



ELSEVIER

Journal of Statistical Planning and
Inference ■■■ (■■■■) ■■■-■■■journal of
statistical planning
and inference

www.elsevier.com/locate/jspi

Allocation rules for adaptive repeated measurements designs[☆]

H.B. Kushner *

*The Nathan S. Kline Institute for Psychiatric Research, Statistical Sciences and Epidemiology Division,
140 Old Orangeburg Road, Orangeburg, NY 10962, USA*

Received 2 November 1999; received in revised form 2 May 2001; accepted 18 October 2001

Abstract

A method for adaptively allocating new subjects to treatment sequences in repeated measurements designs is given. Simulations indicate that the efficiency of the resulting designs increases with N , avoiding the low efficiency and poor power that can result from an unfortunate design choice. An LR test of no treatment differences is presented. © 2001 Elsevier Science B.V. All rights reserved.

MSC: primary 62K05; secondary 62L05

Keywords: Adaptive design; Optimal design; Repeated measurements; Allocation rule; Treatment effects; Efficiency

1. Introduction

1.1. Optimality when the covariance matrix of responses is unknown

An optimal repeated measurements design to test hypotheses about treatment differences may be specified if \mathbf{V} , the covariance matrix of responses, is known — a rare situation in applications. Unfortunately, very low efficiency can result if an experimenter's assumptions about the covariance matrix are erroneous. For example, suppose that a researcher uses a three-period design to test the equality of two treatments, 1 and 2, but erroneously thinks that \mathbf{V} is autoregressive (see (12)) with $0.562 < \rho \leq 0.675$.

[☆] Research partially supported by National Institute of Mental Health Grants #MH 42959 and #P50MH51359.

* Tel.: +1-914-398-6597; fax: +1-914-398-6592.

E-mail address: kushner@nki.rfmh.org (H.B. Kushner).

The optimal design under that assumption (Matthews, 1987) is

		$N/2$	$N/2$	
	Period 1	1	2	
d_1 :	Period 2	1	2	'
	Period 3	2	1	

where an equal number of subjects, $N/2$, is assigned to each of the above two treatment sequences, (1, 1, 2) and (2, 2, 1) If, in reality, \mathbf{V} is autoregressive with $\rho = -0.8$, the efficiency of d_1 is 0.19 (Matthews, 1987, Table 3), which is unacceptably low. To prevent this possibility, Matthews (1987) recommends the design,

	$N/4$	$N/4$	$N/4$	$N/4$	
	1	2	1	2	
d_2 :	2	1	1	2	'
	2	1	2	1	

which assigns $N/4$ subjects to each of the indicated four treatment sequences. The efficiency of design d_2 is greater than 0.9 for all ρ , and is less than 0.97 only if $\rho > 0.6$. Hence, d_2 is a good design under the assumption that \mathbf{V} is autoregressive. On the other hand, if \mathbf{V} not autoregressive, then the efficiency of d_2 can also be quite low. For example, the efficiency of d_2 is only 0.52 if

$$\mathbf{V} = \begin{pmatrix} 2.17460 & -1.03175 & -1.00000 \\ -1.03175 & 0.619048 & 0.555556 \\ -1.00000 & 0.555556 & 0.587302 \end{pmatrix}. \tag{1}$$

1.2. A response-adaptive approach to the problem

To deal with the problem of specifying an efficient design when little is known about \mathbf{V} , this paper develops an adaptive allocation rule (cf. Silvey, 1980, p. 61) for assigning subjects to treatment sequences.

A response-adaptive design (RAD) is one that is modified on the basis of information in the accrued observations in order to achieve a specified goal. See Rosenberger and Lachin (1993) for a description of the procedures and properties of some of these designs. There are many types of RADs. The classical sequential trial is an RAD in which the decision to terminate the accession of new subjects is based on the goal of minimizing the expected sample size (Armitage, 1975). In play-the-winner designs, the goal, an ethical one, is to minimize the number of subjects on an inferior treatment (Zelen, 1969; Wei and Durham, 1978).

This paper concerns adaptive designs whose goal is to improve the precision of estimators of unknown parameters when \mathbf{V} is not known. Schwabe (1987) has studied the problem of estimation of regression coefficients in an experimental situation in which, because responses are univariate, a fixed (classical, deterministic) optimal design can be specified. He showed that the adaptive design of Gebhardt and Heckendorff (1983), in which adaptation occurs once, concluding the experiment, is superior, with

respect to the criterion in A-optimality theory, to any fixed design. Here, we consider multivariate responses for which, because \mathbf{V} is not known, a fixed, optimal repeated measurements (crossover, changeover) design is not available. We propose adaptation rules that can be used throughout the course of the experiment. The rules specify the assignment of subjects to sequences on the basis of updated estimates of \mathbf{V} . The goal is to increase the precision of treatment effect estimators, i.e., to increase the power of the design.

