Optimal crossover designs when carryover effects are proportional to direct effects

Mausumi Bose[a], John Stufken[b,*]

[a]*Indian Statistical Institute, Calcutta, India*
[b]*Department of Statistics, University of Georgia, Athens, GA 30602, USA*

Abstract

Crossover designs are used for a variety of different applications. While these designs have a number of attractive features, they also induce a number of special problems and concerns. One of these is the possible presence of carryover effects. Even with the use of washout periods, which are for many applications widely accepted as an indispensable component, the effect of a treatment from a previous period may not be completely eliminated. A model that has recently received renewed attention in the literature is the model in which first-order carryover effects are assumed to be proportional to direct treatment effects. Under this model, assuming that the constant of proportionality is known, we identify optimal and efficient designs for the direct effects for different values of the constant of proportionality. We also consider the implication of these results for the case that the constant of proportionality is not known.

Keywords: Direct treatment effects; Proportional carryover effects; Universal optimality

1. Introduction

Crossover designs are used in experiments that call for each subject in the study, such as a person or animal, to successively receive multiple treatments over a number of time periods. These designs are widely used in several areas, including clinical trials, learning experiments, animal feeding experiments and agricultural field trials. Practical considerations, such as the possible scarcity of subjects and the ability to compare different treatments on the same subject, which is especially important in case of large natural differences between the subjects, are but two of the reasons why a crossover design can be attractive for certain studies. But using these designs may also introduce a number of possible problems and concerns (cf. Stufken, 1996). One of these is the possibility that an observation on a subject is not only affected by the treatment currently administered to this subject, but also by lingering effects of one or more of the treatments that the subject received in earlier periods. Such effects are called carryover (or residual) effects.

During the last years more and more researchers in this area have adopted the view that, if carryover effects can occur, a crossover design should only be used with sufficiently long periods between the treatment periods so that any effects from the previously administered treatments can vanish. Such intervening periods are known as washout periods. On the other hand, it is not always known when a washout period is "sufficiently long" for the carryover effects to vanish completely, and, even if it is, this requirement may lead to washout periods with a duration that is in conflict with practical or ethical considerations. Whatever view one subscribes to on this issue, the use of washout periods is highly recommended for these studies since, if nothing else, it should help to diminish the size of any carryover effects.

Crossover designs have played a role in experiments for a long time, and the notion of carryover effects is already present in the work of Cochran et al. (1941). The study of optimal crossover designs starts with Hedayat and Afsarinejad (1978), who use the same model as Cochran, Autrey and Cannon. The work by Hedayat and Afsarinejad inspired a number of other researchers, including Cheng and Wu (1980), Kunert (1983, 1984) and Matthews (1987). Stufken (1996) provides a review of selected optimality results. More recently, the work by Kushner (1997, 1998) has spurred a renewed interest in optimal crossover designs and represents a breakthrough that elevates our understanding to an entirely new level. Hedayat and Yang (2003, 2004) also provide additional recent advances.

Most of the work cited hitherto focuses on variants of what may be called traditional modeling of carryover effects (as already present in Cochran et al., 1941). Such models allow for a carryover effect from the treatment in the previous period only, and assume that this effect only depends on that treatment. Much work has been done on the optimality of crossover designs under different models, including the recent work by Afsarinejad and Hedayat (2002) and Kunert and Stufken (2002). For a comparison of optimality results under different models for the carryover effects we refer to Hedayat and Stufken (2003).

The model that we will focus on in this paper is conceptually very simple but technically a bit more complicated. The key assumption is that there is only a carryover effect from the treatment in the previous period and that the size of this effect is proportional to the direct effect of that treatment. This is conceptually quite simple—a treatment with a stronger direct effect will also have a stronger lingering effect. It is technically, however, a more complicated model since it is intrinsically nonlinear.

This model has received attention in the literature before. In particular, recently Kempton et al. (2001) study the model for an unknown constant of proportionality. They obtain the information matrix for direct effects using a linear approximation of the model at the true values of the parameters. Then, by using a combination of analytical reasoning and computer search, for a few special cases they obtain designs that are optimal for the direct effects for different values of the constant of proportionality. The optimality criteria that they use is an integrated A-optimality criterion, where the integration is needed because their information matrix does depend on the true value of the direct effects.

We also consider the proportional carryover effects model, but we assume initially that the constant of proportionality, say θ, is known. This is admittedly not a very realistic assumption, but the results are nevertheless insightful. By making this assumption we are able to obtain closed-form, analytical solutions for optimal designs for the direct effects for ranges of values of θ. These results reveal that there are designs that are optimal for a broad range of plausible values for θ. What we ignore by making this assumption is the loss of efficiency resulting from actually having to estimate θ when it is unknown. But we will show that this loss is small in some cases, leading to designs that are optimal or efficient in the more realistic case that θ is unknown.

The optimality criterion that we consider is that of universal optimality, so that the resulting designs are optimal for direct effects according to any of the usual criteria, such as A-, D- or E-optimality. The class of competing designs considered in this paper is the class of all crossover designs. Unlike Kempton et al. (2001), we do not have to restrict ourselves to smaller subclasses of designs.

In Section 2, we state the model and derive the relevant information matrix. Section 3 contains the optimality results for the case that the constant of proportionality, θ, is known, while optimal designs for various useful parameter combinations for different ranges of θ are obtained and presented in Section 4. Section 5 assesses the implications of the results in the previous sections for the important case that θ is unknown.

