. The designs are given in decreasing order of the mean.

	θ	Order of $E[Y_d\|\theta]$
$n = 1,\ s = 6$	$0.0000 \leq \theta \leq 0.9028$	II, I, III
	$0.9028 \leq \theta \leq 1.0000$	II, III, I
$n = 2,\ s = 12$	$0.0000 \leq \theta \leq 0.9170$	II, III, I
	$0.9170 \leq \theta \leq 0.9986$	II, I, III
	$0.9986 \leq \theta \leq 1.0000$	I, II, III
$n = 3,\ s = 18$	$0.0000 \leq \theta \leq 0.9291$	As previous
	$0.9291 \leq \theta \leq 0.9984$	ordering.
	$0.9984 \leq \theta \leq 1.0000$	
$n = 4,\ s = 24$	$0.0000 \leq \theta \leq 0.9418$	As previous
	$0.9418 \leq \theta \leq 0.9950$	ordering.
	$0.9950 \leq \theta \leq 1.0000$	

amongst the three designs investigated in this section, the pair of Williams squares, design I, will always be the preferred design for the estimation of direct treatment effects.

Examining Table 5.13, we observe that the comments made concerning design selection when the performance measures for X_d are obtained using the MV-criterion are similar to those made when these are obtained using the A-criterion. There are two notable differences. Firstly, when θ is very large, $\theta \geq 0.95$, design I begins to out-perform design II again. This is of little practical importance, however, since it would be unthinkable for a cross-over design to proceed for such a large anticipated value of θ. The more important difference is that the value of θ at which design II begins to out-perform design I is consistently smaller when the performance measures are calculated using the MV-criterion in preference to the A-criterion. The choice of study needs very careful consideration when the value of θ is anticipated to be in the region where a different design is recommended under each criterion.

For all sizes of study considered, Tables 5.12 and 5.14 show that the recommended design for the estimation of first-order carry-over treatment effects under both criteria is the orthogonal residual effects design, design II, provided that the value of $\theta \leq 0.9$.

For small studies and a small value of θ the recommended design for estimating the direct treatment effects is not the same as that for estimating the first-order carry-over effects. In these circumstances, the priorities of the experimenter should be used to choose the design.

5.7 Designs Formed by Changing the Final Period

In the previous section, the performance subject to final period dropout of three different cross-over designs has been investigated. The common feature of the designs

is that they are identical in the first two treatment periods and differ only in the period under threat; the final period. The findings show that, particularly with respect to the estimation of direct treatment effects, the design which has the largest mean performance measure when $\theta = 0.0$ is out-performed by alternative designs for particular values of $\theta > 0$. The work presented in the previous sections of this chapter raises the issue of whether a design with improved robustness to final period dropout can be found by changing the final period of a pair of Williams squares of side three. In order to investigate this an exhaustive computer search of all possible designs for six subjects was undertaken.

The total number of possible three period designs which can be formed using the first two periods of a pair of Williams squares and allowing the final period of each treatment sequence to involve any of the three treatments is $3^6 = 729$. The designs considered in Sections 5.4 and 5.5 are just three of the total number of designs to be investigated. Note that some of the 729 possible designs can be shown to be isomorphic under a permutation of the treatment labels and of the treatment sequences.

An investigation of the mean and variance of the performance measures X_d and Y_d calculated using both the A- and MV-criteria for all 729 designs has been performed for $\theta = 0.0, 0.1, \ldots, 1.0$. The results can be used to establish which of the designs give rise to "better" mean performance measures for different values of θ and to address the following questions:

(i) Which of the designs give "better" mean performance measures for the direct treatment effects?

(ii) Which of the designs give "better" mean performance measures for the carry-over treatment effects?

(iii) Do any of the designs provide "good" mean performance measures for both the direct and carry-over treatment effects?

5.7.1 Direct Treatment Effects

When the probability of final period dropout is $\theta = 0.0$, the design with the largest mean performance measure $E[X_d|\theta]$, using both the A- and MV-criterion, is the pair of Williams squares given in Figure 5.1. This design is therefore, both A- and MV-optimal, within the class of 729 designs under investigation, for estimating the direct treatment effects. It has already been established in Section 5.4, however, that this design does not always give the largest mean performance measures $E[X_d|\theta]$ when $\theta > 0$.

In practice, it will be unusual to proceed with an experiment in which the probability of final period dropout is anticipated to be greater than 0.3. An examination of the mean performance measures obtained for each of the designs under investigation shows that several designs give larger mean performance measures $E[X_d|\theta]$, obtained using either the A- or MV-criterion, than a pair of Williams squares, for some value of θ in the range $0.0 \leq \theta \leq 0.3$. These designs, together with the pair of Williams squares, are listed in Table 5.15. Note that the designs labelled a, b and c in Table 5.15 are designs I, II and III investigated in Section 5.4. Examining these designs we observe that the designs a, b, c, d and h are all designs which are uniform on the periods. In addition, the remaining three designs may be thought of as being "nearly" uniform on the periods since, in the final period of each design, one treatment occurs three times, one treatment occurs twice and the final treatment occurs once. This is a switch of only one treatment.

The relative performances of the eight designs listed in Table 5.15 for estimating the direct treatment effects using both the A- and MV-criteria, are summarised in Table 5.16 for $\theta = 0.0, \ldots, 0.3$. In the table the designs are given in decreasing order of $E[X_d|\theta]$.