The adaptive allocation rule is found by replacing the classical design optimality problem by the following problem: After N subjects have been observed on specified treatment sequences, to what sequences should a small number of additional subjects be assigned so as to maximize the resulting information on treatment effects? The paper provides an approximate solution — an adaptive allocation rule — to this problem. Repeated applications of the allocation rule produce the RADs studied in this paper. The performance of the allocation rules, evaluated by simulated experiments, suggest that they produce designs that avoid the poor efficiencies described in Section 1.1.

1.3. Organization of the paper

Section 2 discusses the notion of the efficiency of an adaptive design. Sections 3 and 4, respectively, give the allocation rule for symmetric and for general designs and examine their efficiency through simulations. The derivation of the allocation rule for symmetric designs is included in Section 3; the reasoning that supports the general allocation rule is presented in Appendix B. Section 5 briefly discusses a likelihood ratio (LR) test of no differences in treatment effects, developed in Appendix C. Section 6 summarizes the paper.

2. Efficiency of an adaptive design

We assume the response model

$$Y_{nj} = \tau_{d(n,j)} + \gamma_{d(n,j-1)} + \pi_j + s_n + \varepsilon_{nj}, \quad 1 \leq n \leq N, \quad 1 \leq j \leq p, \quad (2)$$

for a repeated measurements design having p periods, t treatments and N subjects. The design allocates treatment $1 \leq d(n,j) \leq t$ to the n th subject in the j th period. The parameters τ_i and γ_i are, respectively, treatment and carryover effects; π_j is a period effect; and s_n is a subject effect. The N independent vectors $\varepsilon_n = (\varepsilon_{nj}), 1 \leq j \leq p$, are multivariate normal with mean $\mathbf{0}$, and $p \times p$ covariance matrix, \mathbf{V} . All parameters in (2) are fixed. In the sequel, a prominent role will be played by the matrix, \mathbf{V}^* , defined by

$$\mathbf{V}^* = (v^{jk} - v^j v^k / v^{\text{tot}}), \quad 1 \leq i, j \leq p, \quad (3)$$

where v^{jk} is the jk th entry, v^j the j th row sum, and v^{tot} the total sum, of \mathbf{V}^{-1} . The evaluation of an adaptive design is based on the concept of the Mean Square Error matrix of treatment effect contrasts (Gebhardt and Heckendorff, 1983; Schwabe, 1987).

Set $\tau_{i,c} = \tau_i - (\sum_{j=1}^t \tau_j)/t$, the centered τ_i . If $\hat{\tau}_{i,c}$ are estimators of $\tau_{i,c}$, then \mathbf{MSE}_d , the $t \times t$ Mean Square Error matrix for a design d , is defined by

$$\mathbf{MSE}_d = (E[(\hat{\tau}_{i,c} - \tau_{i,c})(\hat{\tau}_{j,c} - \tau_{j,c})]), \quad 1 \leq i, j \leq t. \tag{4}$$

We will use (4) for ML estimators of $\tau_{i,c}$ when (a) \mathbf{V} is unknown and d is an adaptive design and when (b) \mathbf{V} is known and d is a fixed design. In case (a), the calculation of the ML estimators of $\tau_{i,c}$ is discussed in Appendix C.2. In case (b), the ML estimators of $\tau_{i,c}$ are Generalized Least-Squares estimators and (4) is the covariance matrix of $\hat{\tau}_{i,c}, 1 \leq i \leq t$, which is related to the $t \times t$ treatment effects information matrix, \mathbf{C}_d , by (Kunert, 1985)

$$\mathbf{MSE}_d = \mathbf{C}_d^+, \quad d \text{ a fixed design, } \mathbf{V} \text{ known,}$$

where $+$ denotes the Moore–Penrose inverse. Let \mathcal{U} be the cone of non-negative $t \times t$ matrices such that, for $\mathbf{U} \in \mathcal{U}$, the vector $\mathbf{1}_t$ spans the nullspace of \mathbf{U} . The vector $\mathbf{1}_t$ belongs to the nullspace of \mathbf{MSE}_d , since $\sum_{i=1}^t \tau_{i,c} = \sum_{i=1}^t \hat{\tau}_{i,c} = 0$, and we will assume, in case (a), that $\mathbf{MSE}_d \in \mathcal{U}$. The Φ -efficiency of an adaptive design d is