2. The model and a strategy

Let $\Omega_{t,n,p}$ be the class of all crossover designs with t treatments, p periods and n subjects. For a design $d \in \Omega_{t,n,p}$, a model that has been used extensively is given by

$$y_{ij} = \alpha_i + \beta_j + \tau_{d(i,j)} + \rho_{d(i-1,j)} + \varepsilon_{ij}, \quad i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, n,$$

where $d(i, j)$ stands for the treatment that is assigned to subject j in period i under design d, μ is a general mean, α_i is an effect due to the ith period, β_j is an effect due to the jth unit, $\tau_{d(i,j)}$ is the direct effect due to treatment $d(i, j)$ and

$\rho_{d(i-1,j)}$ is the first-order carryover effect due to treatment $d(i-1, j)$. All the above effects are assumed to be fixed effects and we define $\rho_{d(0,j)} = 0$. The ε_{ij} are independent identically distributed Gaussian random error terms. (The model is often presented with an overall mean μ included, but that is merely a reparametrization of the model given here and makes no difference for our considerations.)

In matrix notation the above model may be written as

$$y = P\alpha + U\beta + T_d\tau + R_d\rho + \varepsilon, \tag{1}$$

where $y = (y_{11}, y_{12}, \ldots, y_{pn})'$, $\alpha = (\alpha_1, \ldots, \alpha_p)'$, $\beta = (\beta_1, \ldots, \beta_n)'$, $\tau = (\tau_1, \ldots, \tau_t)'$, $\rho = (\rho_1, \ldots, \rho_t)'$, $\varepsilon = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{pn})'$; the matrices P and U denote the parts of the design matrix that correspond to period effects and subject effects, respectively; T_d and R_d are the incidence matrices for direct and carryover effects for design d, respectively; and ε is the vector of random errors.

Under the assumption that the carryover effects are proportional to the direct effects, model (1) becomes

$$y = P\alpha + U\beta + T_d\tau + \theta R_d\tau + \varepsilon, \tag{2}$$

where θ is the constant of proportionality. We place no constraint on θ, although it would normally seem plausible for θ to take values in the interval $[-1, 1]$ or even in the interval $[0, 1]$ (see also Kempton et al., 2001).

While model (2) is a nonlinear model, following the discussion in the previous section we will treat θ in Sections 2–4 as a known constant (reducing the model to a linear model) in order to investigate the dependence of the performance of designs for estimating treatment contrasts as a function of θ.

In what follows, we write 1_m and I_m (or occasionally just I) to denote the $m \times 1$ vector of ones and the identity matrix of order m, respectively. For a matrix A, $\omega(A)$ denotes the projection matrix onto the column space of A and $\omega^{\perp}(A)$ denotes $I - \omega(A)$, where I is an identity matrix of appropriate order. $\mathrm{Tr}[A]$ denotes the trace of matrix A.

With the ordering of the elements of y, as shown after model (1), it follows that $P = I_p \otimes 1_n$ and $U = 1_p \otimes I_n$. Writing $X_d(\theta) = T_d + \theta R_d$, it follows that under model (2), for a design d in $\Omega_{t,n,p}$ and a fixed value of θ, the information matrix for τ is given by

$$C_d(\theta) = (X_d(\theta))'\omega^{\perp}(P\ \ U)X_d(\theta). \tag{3}$$

Aside from changes in notation, the difference between the matrix in (3) and the information matrix in Kempton et al. (2001), which is obtained by treating θ as an unknown parameter and using a linear approximation to model (2), is a nonnegative definite matrix that depends on τ and θ. We will return to this in more detail in Section 5.

A design $d \in \Omega_{t,n,p}$ is said to be *uniform over the periods* if each of the treatments occurs equally often (i.e., n/t times) in each period in d. It is said to be *invariant over the periods* if for every treatment, the treatment is replicated equally often in each period (but different treatments need not have the same replication). Clearly, a design that is uniform over the periods is also invariant over the periods. Based on the results in Kiefer (1975), for a fixed value of θ, we will seek designs in $\Omega_{t,n,p}$ that maximize $\mathrm{Tr}[C_d(\theta)]$. Instead of studying $C_d(\theta)$ directly, we will find a judiciously selected slightly simpler matrix that majorizes $C_d(\theta)$ (in the Loewner ordering). This will eventually provide an upper bound for $\mathrm{Tr}[C_d(\theta)]$ that applies to all designs $d \in \Omega_{t,n,p}$ but that does not depend on d. If a design attains this upper bound and if the information matrix for that design is completely symmetric, then, following Kiefer (1975), we can conclude that this design is universally optimal for direct effects for the selected value of θ.

An essential step in this direction is formulated in Lemma 1. The result of Lemma 1 will be analyzed further in Section 3, where the importance of the concept of uniformity over the periods will also be elucidated.

Lemma 1. *For a design* $d \in \Omega_{t,n,p}$,

$$C_d(\theta) \leqslant (X_d(\theta))'\omega^{\perp}(R_d 1_t\ \ U)X_d(\theta) \tag{4}$$

in the sense of the Loewner ordering, with equality for all θ *if and only if* d *is invariant over the periods.*

Proof. By observing that $R_d 1_t$ belongs to the column space of P, it follows that

$$(X_d(\theta))'\omega^{\perp}(P\ \ U)X_d(\theta) = (X_d(\theta))'\omega^{\perp}(R_d 1_t\ \ U)X_d(\theta) - (X_d(\theta))'\omega(\omega^{\perp}(R_d 1_t\ \ U)\omega^{\perp}(R_d 1_t)P)X_d(\theta). \tag{5}$$

The second term on the right-hand side of (5) is always $\geqslant 0$, which shows the validity of (4). This term is equal to 0 if and only if

$$(X_d(\theta))'\omega^{\perp}(R_d 1_t\ U)\omega^{\perp}(R_d 1_t)P = 0. \tag{6}$$

Straightforward computations show that

$$\omega^{\perp}(R_d 1_t\ U)\omega^{\perp}(R_d 1_t)P = \begin{bmatrix} 0_n & 0_n \ \cdots \ 0_n \\ 0_{n(p-1)} & \omega^{\perp}(1_{p-1}) \otimes 1_n \end{bmatrix}.$$