Examining Table 5.16 we observe that when $\theta \geq 0.2$ the pair of Williams squares, design a, no longer gives the largest mean performance measures when calculated under either optimality criterion. In addition, designs a b, d, and h give the largest

Table 5.15: Designs using the first two periods of Figure 5.1 which perform at least as well as Figure 5.1 for the estimation of the direct treatment comparisons when $\theta \leq 0.3$.

a	0	1	2	b	0	1	1	c	0	1	2	d	0	1	2
	1	2	0		1	2	2		1	2	0		1	2	2
	2	0	1		2	0	0		2	0	1		2	0	0
	0	2	1		0	2	2		0	2	2		0	2	1
	1	0	2		1	0	0		1	0	0		1	0	0
	2	1	0		2	1	1		2	1	1		2	1	1

e	0	1	1	f	0	1	1	g	0	1	1	h	0	1	2
	1	2	2		1	2	2		1	2	2		1	2	2
	2	0	0		2	0	1		2	0	0		2	0	1
	0	2	1		0	2	1		0	2	1		0	2	1
	1	0	0		1	0	0		1	0	0		1	0	0
	2	1	1		2	1	0		2	1	0		2	1	0

Table 5.16: Comparisons of the mean of X_d, under the A- and MV-criteria, for designs a-h. The designs are given in decreasing order of the mean.

	A-criterion	MV-criterion
θ	Order of $E[X_d\|\theta]$	Order of $E[X_d\|\theta]$
0.0	a, h, d, c, f, g, b, e	a, h, c, d, b, f, g, e
0.1	a, h, d, f, c, b, g, e	a, h, d, c, b, g, f, e
0.2	h, d, a, b, f, c, e, g	b, d, h, c, a, g, e, f
0.3	d, b, h, f, e, c, g, a	b, d, c, h, e, g, f, a

mean performance measures under at least one criterion, for some value of θ in the given range. Having found four designs which perform well for studies involving just six subjects, an investigation into the relative performance of the four designs for studies involving up to 24 subjects for $0 \leq \theta \leq 1$ was undertaken. A summary of the results for performance measures obtained using both the A- and MV-criteria is given in Tables 5.17 and 5.18 respectively.

Note that the larger the value of n, that is the larger the number of subjects allocated to each treatment sequence, the greater the value of θ at which the change in the ordering of the designs occurs. However, this increase in the value of θ is not rapid. For example, when the performance measures are obtained using the A-criterion and $n = 1$, design a, the pair of Williams squares of side three, ceases to have the largest mean value of X_d when $\theta \geq 0.1070$. When $n = 4$ the mean value of X_d obtained for design a is out-performed when $\theta \geq 0.2134$.

5.7.2 Carry-over treatment effects

Due to Kunert (1984), we know that within the class of designs in which $t = p = 3$ involving six subjects, the universally optimal design for estimating first-order carry-over effects, when dropouts are not considered, is the orthogonal residual effects design given in Figure 5.6. Consequently, this design is universally optimal amongst the 729 designs under investigation when $\theta = 0.0$.

The investigation of the mean performance measures $E[Y_d|\theta]$ obtained for each of the 729 designs shows that, when d^* is the orthogonal residual effects design, then

$$E[Y_{d^*}|\theta] \geq E[Y_d|\theta]$$

for the values of θ in the investigation and when the performance measures are obtained using either the A- or MV-criterion.

Unfortunately, the orthogonal residual effects design does not always give smaller values for $\mathrm{Var}[Y_d|\theta]$ than all the other 729 designs considered. However, the values

Table 5.17: Comparisons of the mean of X_d, under the A-criterion, for designs a, b, d and h and ≤ 24 subjects. The designs are given in decreasing order of the mean.

	θ	Order of $E[X_d\|\theta]$
$n = 1,\ s = 6$	$0.0000 \leq \theta \leq 0.1070$	a, h, d, b
	$0.1070 \leq \theta \leq 0.1602$	h, a, d, b
	$0.1602 \leq \theta \leq 0.2060$	h, d, a, b
	$0.2060 \leq \theta \leq 0.2104$	h, d, b, a
	$0.2104 \leq \theta \leq 0.2569$	d, h, b, a
	$0.2569 \leq \theta \leq 0.3098$	d, b, h, a
	$0.3098 \leq \theta \leq 1.0000$	b, d, h, a
$n = 2,\ s = 12$	$0.0000 \leq \theta \leq 0.1755$	As previous
	$0.1755 \leq \theta \leq 0.2429$	ordering.
	$0.2429 \leq \theta \leq 0.2942$	
	$0.2942 \leq \theta \leq 0.2961$	
	$0.2961 \leq \theta \leq 0.3413$	
	$0.3413 \leq \theta \leq 0.3854$	
	$0.3854 \leq \theta \leq 1.0000$	
$n = 3,\ s = 18$	$0.0000 \leq \theta \leq 0.2009$	As previous
	$0.2009 \leq \theta \leq 0.2745$	ordering.
	$0.2745 \leq \theta \leq 0.3290$	
	$0.3290 \leq \theta \leq 0.3309$	
	$0.3309 \leq \theta \leq 0.3763$	
	$0.3763 \leq \theta \leq 0.4193$	
	$0.4193 \leq \theta \leq 1.0000$	
$n = 4,\ s = 24$	$0.0000 \leq \theta \leq 0.2134$	As previous
	$0.2134 \leq \theta \leq 0.2904$	ordering.
	$0.2904 \leq \theta \leq 0.3469$	
	$0.3469 \leq \theta \leq 0.3476$	
	$0.3476 \leq \theta \leq 0.3950$	
	$0.3950 \leq \theta \leq 0.4383$	
	$0.4383 \leq \theta \leq 1.0000$	

Table 5.18: Comparisons of the mean of X_d, under the MV-criterion, for designs a, b, d and h and ≤ 24 subjects. The designs are given in decreasing order of the mean.

	θ	Order of $E[X_d\|\theta]$
$n = 1,\ s = 6$	$0.0000 \leq \theta \leq 0.1192$	a, h, d, b
	$0.1192 \leq \theta \leq 0.1438$	h, a, d, b
	$0.1438 \leq \theta \leq 0.1446$	h, d, a, b
	$0.1446 \leq \theta \leq 0.1471$	h, d, b, a
	$0.1471 \leq \theta \leq 0.1598$	h, b, d, a
	$0.1598 \leq \theta \leq 0.1684$	b, h, d, a
	$0.1684 \leq \theta \leq 0.7869$	b, d, h, a
	$0.7869 \leq \theta \leq 0.8641$	b, d, a, h
	$0.8641 \leq \theta \leq 0.9556$	b, a, d, h
	$0.9556 \leq \theta \leq 0.9595$	a, b, d, h
	$0.9595 \leq \theta \leq 1.0000$	a, b, h, d
$n = 2,\ s = 12$	$0.0000 \leq \theta \leq 0.1513$	a, h, d, b
	$0.1513 \leq \theta \leq 0.2098$	h, a, d, b
	$0.2098 \leq \theta \leq 0.2248$	h, d, a, b
	$0.2248 \leq \theta \leq 0.2553$	h, d, b, a
	$0.2553 \leq \theta \leq 0.2582$	d, h, b, a
	$0.2582 \leq \theta \leq 0.2623$	d, b, h, a
	$0.2623 \leq \theta \leq 0.8193$	b, d, h, a
	$0.8193 \leq \theta \leq 0.8880$	b, d, a, h
	$0.8880 \leq \theta \leq 0.9613$	b, a, d, h
	$0.9613 \leq \theta \leq 0.9679$	a, b, d, h
	$0.9679 \leq \theta \leq 1.0000$	a, b, h, d

Table 5.18: continued.

	θ	Order of $E[X_d\|\theta]$
$n = 3$, $s = 18$	$0.0000 \leq \theta \leq 0.1636$	a, h, d, b
	$0.1636 \leq \theta \leq 0.2402$	h, d, a, b
	$0.2402 \leq \theta \leq 0.2621$	a, h, d, b
	$0.2621 \leq \theta \leq 0.2929$	h, d, b, a
	$0.2929 \leq \theta \leq 0.3037$	d, h, b, a
	$0.3037 \leq \theta \leq 0.3195$	d, b, h, a
	$0.3195 \leq \theta \leq 0.8435$	b, d, h, a
	$0.8435 \leq \theta \leq 0.9049$	b, d, a, h
	$0.9049 \leq \theta \leq 0.9661$	b, a, d, h
	$0.9661 \leq \theta \leq 0.9751$	a, b, d, h
	$0.9751 \leq \theta \leq 1.0000$	a, b, h, d
$n = 4$, $s = 24$	$0.0000 \leq \theta \leq 0.1731$	As previous
	$0.1731 \leq \theta \leq 0.2596$	ordering.
	$0.2596 \leq \theta \leq 0.2852$	
	$0.2852 \leq \theta \leq 0.3138$	
	$0.3138 \leq \theta \leq 0.3288$	
	$0.3288 \leq \theta \leq 0.3509$	
	$0.3509 \leq \theta \leq 0.8648$	
	$0.8648 \leq \theta \leq 0.9189$	
	$0.9189 \leq \theta \leq 0.9671$	
	$0.9671 \leq \theta \leq 0.9781$	
	$0.9781 \leq \theta \leq 1.0000$	

of $\text{Var}[Y_d|\theta]$ obtained for the orthogonal residual effects design are never large or very different from those obtained for any other design, particularly when θ is small.

5.8 Discussion

In this chapter a study has been presented of the performance of a selection of three treatment, three period designs each involving six different treatment sequences and subject to final period dropout. This has shown that, of the designs investigated for the estimation of first-order carry-over treatment effects, designs formed from replicates of the orthogonal residual effects design of Figure 5.6 are the most robust to final period dropout.

With respect to the estimation of direct treatment effects, the study has shown that the design recommended will depend upon the anticipated probability of final period dropout. This is because the design with the largest value of $E[X_d|\theta]$ is not the same for all possible values of θ. Of the designs investigated when $\theta \leq 0.14$, the design formed from one or more replicates of a pair of Williams squares of side three will always give the largest mean performance measures. As the value of θ increases there will reach a point at which the design will be out-performed by other designs. This value of θ will depend upon the number of subjects allocated to the study. Increasing the number of subjects does not generally affect the overall ordering amongst the performance measures of competing designs it just increases the value of θ at which the ordering changes.

If the aim of a study is to investigate both the direct and first-order carry-over treatment effects, designs formed from replicates of the orthogonal residual effects design are realistic alternatives to the currently favoured designs formed from a pair of Williams squares. Although for small values of θ designs formed from a pair of Williams squares provide large mean values of X_d, the mean values of Y_d are poor in comparison to those of other designs. Regardless of the anticipated value of θ, designs formed from the orthogonal residual effects design have been found to be

the most robust to final period dropout for the estimation of first-order carry-over treatment effects. When θ is small, although the designs do not provide the largest mean values of X_d, the values obtained are very close to those of designs formed from pairs of Williams squares. In addition, for each size of study considered, there is a large region of θ values for which designs formed from replicates of the orthogonal residual effects design give the largest mean values of X_d.

All the designs investigated in this chapter involve six different treatment sequences. Further investigation is required of designs involving different numbers of treatment sequences. These need not be uniform balanced designs since, as shown in this chapter, these will not necessarily be the most robust to final period dropout. Designs formed by combining the treatment sequences of a uniform balanced design with an orthogonal residual effects design are particularly interesting candidates for further investigation. An example of such a design is the eighteen treatment sequences formed using Proposition 5.1. Kunert (1984) has shown this design to be universally better for the estimation of direct treatment effects than a uniform balanced design when dropouts are not considered.

In this chapter designs for three treatment, three period studies have been discussed and some areas for further investigation have been identified. In Chapter 6 further areas for future research are described and conclusions from the current work are given.

Chapter 6

Extensions and Future Work

6.1 Introduction

In this chapter issues which have arisen in this research are discussed and directions for future work are identified. Finally, conclusions from the research project are presented.