Hence, (6) holds for all θ, if and only if

$$T_d' \begin{bmatrix} 0_n & 0_n \ \cdots \ 0_n \\ 0_{n(p-1)} & \omega^{\perp}(1_{p-1}) \otimes 1_n \end{bmatrix} = R_d' \begin{bmatrix} 0_n & 0_n \ \cdots \ 0_n \\ 0_{n(p-1)} & \omega^{\perp}(1_{p-1}) \otimes 1_n \end{bmatrix} = 0_{t\times p}. \tag{7}$$

Writing T_{di} for the $n \times t$ submatrix of T_d corresponding to observations in period i, $i = 1, \ldots, p$, it follows easily that (7) holds if and only if

$$T_{d1}' 1_n = T_{d2}' 1_n = \cdots = T_{dp}' 1_n,$$

which is equivalent to d being invariant over the periods. □

3. Optimality result

From the sufficient conditions in Kiefer (1975), given θ, a design $d^* \in \Omega_{t,n,p}$ will be universally optimal in $\Omega_{t,n,p}$ for direct effects if $C_{d^*}(\theta)$ is completely symmetric and maximizes $\mathrm{Tr}[C_d(\theta)]$ over all designs $d \in \Omega_{t,n,p}$. So, to search for universally optimal designs, we consider the upper bound of $C_d(\theta)$ as given in Lemma 1 and try to maximize the trace of this bound, keeping in mind that only a design that is invariant over the periods can attain equality in (4) for all values of θ.

For simplicity of notation we will use A_t to denote the matrix $\omega^{\perp}(1_t)$. We also define

$$C_{d11} = T_d'\omega^{\perp}(U)T_d, \quad C_{d12} = T_d'\omega^{\perp}(U)R_d, \quad C_{d22} = R_d'\omega^{\perp}(U)R_d. \tag{8}$$

The following lemma builds on the result of Lemma 1.

Lemma 2. *For any θ and any design $d \in \Omega_{t,n,p}$,*

$$\mathrm{Tr}[C_d(\theta)] \leqslant \mathrm{Tr}[C_{d11}] + 2\theta\,\mathrm{Tr}[C_{d12}] + \theta^2\,\mathrm{Tr}[A_t C_{d22}], \tag{9}$$

and equality holds for every θ if d is uniform over the periods.

Proof. Since $C_d(\theta)1_t = 0_t$, we have that $\mathrm{Tr}[C_d(\theta)] = \mathrm{Tr}[A_t C_d(\theta)]$. Using Lemma 1, this means that

$$\mathrm{Tr}[C_d(\theta)] = \mathrm{Tr}[A_t C_d(\theta)] \leqslant \mathrm{Tr}[A_t (X_d(\theta))'\omega^{\perp}(R_d 1_t\ U)X_d(\theta)]. \tag{10}$$

To evaluate this expression we first observe that

$$(X_d(\theta))'\omega^{\perp}(R_d 1_t\ U)X_d(\theta) = (X_d(\theta))'\omega^{\perp}(U)X_d(\theta) - (X_d(\theta))'\omega(\omega^{\perp}(U)R_d 1_t)X_d(\theta). \tag{11}$$

Using the notation in (8), the first term on the right-hand side of (11) can be written as

$$(X_d(\theta))'\omega^{\perp}(U)X_d(\theta) = C_{d11} + \theta(C_{d12} + C_{d12}') + \theta^2 C_{d22},$$

so that

$$\mathrm{Tr}[A_t (X_d(\theta))'\omega^{\perp}(U)X_d(\theta)] = \mathrm{Tr}[C_{d11}] + 2\theta\,\mathrm{Tr}[C_{d12}] + \theta^2\,\mathrm{Tr}[A_t C_{d22}]. \tag{12}$$

Since $\mathrm{Tr}[A_t (X_d(\theta))'\omega(\omega^{\perp}(U)R_d 1_t)X_d(\theta)]$ is clearly nonnegative, combining (10)–(12) yields the desired result in (9).

Equality in (9) holds if $\mathrm{Tr}[A_t(X_d(\theta))'\omega(\omega^{\perp}(U)R_d 1_t)X_d(\theta)]=0$. This is the case if and only if $A_t(X_d(\theta))'\omega^{\perp}(U)R_d 1_t=0$, i.e., if and only if $(X_d(\theta))'\omega^{\perp}(U)R_d 1_t$ is a multiple of 1_t. Since $X_d(\theta)=T_d+\theta R_d$, if this latter condition is to hold for every θ, then

$$T_d'\omega^{\perp}(U)R_d 1_t \quad \text{and} \quad R_d'\omega^{\perp}(U)R_d 1_t$$

must each be a multiple of 1_t. This holds for a design that is uniform over the periods. Since such a design also attains equality in (10), the claim in Lemma 2 follows. □

It can be shown that the only designs that attain equality in both Lemmas 1 and 2 are designs that are uniform over the periods.

The bound in Lemma 2 is very important. It enables us to proceed along the lines of Kushner (1997, 1998) and Kunert and Stufken (2002) in order to find an attainable upper bound for $\mathrm{Tr}[C_d(\theta)]$ that is independent of d. In order to pursue this we need additional notation.

For a design $d \in \Omega_{t,n,p}$, let s_j denote the sequence of treatments for subject j in design d, $j=1,\ldots,n$. For $1 \leqslant u \leqslant v \leqslant 2$, and any sequence s, we write C_{uv}^{s} to denote the matrix C_{d_0uv} as defined in (8) with d_0 denoting the design in $\Omega_{t,1,p}$ consisting of only the single sequence s. It is easy to verify that

$$C_{duv}=\sum_{j=1}^{n} C_{uv}^{s_j} \quad \text{for all } u, v,$$

and hence that

$$\mathrm{Tr}[C_{d11}]=\sum_{j=1}^{n}\mathrm{Tr}[C_{11}^{s_j}], \quad \mathrm{Tr}[C_{d12}]=\sum_{j=1}^{n}\mathrm{Tr}[C_{12}^{s_j}], \quad \mathrm{Tr}[A_tC_{d22}]=\sum_{j=1}^{n}\mathrm{Tr}[A_tC_{22}^{s_j}]. \tag{13}$$

This implies that, for a given θ, to maximize the right-hand side of (9), the design must only use sequences s that maximize

$$\mathrm{Tr}[C_{11}^{s}]+2\theta\,\mathrm{Tr}[C_{12}^{s}]+\theta^2\,\mathrm{Tr}[A_tC_{22}^{s}]. \tag{14}$$

We will denote the quantity in (14) by $B_s(\theta)$, and refer to a sequence as *admissible* (for that value of θ) if it maximizes $B_s(\theta)$ over all possible sequences s for given t and p. Observe that if a sequence can be obtained from a second sequence by a permutation of the treatments, then the two sequences yield the same value for $B_s(\theta)$. We will call any two sequences that can be obtained from each other through such a permutation *equivalent*. In order to find the maximum of $B_s(\theta)$ over s we need to compute $B_s(\theta)$ for only one sequence s from each equivalence class. We will refer to an equivalence class as admissible if the sequences in that class are admissible.

A final concept that we need is that of neighbor balance. A design d in $\Omega_{t,n,p}$ is said to be *neighbor balanced* if, for two selected treatments, the number of times that d assigns the two treatments to the same subject in consecutive periods, irrespective of order, is the same for any two treatments. Note that this definition does not formulate any requirements on how often the same treatment should be assigned to a subject in consecutive periods.

We are now ready to formulate the main result.

Theorem 1. *Consider model* (2) *for a given value of* θ. *Let* d^* *be a design in* $\Omega_{t,n,p}$, *such that*

(i) d^* *is uniform over the periods*,
(ii) *all sequences in* d^* *are admissible* (*for this value of* θ), *and*
(iii) *all of the matrices* C_{d^*11}, $C_{d^*12}+C_{d^*12}'$ *and* C_{d^*22} *are completely symmetric.*

Then, for the given value of θ, d^* *is universally optimal in* $\Omega_{t,n,p}$ *for direct treatment effects.*

Proof. For any design $d \in \Omega_{t,n,p}$ we have that

$$\begin{aligned}\mathrm{Tr}[C_{d^*}(\theta)] &= \mathrm{Tr}[C_{d^*11} + 2\theta C_{d^*12} + \theta^2 A_t C_{d^*22}] = n \max_s B_s(\theta) \\ &\geqslant \mathrm{Tr}[C_{d11} + 2\theta C_{d12} + \theta^2 A_t C_{d22}] \geqslant \mathrm{Tr}[C_d(\theta)]. \end{aligned} \tag{15}$$

The first equality in (15) follows from Lemma 2 and property (i) in Theorem 1; the second equality is a consequence of the additivity property in (13) and property (ii) in Theorem 1. The first inequality in (15) follows again from the additivity property, while the second inequality is simply from Lemma 2. Thus, the information matrix for d^* has maximal trace over all designs in $\Omega_{t,n,p}$.

The result of Theorem 1 follows now from Kiefer (1975) if we can conclude that $C_{d^*}(\theta)$ is a completely symmetric matrix. But

$$\begin{aligned}C_{d^*}(\theta) &= C_{d^*11} + \theta(C_{d^*12} + C'_{d^*12}) + \theta^2 C_{d^*22} - (X_{d^*}(\theta))'\omega(\omega^{\perp}(U)R_{d^*}1_t)X_{d^*}(\theta) \\ &= C_{d^*11} + \theta(C_{d^*12} + C'_{d^*12}) + \theta^2 C_{d^*22} - \delta 1_t 1_t'\end{aligned} \tag{16}$$

for some constant δ. For the first equality in (16) we have used Lemma 1 and the expression in (11), while the second equality is a consequence of property (i) in Theorem 1. Property (iii) implies now that $C_{d^*}(\theta)$ is completely symmetric. □

Note that properties (i)–(iii) in Theorem 1 are sufficient for the conclusion but not necessary. For example, if $\theta = 0$ all that is needed in (iii) is that C_{d^*11} is a completely symmetric matrix.

In applying Theorem 1, property (i) is of course the easiest to verify. In the next section we will derive the admissible sequences for selected values of t and p, concentrating on $2 \leqslant p \leqslant 4$. Property (iii) is a lot easier to verify than it might seem at first. It is always satisfied for $t = 2$. For $t \geqslant 3$ it is of great help to observe that both C_{d11} and C_{d22} can be interpreted as information matrices for block designs, with subjects as blocks, under the usual additive model with block effects and treatment effects. Finally, for a design that satisfies all other properties, the matrix $C_{d12} + C'_{d12}$ is completely symmetric if and only if the design is neighbor balanced.

4. Admissible classes and optimal designs for various combinations of the number of treatments and periods

In this section we will identify optimal designs for various combinations of t and p by using the result in Theorem 1. We pay special attention to cases with $2 \leqslant p \leqslant 4$, which tend to be used more often in practice. In view of Theorem 1, the initial focus will be on identifying all admissible equivalence classes. Using sequences from these classes only, we will then obtain designs that possess properties (i) and (iii) in Theorem 1.

4.1. The case $p = 2,\ t \geqslant 2$

There are only two equivalence classes in this case. Representative sequences from these classes and the corresponding values of $B_s(\theta)$ are shown in Table 1.

Since $0.25(\theta - 2)^2$ is nonnegative for all θ, it is clear that equivalence class (2) is admissible for all θ. When $\theta = 2$, both classes are admissible. However, the fact that $\max_s B_s(\theta) = 0$ when $\theta = 2$ means that $\tau_1 - \tau_2$ is not estimable in that case.