6.2 Assessing Designs When Multi-period Dropout May Occur

In Chapter 2 methods for assessing the performance of cross-over designs in the presence of final period dropout were presented. In some practical situations it is unrealistic to assume that the final period is the only stage at which dropouts may occur. In this section we show how to extend the methods presented earlier in this thesis to assess the performance of cross-over designs when subjects may drop out in any period of the study. Note that, when a subject drops out in a particular period, it is assumed that the current and all subsequent observations on the subject are not made, that is that the subject does not return to the study.

A similar approach to that of Section 2.4 can be adopted, provided the following

extensions are made:

(i) The set of all possible implementable designs, D, is extended to include each of the designs d_l where $l = (l_{0,1} \ldots l_{p,1} \ldots l_{0,m} \ldots l_{p,m})$ in which $l_{i,j}$ denotes the number of subjects on sequence j who complete i periods and then drop out during period $i+1$ and $l_{p,j}$ denotes the number of subjects on sequence j who complete the study, where $i = 0, \ldots, p-1$ and $j = 1, \ldots, m$.
Note that the total number of designs in D is then given by $|D| = \left[\binom{n+p}{p} \right]^m$.

(ii) The probability of realising each of the implementable designs is obtained from a multinomial distribution, rather than a binomial distribution, as follows.

Consider a planned design $d(t, m, n, p)$ and assume that each subject has a fixed probability θ_i of completing i periods and then dropping out during period $i+1$ ($i = 0, \ldots, p-1$), where $\theta_p = 1 - \sum_{i=0}^{p-1} \theta_i$ denotes the probability that a subject completes the study. Suppose that there are $l_{i,j}$ subjects ($i = 0, \ldots, p-1$; $j = 1, \ldots, m$) on sequence j who complete i periods and then drop out during period $i+1$ resulting in an implemented design, d_l. Then the probability that d_l is realised is given by

$$P(l|\theta_0, \ldots, \theta_{p-1}) = \prod_{j=1}^{m} \frac{n!}{l_{0,j}! \ldots l_{p,j}!} \theta_0^{l_{0,j}} \ldots \theta_p^{l_{p,j}}. \tag{6.1}$$

Hence using equation (6.1) we can calculate the probability that d_l is the realised design for each $d_l \in D$.

The mean and variance of the performance measures X_d and Y_d, given in Definitions 2.4 and 2.5, can be used to provide summary measures for the performance of the planned design under repeated use in experiments. Note that the probability distributions for X_d and Y_d for given values of θ_i ($i = 0, \ldots, p-1$) are obtained from (6.1) as

$$P(X_d = x|\theta_0, \ldots, \theta_{p-1}) = \sum_{l \in L} P(l|\theta_0, \ldots, \theta_{p-1}), \tag{6.2}$$

where $L = \{l; d_l \in D,\ X_d(d_l) = x\}$ and

$$P(Y_d = y|\theta_0, \ldots, \theta_{p-1}) = \sum_{l \in L} P(l|\theta_0, \ldots, \theta_{p-1}), \tag{6.3}$$

where $L = \{l; d_l \in D,\ Y_d(d_l) = y\}$.

For experiments of a realistic size, obtaining summary performance measures for a planned design from (6.2) and (6.3) can involve a prohibitive amount of computation. Fortunately, the combinatorial theory set out in Chapter 3 can be extended to provide significant computational reductions when assessing cross-over designs in the presence of multi-period dropouts. For example, Table 6.1 lists the number of implementable designs and the corresponding number of equivalence classes into which D can be partitioned for a Williams square of side four and two periods of dropout. From Table 6.1 we observe that the computational reduction achieved is approximately three quarters. The above results have been reported in Low, Lewis, McKay and Prescott (1994).

Table 6.1: Number of equivalence classes and implementable designs for designs based on a Williams square of side four and two periods of dropout.

Number of subjects.	Number of implementable designs.	Number of combinatorial equivalence classes.
12	10000	2530
16	50625	12720
20	194481	48741
24	614656	153874
28	1679616	420246
32	4100625	1025685
36	9150625	2288440

Work remaining to be done in the future includes the implementation of these extensions, including the computational savings, in order to investigate the performance of different cross-over designs. An issue of particular interest is whether

those designs which have been identified as robust under final period dropout are also robust to dropouts occurring earlier in the study.

6.3 Treatment Related Dropout

In Section 1.4, three categories of dropout for repeated measurement studies were given using the definitions of Diggle and Kenwood (1994): completely random dropout, random dropout and informative dropout. Throughout this thesis, it has been assumed that each subject has a fixed probability θ of dropping out in the final period unrelated to the proposed treatment, that is, that the dropout process is completely random.

The other two dropout processes relate to situations in which the reason for a subject dropping out is treatment related. If a study is undertaken and dropouts occur which can be shown to be treatment related, then this finding will be one of the most important outcomes of the investigation. In practice, it is unlikely that a study will be carried out if it is strongly believed a priori that one or more of the active treatments has known side-effects which might lead to subjects dropping out.

In many clinical investigations it is required to investigate the efficacy of a drug through comparison with a placebo treatment, that is an inactive substance, as well as with other active drugs. When a placebo is included in a trial it may be reasonable to assume that the probability of a subject dropping out during or immediately after a period of placebo treatment is greater than during any other period; for example, the probability that a subject drops out during the final period might be θ when the final treatment administered is active and 2θ when the final treatment is placebo. The methodology presented in this thesis can be easily extended to cover situations where this form of treatment related dropout is anticipated. However, the computational reductions described in Chapter 3 will not be as large as in the case when the dropout process is completely random. This is because not all designs which are combinatorially equivalent will have the same probability of being

implemented when the probability of dropping out is treatment related.

Future directions for work in this area includes assessing the performance of different designs for experiments in which a placebo is included, comparing the different designs available and, in particular, establishing whether or not the proposed inclusion of a placebo alters the recommendations concerning design selection.