For any other value of θ we should only use sequences that belong to class (2). If $t = 2$, a design that uses the sequences 12 and 21 equally often is universally optimal for direct effects according to Theorem 1. For odd $t \geqslant 3$

Table 1
The equivalence classes for $p = 2,\ t \geqslant 2$

Class	Sequence s	$\mathrm{Tr}(C_{11}^s)$	$\mathrm{Tr}(C_{12}^s)$	$\mathrm{Tr}(A_t C_{22}^s)$	$B_s(\theta)$
(1)	11	0	0	0	0
(2)	12	1	−0.5	0.25	$0.25(\theta-2)^2$

we can always construct a neighbor balanced design that possesses the properties in Theorem 1 provided that n is a multiple of $t(t-1)/2$. For even t, the same is true provided that n is a multiple of $t(t-1)$.

Example 1. For $t=3$ and $n=3\lambda$ for a positive integer λ, let d^* be the design that uses each of the sequences 12, 23 and 31 for λ of the subjects. Then d^* is universally optimal for direct effects in $\Omega_{3,3\lambda,2}$ for any θ. For $t=4$ and $n=12\lambda$, the design that uses each of the sequences 12, 21, 13, 31, 14, 41, 23, 32, 24, 42, 34 and 43 for λ of the subjects is universally optimal for direct effects in $\Omega_{4,12\lambda,2}$.

4.2. The case $p=3,\ t=2$

Table 2 presents the four equivalence classes for this case.

The entries in the column for $B_s(\theta)$ in Table 2 show that class (4) is the only admissible class if $\theta>0$ and that only class (3) is admissible if $\theta<0$. If $\theta=0$, classes (2)–(4) are all admissible. The latter is not surprising since it corresponds to the well-known case of the common model for a row–column design (without carryover effects), and it is known that sequences from all three classes can then occur in an optimal design.

The construction of universally optimal designs is particularly simple for this case, and is illustrated in Example 2.

Example 2. Let d^* be the design with each of the sequences 122 and 211 assigned to $\lambda\geqslant 1$ of the 2λ subjects. Then, for any $\theta\geqslant 0$, we conclude from Theorem 1 that d^* is universally optimal in $\Omega_{2,2\lambda,3}$. Similarly, a design that applies only the sequences 121 and 212, using both equally often, is universally optimal in $\Omega_{2,2\lambda,3}$ if $\theta\leqslant 0$.

Note that design d^* in this example is the strongly balanced design that is also known to be universally optimal for model (1).

4.3. The case $p=3,\ t\geqslant 3$

There are five equivalence classes for this combination of the number of treatments and periods. Only class (5) in Table 3 is new compared to the previous case.

From the discussion in Section 4.2, it is clear that when identifying admissible sequences we only need to compare class (5) to classes (4) (if $\theta>0$) and (3) (if $\theta<0$).

After simple algebra, it follows that (4) is the only admissible class when $0.52<\theta<11.48$, (3) is the only admissible class when $-4.73<\theta<-1.27$, while (5) is admissible for other values of θ. In particular, if $\theta=0$ (the case in which

Table 2
The equivalence classes for $p=3,\ t=2$

Class	Sequence s	$\mathrm{Tr}(C_{11}^s)$	$\mathrm{Tr}(C_{12}^s)$	$\mathrm{Tr}(A_t C_{22}^s)$	$B_s(\theta)$
(1)	111	0	0	0	0
(2)	112	1.33	−0.33	0.33	$1.33-0.66\theta+0.33\theta^2$
(3)	121	1.33	−1	1	$1.33-2\theta+\theta^2$
(4)	122	1.33	0	1	$1.33+\theta^2$

Table 3
The equivalence classes for $p=3,\ t\geqslant 3$

Class	Sequence s	$\mathrm{Tr}(C_{11}^s)$	$\mathrm{Tr}(C_{12}^s)$	$\mathrm{Tr}(A_t C_{22}^s)$	$B_s(\theta)$
(1)	111	0	0	0	0
(2)	112	1.33	−0.33	0.33	$1.33-0.66\theta+0.33\theta^2$
(3)	121	1.33	−1	1	$1.33-2\theta+\theta^2$
(4)	122	1.33	0	1	$1.33+\theta^2$
(5)	123	2	−0.67	1.11	$2-1.34\theta+1.11\theta^2$

Table 4
The equivalence classes for $p=4,\ t=2$

Class	Sequence s	$\mathrm{Tr}(C_{11}^s)$	$\mathrm{Tr}(C_{12}^s)$	$\mathrm{Tr}(A_t C_{22}^s)$	$B_s(\theta)$
(1)	1111	0	0	0	0
(2)	1112	1.5	−0.25	0.375	$1.5-0.50\theta+0.375\theta^2$
(3)	1121	1.5	−0.75	1.375	$1.5-1.5\theta+1.375\theta^2$
(4)	1211	1.5	−0.75	1.375	$1.5-1.5\theta+1.375\theta^2$
(5)	1122	2	0.5	1.375	$2+\theta+1.375\theta^2$
(6)	1212	2	−1.5	1.375	$2-3\theta+1.375\theta^2$
(7)	1221	2	−0.5	1.375	$2-\theta+1.375\theta^2$
(8)	1222	1.5	0.25	1.375	$1.5+0.5\theta+1.375\theta^2$

carryover effects are assumed to be negligible), (5) is the only admissible class. This is again as expected based on existing results for row–column designs.

Optimal designs are presented in the following examples.

Example 3. If $0.52<\theta<11.48$, an optimal design can be obtained by adding a period to the optimal designs in Example 1 in which the same treatment is used as in the second period. Thus, if $t=3$, the design d^* that uses each of the sequences 122, 233 and 311 for λ subjects is universally optimal for such θ in $\Omega_{3,3\lambda,3}$. Similarly, if $-4.73<\theta<-1.27$, then we can again obtain optimal designs by starting with an optimal design in Example 1. In this case, however, the third period should be constructed by repeating the first period. Thus, if $t=3$, using each of 121, 232 and 313 for λ subjects results for such values of θ in a universally optimal design in $\Omega_{3,3\lambda,3}$.

Example 4. The more interesting case is that where θ is near zero. If $-1.27<\theta<0.52$, an optimal design can be obtained by using each sequence that is equivalent to 123 equally often. Since there are $t(t-1)(t-2)$ sequences in this equivalence class, this would quickly require the use of a huge number of sequences. For $t=3$, instead of using all six sequences, a design that uses each of 123, 231 and 312 equally often is by Theorem 1 also optimal. For $t=4$, instead of using all 24 sequences equally often, an optimal design that uses 12 distinct sequences can be constructed. The 12 sequences 123, 312, 231, 421, 142, 214, 134, 413, 341, 432, 243 and 324 of the totally balanced design in Kunert and Stufken (2002) can be used for this purpose.

4.4. The case $p=4,\ t=2$

For this case there are eight equivalence classes, which are all shown in Table 4.

It is immediately clear from Table 4 by looking at the column for $B_s(\theta)$ that the only admissible class for $\theta>0$ is class (5). Similarly, class (6) is the only admissible class for $\theta<0$. If $\theta=0$, classes (5)–(7) are all admissible—which, once more, does not come as a surprise based on known results for row–column designs. Identification of optimal designs is again extremely simple, as shown in Example 5.

Example 5. Let d_1^* be the design that uses sequences 1122 and 2211 equally often, say λ times. Then, by Theorem 1, d_1^* is universally optimal for direct effects in $\Omega_{2,2\lambda,4}$ if $\theta\geqslant 0$. Similarly, design d_2^* that uses each of the sequences 1212 and 2121 λ times is universally optimal for direct effects in $\Omega_{2,2\lambda,4}$ if $\theta\leqslant 0$.

4.5. The case $p=4,\ t=3$

For this case there are 14 equivalence classes. In addition to the eight classes in Table 4, which also apply now, we must also consider the six classes in Table 5.

For $\theta>0$, it is clear that class (14) has a larger $B_s(\theta)$ value than any other class in Table 5. Simple algebra shows that this value also dominates the $B_s(\theta)$ value for class (5) in Table 4, which was best in that table. Hence, if $\theta>0$, then class (14) is the only admissible class. Similarly, class (10) is the only admissible class if $-1<\theta<0$, while (11) and (13) are the only admissible classes if $\theta<-1$. For $\theta=0$, all of the classes (9)–(14) are admissible.

Table 5
Additional equivalence classes for $p=4,\ t=3$

Class	Sequence s	$\mathrm{Tr}(C_{11}^s)$	$\mathrm{Tr}(C_{12}^s)$	$\mathrm{Tr}(A_t C_{22}^s)$	$B_s(\theta)$
(9)	1123	2.5	-0.25	1.5	$2.5-0.5\theta+1.5\theta^2$
(10)	1213	2.5	-1.25	1.5	$2.5-2.5\theta+1.5\theta^2$
(11)	1231	2.5	-1.0	2	$2.5-2\theta+2\theta^2$
(12)	1223	2.5	-0.25	1.5	$2.5-0.5\theta+1.5\theta^2$
(13)	1232	2.5	-1	2	$2.5-2\theta+2\theta^2$
(14)	1233	2.5	0	2	$2.5+2\theta^2$

Table 6
An additional equivalence class for $p=4,\ t\geqslant 4$

Class	Sequence s	$\mathrm{Tr}(C_{11}^s)$	$\mathrm{Tr}(C_{12}^s)$	$\mathrm{Tr}(A_t C_{22}^s)$	$B_s(\theta)$
(15)	1234	3	-0.75	2.0625	$3-1.5\theta+2.0625\theta^2$

Example 6. Let d_1^* and d_2^* be the designs that use sequences 1233, 2311 and 3122 and sequences 1213, 2321 and 3132 equally often, respectively. Then d_1^* is optimal for direct effects if $\theta>0$, while d_2^* is optimal if $-1<\theta<0$. If $\theta<-1$, then an optimal design may be constructed by using each of the sequences 1231, 2312 and 3123 equally often.

4.6. The case $p=4,\ t\geqslant 4$

For this case there are 15 equivalence classes, including the 14 classes in Tables 4 and 5. The only remaining case is shown in Table 6.

A comparison of the $B_s(\theta)$ values shows that class (14) is the only admissible class if $0.34<\theta<23.66$, while class (15) is admissible for other $\theta>0$. When $\theta<0$, both (11) and (13) are admissible if $-6.83<\theta<-1.17$ while (15) is admissible for other $\theta<0$. In particular, class (15) is the only admissible class if $-1.17<\theta<0.34$.

Example 7. Whenever class (15) is admissible, the design that uses each sequence from that equivalence class equally often is optimal for direct effects by Theorem 1. But this class contains $t(t-1)(t-2)(t-3)$ different sequences. Optimal designs can typically be constructed using fewer sequences. For example, if $t=4$, the design d^* that uses the sequences 1234, 2413, 3142 and 4321 equally often is optimal for direct effects if θ takes a value where class (15) is admissible.

Example 8. Starting with an optimal 3-period design with $t\geqslant 4$ as in Example 4, if we add a fourth period by repeating the treatment in the third period in period 4, then we obtain an optimal 4-period design for $0.34<\theta<23.66$.