6.4 Investigations for Alternative Models

The investigations carried out throughout this thesis have assumed that the observations follow the simple carry-over model of equation (1.1). The methodology for examining the robustness of cross-over designs to dropouts described in Chapter 2 is not dependent upon the use of this particular model and could be used in conjunction with any appropriate model. Examples are models which include additional carry-over effects from treatments in earlier periods or interaction terms such as that between the direct treatment and carry-over terms. Alternatively, the assumption of independently and identically distributed random errors could be replaced by some form of correlated error structure.

A review of the various models which could be adopted for the observations from a cross-over study has been given in Section 1.3 and these could be used for future assessment of the performance of different designs.

It is particularly difficult to predict in advance an appropriate form of correlated error structure. In recent years several authors have investigated the robustness of cross-over designs to different forms of correlated error structures. The main purpose of this work is to establish which designs perform well under a variety of different error structures so that a mis-specification of the error structure during the planning stage will not have drastic consequences on the realised experiment. One important area of future work will be to try to find designs which are robust to both correlated error structures and dropouts.

6.5 Derivation of Universally Optimal Designs

Several of the important advances in the design of experiments have been made possible by the work on universal optimality due to Kiefer (1975). An interesting area for future research is to apply the techniques of Kiefer to establish which designs have the maximum mean performance either for direct or carry-over treatment comparisons, as defined in equation (2.4) and (2.6).

The following lemma, which is analogous to Proposition 1 Kiefer (1975), shows that the main problem to be addressed is that of finding a design which maximises the trace of a matrix. The matrix is a weighted sum of the information matrices of all the implemented designs arising from a planned design, either for the estimation of the direct or carry-over effects, A_α or A_λ.

Lemma 6.1 Let $\mathcal{F}$ be the set of all cross-over designs involving t treatments, p periods and s subjects. Let Z be the set of non-negative definite matrices having zero row and column sums such that

$$Z = \{z; z = \sum_{l \in L} A(d_l)P(l|\theta), \quad \text{for some } d \in \mathcal{F}\},$$

where A is A_α or A_λ and $\{d_l; l \in L\}$ is the set of implementable designs arising from a planned design d.

Let Z contain a matrix z^* such that

(a) z^* is completely symmetric, and

(b) $tr(z^*) = max_{z \in Z} tr(z)$.

Then

$$\Phi(z^*) \leq \Phi(z) \quad \text{for all} \quad z \in Z,$$

for Φ as defined in Definition 1.3.

In the presence of final period dropout, very few designs will give rise to a matrix z which is completely symmetric when $\theta \neq 0$. However due to its particular com-

binatoric properties, any complete set of mutually orthogonal Latin squares which possess the additional property of balance will always give rise to completely symmetric z, irrespective of the probability of final period dropout θ. Future work is required to establish the class of cross-over design over which a set of balanced mutually orthogonal Latin squares can be shown to be optimal when final period dropout is anticipated; that is the class of designs over which condition (b) is satisfied. A full investigation of other designs which satisfy the above criteria is also required.

The development of optimality results for cross-over designs in the presence of dropouts is a particular challenging area of investigation in which there is a great deal of potential for future research.

6.6 Conclusions

In this final section the conclusions arising from the research presented in this thesis are outlined.

One of the most important decisions taken during the planning stage of any experiment is selecting an appropriate design. A problem frequently encountered in the context of clinical cross-over studies is that subjects fail to complete their allotted sequence of treatments. Even though this problem is widely acknowledged it is usual to select a design on the basis of performance measures which assume no subjects drop out. The aim of the work undertaken in this thesis has been to address this problem by developing methods for assessing cross-over designs when dropouts may occur, investigate the robustness of a variety of the most frequently employed cross-over designs and make recommendations concerning design selection.

In Chapter 1 the problem of dropouts was presented and the direction of the research outlined. The particular features of cross-over experiments were described and some of the controversies concerning their use discussed. The different models which can be assumed for the observations were reviewed and some of the arguments

against the use of the most commonly used model, namely the simple carry-over model, were considered.

In Chapter 2 a method for assessing cross-over designs in the presence of final period dropout was presented together with criteria for choosing between different competing designs. These were illustrated using examples involving Williams squares of side four.

Assessing the performance of cross-over designs using the criteria developed in Chapter 2 requires a considerable amount of computation. In Chapter 3 ways of reducing this by applying results from combinatorial theory were presented and illustrated.

In Chapter 4 a study of the performance subject to final period dropout of three different four treatment, four period, uniform balanced designs was given. This showed that, although each of the designs investigated performs equally well when the probability of dropout is not considered, each design is not equally robust to the probability of final period dropout. From this study we conclude that a single Williams square of side four should be avoided since its mean performance measures are not as high as can be obtained using alternative designs and there is a non-zero probability that the implemented design will be disconnected. When the maximum number of treatment sequences is eight we recommend that a complementary pair of Williams squares should be used. When the maximum number of treatment sequences available is 12 we recommend using a design based on a complete set of balanced mutually orthogonal Latin squares.

In Chapter 5 the performance subject to final period dropout of three treatment, three period designs was investigated. Unlike Chapter 4, attention was not restricted to uniform balanced designs. A review of the optimality of designs when the restriction of uniformity is removed was presented and these results were used to identify which particular designs to investigate. Attention was drawn to the fact that, if final period dropouts occur, the realised experiment cannot be uniform balanced. Since it is known that when dropouts are not considered uniform balanced designs

are not necessarily optimal outside the class of uniform designs, it is sensible to consider non-uniform designs for use in experiments in which final period dropouts are anticipated. A study of the performance subject to final period dropout of three different designs was undertaken and the results for designs involving up to 36 subjects presented. These showed that the recommended design may change as the probability of final period dropout increases.

Since the designs compared were identical in the first two periods, an investigation was made into the "best" third period to employ using an exhaustive search of all the 729 possible designs. An examination of the mean and variance of the performance measures X_d and Y_d for each of the possible 729 designs showed that, with reference to the estimation of first-order carry-over treatment effects, irrespective of the value of θ, an orthogonal residual effects design is recommended. Conclusions concerning the estimation of direct treatment effects depend upon the anticipated value of θ. Tables indicating the recommended design for ranges of θ values were given, together with a discussion concerning which designs to use if efficient estimates of both the direct and first-order carry-over treatment effects are required.