It is interesting to compare the results in this subsection with those in Kempton et al. (2001). Using their approach, which we briefly described in Sections 1 and 2, and a computer search they looked for the best 4-period designs for four treatments and 12 subjects. They found two 4-period designs to be dominant. One of these, design 1, uses 12 sequences that are equivalent to 1234, while the other, design 2, uses 12 sequences that are equivalent to 1233. Since both of their designs satisfy properties (i) and (iii) of our Theorem 1, we would conclude that, when restricting θ to the interval $[-1,1]$, design 1 is optimal for $-1<\theta<0.34$ and design 2 is optimal for $0.34<\theta<1$. Their conclusion agrees largely with ours, except that they found the crossing point to be near $\theta=0.42$. Moreover, near the crossing point (for $0.37<\theta<0.48$) they found two designs that combined sequences from classes (14) and (15) to perform slightly better than either of design 1 or 2. We return to their approach in Section 5.

4.7. The case $p = 6,\ t = 3$

We provide a brief discussion of this special case where a rather special design, called a strongly balanced uniform design, exists. An example of such a design in $\Omega_{3,9,6}$ consists of the sequences 112233, 223311, 331122, 122331, 233112, 311223, 132132, 213213 and 321321 is known to be universally optimal under model (1) (cf. Cheng and Wu, 1980). It contains sequences from three equivalence classes, but it can be shown that, if $\theta > 0$, only the class consisting of the six sequences that are equivalent to 112233 is admissible. The design that uses each of the sequences 112233, 223311 and 331122 thrice is universally optimal in $\Omega_{3,9,6}$ if $\theta > 0$. Thus, strongly balanced uniform designs do not necessarily remain optimal under the proportional carryover effects model.

5. But θ is not known!

In applications we do not know θ and would have to estimate it. This results in a loss of information for the estimation of treatment contrasts, a loss that can be more severe with some designs than with other designs. The question is therefore justified whether the results in the previous sections, which are obtained under the assumption that θ is known, provide any guidance at all for the important case that θ is not known.

The short answer is that the results provide useful guidance. In some special cases they suggest designs that are also optimal for the direct effects when θ is unknown. But in other cases they need to be used with caution. This section provides a more thorough, rigorously supported answer to this question.

If θ is treated as an unknown parameter then model (2) is a nonlinear model. Following Kempton et al. (2001), using a Taylor series expansion we approximate this model by a linear model in the neigborhood of the true parameter values θ_0 and τ_0. We then obtain the information matrix for the direct effects τ under this linear approximation. Using our notation it can be shown that the information matrix obtained in this way (see Kempton et al., 2001) can be written as

$$C_d(\theta_0) = (X_d(\theta_0))'\omega^{\perp}(P\ \ U\ \ R_d\tau_0)X_d(\theta_0). \tag{17}$$

We will write $C_d^N(\theta_0)$ for this information matrix to distinguish it from the matrix in (3), which we will denote in this section by $C_d^L(\theta_0)$.

It is immediately clear from the expressions in (3) and (17) that $C_d^N(\theta_0) \leqslant C_d^L(\theta_0)$ for any design d, no matter what the true value τ_0 is. It is also clear from these expressions and the observation in the first line of the proof of Lemma 1 that the two matrices are identical if all treatment effects are the same. Note also that the difference between the two information matrices is not a function of the size of τ_0, but only of its direction.

For a design that is uniform over the periods it follows, analogously as in the proof of Lemma 1 (we omit the details), that

$$C_d^N(\theta_0) = (X_d(\theta_0))'\omega^{\perp}(U\ \ R_d 1_t\ \ R_d\tau_0)X_d(\theta_0). \tag{18}$$

Consequently, for such a design we have that

$$\begin{aligned} C_d^L(\theta_0) - C_d^N(\theta_0) &= (X_d(\theta_0))'(\omega^{\perp}(U\ \ R_d 1_t) - \omega^{\perp}(U\ \ R_d 1_t\ \ R_d\tau_0))X_d(\theta_0) \\ &= (X_d(\theta_0))'(\omega(U\ \ R_d 1_t\ \ R_d\tau_0) - \omega(U\ \ R_d 1_t))X_d(\theta_0) \\ &= (X_d(\theta_0))'(\omega(\omega^{\perp}(U\ \ R_d 1_t)R_d\tau_0))X_d(\theta_0). \end{aligned} \tag{19}$$

If, for some θ_0 and τ_0, this difference is 0 and d is a design that is universally optimal for direct effects under Theorem 1 for $\theta = \theta_0$, then d is also universally optimal for direct effects if we treat θ as unknown and θ_0 and τ_0 are the true values. It would be especially useful to know if there are designs d for which the difference in (19) is 0 for some θ_0 irrespective of the value of τ_0, where we assume without loss of generality that $\tau_0' 1_t = 0$. We will identify such designs.

One of the properties required in Theorem 1 is that $C_{d12} + C_{d12}'$ is completely symmetric. If θ is unknown it turns out that we need to strengthen this to the requirement that C_{d12} itself is completely symmetric. Designs with this stronger property tend to be better for the latter case. It is not difficult to obtain designs with this stronger property that satisfy all the other conditions in Theorem 1 from the designs that were constructed in Section 4. To illustrate this point, in Example 4 we considered the neighbor balanced design consisting of sequences 123, 231 and 312. If we add the three sequences 213, 132 and 321, which are obtained by permuting 1 and 2 in the first three sequences, then the resulting design that uses all six sequences equally often has a completely symmetric matrix C_{d12} and satisfies all other conditions of Theorem 1 for all values of θ for which the original design satisfied these conditions. (The resulting design is the well-known uniform design that is balanced for carryover effects.) This idea of adding sequences to a given design that are obtained by applying certain treatment permutations to the sequences of the given design can be used in general to obtain a new design that has a completely symmetric matrix C_{d12}.

We are now ready to state the main result of this section as a theorem and will formulate an interesting special case in the corollary that succeeds it. These results identify designs for which the difference in (19) is 0 for any τ_0 and are an immediate consequence of the previous discussion.

Theorem 2. *Consider the nonlinear model* (2). *Let d^* be a design in $\Omega_{t,n,p}$, and suppose that the true* (*unknown*) *value of θ is $\theta_0 = -\mathrm{Tr}[C_{d^*12}]/\mathrm{Tr}[A_t C_{d^*22}]$. If*

(i) *d^* is uniform over the periods,*
(ii) *all sequences in d^* are admissible for θ_0, and*
(iii) *all of the matrices C_{d^*11}, C_{d^*12} and C_{d^*22} are completely symmetric,*

then d^* is universally optimal in $\Omega_{t,n,p}$ for direct treatment effects irrespective of the true value of τ.

Corollary 1. *If $d^* \in \Omega_{t,n,p}$ satisfies the conditions in Theorem* 1 *for known $\theta = 0$ and $C_{d^*12} = 0$, then d^* is also universally optimal for direct effects under model* (2) *when θ is not known and its true value is $\theta_0 = 0$, no matter what the true value of τ is.*

The condition $C_{d^*12} = 0$ is satisfied by a number of well-known strongly balanced designs (see also Cheng and Wu, 1980), and we will encounter some of these in the next examples.

This special case of $\theta_0 = 0$ is a very important one if we believe that there are only very small or no carryover effects—which, as mentioned in the Introduction, is a desirable goal when using crossover designs. Corollary 1 provides us with a tool to identify designs that are universally optimal for direct effects, irrespective of the true value of τ, when the true (unknown) value of θ is 0. We highly recommend these designs if the carryover effects are expected to be small relative to the direct treatment effects.

Example 9. Let $p = 3$ and $t = 2$. Based on Theorem 1, the design d^* that uses the sequences 112 and 221, each for half of the subjects, is optimal if it is known that $\theta = 0$ (see Section 4.2). Since this design satisfies $C_{d^*12} = 0$, it is also optimal, no matter what the true value of τ is, for model (2) when θ is unknown and its true value is 0.

Example 10. Consider the case $p = 4$ and $t = 2$. From Section 4.4 and Corollary 1, it follows that the design d^* that uses each of the sequences 1122, 2211, 1221, 2112 for a quarter of the subjects is optimal for model (2) when the unknown true value of θ is 0.

Example 11. For $p = 4$, $t = 3$, it follows from Section 4.5 and Corollary 1 that the design d^* that uses each of the sequences 1233, 2311, 3122, 2133, 3211, and 1322 equally often is universally optimal for direct effects when the unknown true value of θ in model (2) is 0, irrespective of the unknown true value of τ.

For other values of θ_0 than that in Theorem 2, an optimal design may depend on the true value τ_0 of τ, making the problem much more difficult. The discussion about the results by Kempton, Ferris and David in Section 4.6 does, however, suggest that the results in Section 4 provide in that situation also useful guidance for the case that θ is unknown.

Acknowledgment

We thank an anonymous referee for raising questions that resulted in an improved version of Section 5.

Note added in proof

Recently Bailey and Kunert (Biometrika 2006, 93, 613–625) were able to extend the work Kempton, Ferris and David (2001) and obtain general optimality results without assuming θ to be known. In particular, they found that the totally balanced designs of Kunert and Stufken, 2002 often perform very well.

References

Afsarinejad, K., Hedayat, A.S., 2002. Repeated measurement designs for a model with self and mixed carryover effects. J. Statist. Plann. Inference 106, 449–459.

Cheng, C.S., Wu, C.F.J., 1980. Balanced repeated measurements designs. Ann. Statist. 8, 1272–1283 (Correction 1983, 11, 349.).

Cochran, W.G., Autrey, K.M., Cannon, C.Y., 1941. A double change-over designs for dairy cattle feeding experiments. J. Dairy Sci. 24, 937–951.

Hedayat, A.S., Afsarinejad, K., 1978. Repeated measurement designs II. Ann. Statist. 6, 619–628.

Hedayat, A.S., Stufken, J., 2003. Optimal and efficient crossover designs under different assumptions about the carryover effects. J. Biopharm. Statist. 13, 519–528.

Hedayat, A.S., Yang, M., 2003. Universal optimality of balanced uniform crossover designs. Ann. Statist. 31, 978–983.

Hedayat, A.S., Yang, M., 2004. Universal optimality of selected crossover designs. J. Amer. Statist. Assoc., 99, 461–466.

Kempton, R.A., Ferris, S.J., David, O., 2001. Optimal change-over designs when carry-over effects are proportional to direct effects of treatments. Biometrika 88, 391–399.

Kiefer, J., 1975. Construction and optimality of generalized Youden designs. In: Srivastava, J.N. (Ed.), A Survey of Statistical Designs and Linear Models. North-Holland, Amsterdam, pp. 333–353.

Kunert, J., 1983. Optimal design and refinement of the linear model with applications to repeated measurements designs. Ann. Statist. 11, 247–257.

Kunert, J., 1984. Optimality of balanced uniform repeated measurement designs. Ann. Statist. 12, 1006–1017.

Kunert, J., Stufken, J., 2002. Optimal crossover designs in a model with self and mixed carryover effects. J. Amer. Statist. Assoc. 97, 898–906.

Kushner, H.B., 1997. Optimal repeated measurements designs: the linear optimality equations. Ann. Statist. 25, 2328–2344.

Kushner, H.B., 1998. Optimal and efficient repeated-measurements designs for uncorrelated observations. J. Amer. Statist. Assoc. 93, 1176–1187.

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, 311–320 (Correction 1988, 75, 396.).

Stufken, J., 1996. Optimal crossover designs. In: Ghosh, S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 13. North-Holland, Amsterdam, pp. 63–90.