The research presented in this thesis indicates that when choosing a design for studies in which subjects may drop out the robustness of competing designs to dropouts should be considered. Further work is required to establish designs which are robust to subjects dropping out during any period of a study, to establish designs which are robust to the problem of treatment related dropout and to find designs which are robust to dropouts for studies in which it is appropriate to assume the observations follow models other than the simple carry-over model.

Appendix A

Computer Program

This appendix contains a computer program written in SAS/IML which evaluates the mean and variance of the performance measures X_d and Y_d, under both the A- and MV-criteria, for a cross-over design when the observations are assumed to follow the simple carry-over model. This particular program will assess the robustness to final period dropout of designs based on n replicates of a complementary pair of Williams squares of side four for $\theta = 0.0, 0.1, \ldots, 1.0$.

```
/* Program to obtain summary measures for robustness of     */
/* designs using complementary pairs of Williams squares    */
/* of side four                                             */

options linesize=70;

DATA missing;

/* To input file of dropout identifiers and equivalence      */
/* class sizes      */
INFILE 'dropout.data';
INPUT d1 d2 d3 d4 d5 d6 d7 d8 size;

PROC IML;

START spec(m,n,p,s,Xp,Xs,Xalpha,Xlambda);
/* This subroutine contains the design specifications        */
/* p= number of periods                                      */
/* t= number of treatments                                   */
/* m= number of treatment sequences                          */
/* n= number of subjects per treatment sequence              */
/* s= total number of subjects                               */
/* D= Design in block form                                   */
/* Xp= Design matrix holding period effects                  */
/* Xs= Design matrix holding subject effects                 */
/* Xalpha= Design matrix holding direct treatment effects    */
/* Xlambda= Design matrix holding carry-over effects         */

p=4;  t=4; m=8;  n=2; s=m#n;
```

```
D= J(n,1)@{1 2 4 3}//
   J(n,1)@{2 3 1 4}//
   J(n,1)@{3 4 2 1}//
   J(n,1)@{4 1 3 2}//
   J(n,1)@{1 3 2 4}//
   J(n,1)@{2 1 4 3}//
   J(n,1)@{3 4 1 2}//
   J(n,1)@{4 2 3 1};

Xs= I(s)@J(p,1);
Xp= J(s,1)@I(p);

/* To generate the design matrix Xalpha                    */
DD=SHAPE(D,s#p);
Xalpha=DESIGN(DD);

/* To generate the design matrix Xlambda                   */
ZERO=J(s,1,0);
R=ZERO || D(|,1:(p-1)|);
RR=SHAPE(R,s#p);
RRR=DESIGN(RR);
Xlambda=RRR(|,2:(t+1)|);

FREE ZERO D DD R RR RRR;
FINISH;
```

```
START lossmat(n,p,q1,q2,q3,q4,q5,q6,q7,q8,L);
/*This subroutine generates the loss matrix */

DROP=SHAPE(0,p-1,1);
DR=I(p-1)||DROP;
FREE DROP;

START   submat(qi,Li,n,p,DR);

IF qi=0 THEN DO;
          Li=I(n*p);
  END;

          ELSE DO;

          IF qi=n THEN DO;
              Li=I(n)@DR;
              END;

              ELSE DO;
              Li=BLOCK(I(n-qi)@I(p) , I(qi)@DR);
          END;

END;

FINISH;

RUN submat(q1,L1,n,p,DR);
RUN submat(q2,L2,n,p,DR);
```

```
RUN submat(q3,L3,n,p,DR);
RUN submat(q4,L4,n,p,DR);
RUN submat(q5,L5,n,p,DR);
RUN submat(q6,L6,n,p,DR);
RUN submat(q7,L7,n,p,DR);
RUN submat(q8,L8,n,p,DR);

L=BLOCK(L1,L2,L3,L4,L5,L6,L7,L8);

FINISH;

START assess(m,n,p,s,Q,dvec,L,Xp,Xs,Xalpha,Xlambda,ADT,ACO,MVDT,MVCO);
/* This subroutine calculates the information matrices and the  */
/* and the preformance measures for direct and carry-over       */
/* effects under the A- and MV-criteria for each non-equivalent */
/* implementable design.                                        */
/* Aalp=information matrix for direct treatment effects         */
/* Alam=information matrix for carry-over effects               */
/* C=matrix holding the contrasts of interest                   */

/* To calculate the matrix to adjust for subjects and periods   */

LXp=L*Xp;
LXs=L*Xs;
prLXs=LXs*INV(LXs`*LXs)*LXs`;
W=(I((s*p)-Q)-prLXs);
K=W-W*LXp*GINV(LXp`*W*LXp)*LXp`*W;

/* To calculate the information matrices                        */
```

```
Lxalpha=L*Xalpha;
LXlambda=L*Xlambda;
in1=LXalpha'*K*LXalpha;
in2=LXalpha'*K*LXlambda;
in3=LXlambda'*K*LXlambda;

Aalp=in1-in2*GINV(in3)*in2';
Alam=in3-in2'*GINV(in1)*in2;

C={1 -1 0 0,
   1 0 -1 0,
   1 0 0 -1,
   0 1 -1 0,
   0 1 0 -1,
   0 0 1 -1};

VARDT=C*GINV(Aalp)*C';
VARCO=C*GINV(Alam)*C';

/* To calculate the reciprical average variances    */
ADT=6/TRACE(VARDT);
ACO=6/TRACE(VARCO);

/* To calculate the reciprocal of the maximum variance        */
VDT=VARDT(|1,1|)//VARDT(|2,2|)//VARDT(|3,3|)//VARDT(|4,4|)//
    VARDT(|5,5|)//VARDT(|6,6|);
VCO=VARCO(|1,1|)//VARCO(|2,2|)//VARCO(|3,3|)//VARCO(|4,4|)//
    VARCO(|5,5|)//VARCO(|6,6|);
```

```
MVARDT=MAX(VDT);
MVARCO=MAX(VCO);

MVDT=INV(MVARDT);
MVCO=INV(MVARCO);

FREE  LXp LXs prLXp prLXs LXalpha LXlambda K W in1 in2 in3 C;
FREE  VAECO VARDT VDT VCO MVARDT MVARCO;
FINISH;

START dropout(n,q1,q2,q3,q4,q5,q6,q7,q8,size,theta,PrL);
/* This subroutine calculates the probability of implementation */
/* for each implemntable design */

START facto(z,fact);
/* This subrotine calculates factorials                          */

j=z;              /* j is a dummy variable       */

IF z=1| z=0 THEN DO;
            fact=1;
            END;

            ELSE DO i=1 to (z-1) by 1;
            fact=j*(z-i);
            j=fact;
            END;
END;
FREE j;
```

```
FINISH;

START sequence(qi,n,theta,Pqi);
/* This subroutine calculates the probability of dropout per    */
/* treatment sequeunce                                          */

 nqi=n-qi;
 RUN facto(n,factn);
 RUN facto(qi,factqi);
 RUN facto(nqi,factnqi);
 Pqi=(factn/(factqi#factnqi))#(theta##qi)#((1-theta)##nqi);

FREE factn factqi factnqi nqi ;
FINISH;

RUN sequence(q1,n,theta,Pq1);
RUN sequence(q2,n,theta,Pq2);
RUN sequence(q3,n,theta,Pq3);
RUN sequence(q4,n,theta,Pq4);
RUN sequence(q5,n,theta,Pq5);
RUN sequence(q6,n,theta,Pq6);
RUN sequence(q7,n,theta,Pq7);
RUN sequence(q8,n,theta,Pq8);

PrL=Pq1#Pq2#Pq3#Pq4#Pq5#Pq6#Pq7#Pq8#size;

FREE Pq1 Pq2 Pq3 Pq4 Pq5 Pq6 Pq7 Pq8;
FINISH;
```

```
START main;
RUN spec(m,n,p,s,Xp,Xs,Xalpha,Xlambda);
use missing;
read all into xdrop;
k=NROW(xdrop);
do a= 1 to k by 1;
q1=xdrop(|a,1|);
q2=xdrop(|a,2|);
q3=xdrop(|a,3|);
q4=xdrop(|a,4|);
q5=xdrop(|a,5|);
q6=xdrop(|a,6|);
q7=xdrop(|a,7|);
q8=xdrop(|a,8|);
size=xdrop(|a,9|);
dvec=q1||q2||q3||q4||q5||q6||q7||q8;
Q=q1+q2+q3+q4+q5+q6+q7+q8;  /* total number of dropouts          */

RUN lossmat(n,p,q1,q2,q3,q4,q5,q6,q7,q8,L);
RUN assess(m,n,p,s,Q,dvec,L,Xp,Xs,Xalpha,Xlambda,ADT,ACO,MVDT,MVCO);

/* To create an output dataset for a-opt and mv-opt measures*/;
AOUT=ADT||(ADT##2)||ACO||(ACO##2);
MVOUT=MVDT||(MVDT##2)||MVCO||(MVCO##2);

IF a=1 THEN DO;
        ARES=AOUT;
        MVRES=MVOUT;
        END;
```

```
        ELSE DO;
        ASTATS=ARES//AOUT;
        MVSTATS=MVRES//MVOUT;
        ARES=ASTATS;
        MVRES=MVSTATS;
        END ;

END;

DO theta=0.1 to 0.9 by 0.1;

DO b= 1 to k by 1;

q1=xdrop(|b,1|);
q2=xdrop(|b,2|);
q3=xdrop(|b,3|);
q4=xdrop(|b,4|);
q5=xdrop(|b,5|);
q6=xdrop(|b,6|);
q7=xdrop(|b,7|);
q8=xdrop(|b,8|);
size=xdrop(|b,9|);

RUN dropout(n,q1,q2,q3,q4,q5,q6,q7,q8,size,theta,PrL);

IF b=1  THEN DO;
        prob=PrL;
        END;
```

```
        ELSE DO;
        newp=prob//PrL;
        prob=newp;
        END;
END;

/* TO OBTAIN OUTPUT FOR A-OPT MEASURES */
ARS=prob'*ARES;

MADT=ARS(|1,1|);
VARADT=ARS(|1,2|)-MADT##2;

MACO=ARS(|1,3|);
VARACO=ARS(|1,4|)-MACO##2;

AFINAL=theta||MADT||VARADT||MACO||VARACO;

/* TO OBTAIN OUTPUT FOR MV-OPT MEASURES */
MVRS=prob'*MVRES;

MEMVDT=MVRS(|1,1|);
VARMVDT=MVRS(|1,2|)-MEMVDT##2;

MEMVCO=MVRS(|1,3|);
VARMVCO=MVRS(|1,4|)-MEMVCO##2;

MVFINAL=theta||MEMVDT||VARMVDT||MEMVCO||VARMVCO;
```

```
IF theta=0.1 THEN DO;
        CREATE ARESULTS FROM AFINAL;
        APPEND FROM AFINAL;
        CREATE MVRESULTS FROM MVFINAL;
        APPEND FROM MVFINAL;
        END;

        ELSE DO;
        SETOUT ARESULTS;
        APPEND FROM AFINAL;
        SETOUT MVRESULTS;
        APPEND FROM MVFINAL;
END;

END;

FINISH;

RUN main;
CLOSE ARESULTS;
CLOSE MVRESULTS;

PROC print DATA=ARESULTS;
PROC print DATA=MVRESULTS;
```

References

[1] Abeyasekera, S. and Curnow, R.N., (1984), *"The desirability of adjusting for residual effects in a crossover design."* **Biometrics, 40, pp 1071–1078.**

[2] Balaam, L.N., (1968), *"A two-period design with t^2 experimental units."* **Biometrics, 24, pp 61–73.**

[3] Baker, N., Mews, R.J., Huitson, A. and Poloniecki, J., (1982), *"The two period crossover trial."* **BIAS, 9, pp 67–116.**

[4] Cheng, C.S. and Wu, C.F., (1980), *"Balanced repeated measurements designs."* **The Annals of Statistics, 8, pp 1272–1283.**

[5] Cohen, D.I.A., (1978), **Basic Techniques of Combinatorial Theory.** New York: Wiley.

[6] Cox, D.R., (1984), *"Interaction."* **International Statistics Review, 52, pp 1–31.**

[7] Diggle, P. and Kenward, M.G., (1994), *"Informative dropout in longitudinal data analysis."* **Journal of the Royal Statistical Society, Series C, 43, pp 49–93.**

[8] Fleiss, J.L., (1986), *"On multi-period crossover studies. Letters to the editor."* **Biometrics, 42, pp 449–450.**

[9] Fleiss, J.L., (1989), *"A critique of recent research on the two treatment crossover design."* **Cont. Clinical Trials, 10, pp 237–243.**

[10] Gough, K., (1994), *"Discussion of the paper of Diggle and Kenward."* **Journal of the Royal Statistical Society, Series C, 43, pp 72–93.**

[11] The Group for the Evaluation of Cinromide in the Lennox-Gastaut Syndrome, (1989), *"Double-blind, placebo-controlled evaluation of cinromide in patients with Lennox-Gastaut syndrome."* **Epilepsia, 30, pp 422–429.**

[12] Hedayat, A. and Afsarinejad, K., (1975), *"Repeated measurements designs, I."* In **A Survey of Statistical Design and Linear Models.** Ed. J.N. Sirivastava. Amsterdam: North-Holland.

[13] Hedayat, A. and Afsarinejad, K., (1978), *"Repeated measurements designs II."* **The Annals of Statistics, 6, pp 619–628.**

[14] Herzberg, A.M. and Andrews, D.F., (1976), *"Some considerations in the optimal design of experiments in non-optimal situations."* **Journal of the Royal Statistical Society, Series B, 38, pp 284–289.**

[15] Jones, B. and Donev A.N., (1994), *"Optimal designs for repeated measurements experiments."* **Abstracts of the Royal Statistical Society Conference, Newcastle.**

[16] Jones, B. and Kenward, M.G., (1989), **Design and Analysis of Cross-Over Trials.** London: Chapman and Hall.

[17] Kiefer, J., (1975), *"Construction and optimality of generalised Youden designs."* In **A Survey of Statistical Designs and Linear Models.** Ed. J.N. Srivastava. Amsterdam: North-Holland.

[18] Kunert, J., (1983), *"Optimal design and refinement of the linear model with applications to repeated measurement designs."* **The Annals of Statistics, 11, pp 247–257.**

[19] Kunert, J., (1984), *"Optimality of balanced uniform repeated measurments designs."* **The Annals of Statistics, 12, pp 1006–1017.**

[20] Kunert, J., (1985), *"Optimal repeated measurements designs for correlated observations and analysis by weighted least squares."* **Biometrika, 72, pp 375–389.**

[21] Kunert, J., (1991), *"Crossover designs for two treatments and correlated errors."* **Biometrika, 78, pp 315–324.**

[22] Laserre, V., (1991), *"Determination of optimal designs using linear models in crossover trials."* **Statistics in Medicine, 10, pp 909–924.**

[23] Low, J.L., Lewis S.M., Mckay B.D. and Prescott P., (1994), *"Computational issues for cross-over designs subject to dropout."* **Proceedings in Computational Statistics, pp 423–428.**

[24] **MAPLE V Library Reference Manual (1991).** Eds. B.W. Char, K.O. Geddes, G.H. Gonnet, B.L. Leong, M.B. Monagan and S.M. Watt. Berlin: Springer-Verlag.

[25] Magda, C.G., (1980), *"Circular balanced repeated measurements designs."* **Commun. Statist.- Theory & Meth. A9, pp 1901–1918.**

[26] Matthews, J.N.S., (1987), *"Optimal crossover designs for the comparison of two treatments in the presence of carryover effects and autocorrelated errors."* **Biometrika, 74, pp 311–320.**

[27] Matthews, J.N.S., (1988), *"Recent developements in crossover designs."* **International Statistical Review, 56, pp 117–127.**

[28] Matthews, J.N.S., (1990), *"The analysis of data from crossover designs: the efficiency of ordinary least squares."* **Biometrics, 46, pp 689–696.**

[29] Matthews, J.N.S., (1993), *"Modelling and optimality in the design of crossover studies for medical applications."* **Submitted for publication.**

[30] Patterson, H.D. and Lucas, H.L., (1962), *"Changeover designs."* **North Carolina Agricultural Experiment Station Bulletin, no. 147.**

[31] Searle, S.R., (1971), **Linear Models.** New York: Wiley.

[32] Senn, S.J., (1992), *"Is the simple carry-over model useful?"* **Statistics in Medicine, II, pp 715–726.**

[33] Senn, S.J., (1993), **Cross-over trials in clinical research.** London: Wiley.

[34] Slomson, A., (1991), **Introduction to Combinatorics.** London: Chapman and Hall.

[35] Street, D., (1989), *"Combinatorial problems in repeated measurements designs."* **Discrete Mathematics, 77, pp 323–343.**

[36] Willan, A.R. and Pater, J.L., (1986), *"Carryover and the two-period crossover clinical trial."* **Biometrics, 42, pp 593–599.**

[37] Williams, E.J., (1949), *"Experimental designs balanced for the estimation of residual effects of treatments."* **Australian Journal of Scientific Research, 2, pp 149–168.**