$$\text{eff}_\Phi(d) = \Phi(\mathbf{MSE}_d^+) / \Phi(\mathbf{C}_{d^*}), \tag{5}$$

where Φ is a standard information function defined on \mathcal{U} (Pukelsheim, 1993, pp. 114–119) and d^* is a fixed design that is optimal with respect to the true \mathbf{V} (cf. Schwabe, 1990). If d^* is optimal in approximate design theory, then $\mathbf{C}_{d^*} = Nb(\mathbf{V}^*)(t\mathbf{I}_t - \mathbf{J}_t)/t(t-1)$, where $b(\mathbf{V}^*)$, a scalar function of the matrix \mathbf{V}^* for fixed p and t , is described in Appendix A. In this case, which we assume in this paper, (5) simplifies to

$$\text{eff}_\Phi(d) = t(t-1)\Phi(\mathbf{MSE}_d^+) / Nb(\mathbf{V}^*)\Phi(t\mathbf{I}_t - \mathbf{J}_t). \tag{6}$$

When Φ is the information function used in A-, D- or E-optimality analysis, we denote the efficiency defined in (6) by, respectively, $\text{eff}_A(d), \text{eff}_D(d)$ and $\text{eff}_E(d)$:

$$\text{eff}_A(d) = (t-1)^2 / b(\mathbf{V}^*)N \text{Tr}(\mathbf{MSE}_d), \tag{7a}$$

$$\begin{aligned} \text{eff}_D(d) &= (t-1) / b(\mathbf{V}^*)N (\text{Tr}_{t-1}(\mathbf{MSE}_d))^{1/(t-1)} \\ &= (t-1) / b(\mathbf{V}^*)N (t|(\mathbf{MSE}_d)_{t-1}|)^{1/(t-1)}, \end{aligned} \tag{7b}$$

$$\text{eff}_E(d) = (t-1) / b(\mathbf{V}^*)N \cdot \text{maximum eigenvalue of } \mathbf{MSE}_d, \tag{7c}$$

where $\text{Tr}(\mathbf{R})$ denotes the trace of the square matrix \mathbf{R} , $\text{Tr}_k(\mathbf{R})$ the sum of its $k \times k$ principal determinants, and $|\mathbf{R}_{t-1}|$ its upper left $(t-1) \times (t-1)$ principal determinant. (All the t principal $(t-1) \times (t-1)$ determinants of $\mathbf{MSE}_d \in \mathcal{U}$ are equal.) Unlike the efficiency of a fixed design, the efficiency of an adaptive design can be greater than one, indicating that the adaptive design estimates $\tau_{i,c}$ more accurately than a fixed optimal design. When $t=2$, or when d is symmetric (see Section 3), $\text{eff}_\Phi(d)$ does not depend on Φ . The scalar function $b(\mathbf{V}^*)$ depends on \mathbf{V} since \mathbf{V}^* depends on \mathbf{V} . The matrix \mathbf{MSE}_d depends on the treatment effects and on \mathbf{V} (Appendix C.4). Hence the efficiencies (7) also depend on these unknown parameters.

3. The allocation rule for symmetric designs

3.1. Background on symmetric designs

For $t = 2$, symmetric designs are the dual-balanced designs, such as d_1 and d_2 , introduced by Laska and Meisner (1985). For any t , they are defined as follows. Let $\tau = (t_1, t_2, \dots, t_p)$ denote a treatment sequence, where for $1 \leq j \leq p$ the integers $1 \leq t_j \leq t$ index the treatments. Set $(t)_m = t \cdot (t-1) \cdot \dots \cdot (t-m+1)$, where m is the number of treatments in τ . A symmetry block, denoted $\langle \tau \rangle$, consists of the $(t)_m$ treatment sequences obtained by all possible relabelings of the treatments in τ . In a symmetric design, the same number of subjects is assigned to each treatment sequence in a given symmetry block (Kushner, 1997b). Let n_τ be the number of subjects assigned to τ and let N_τ denote the total number of subjects assigned to symmetry block $\langle \tau \rangle$. Thus $N_\tau = (t)_m n_\tau$. (This divisibility condition on N_τ is not required in the general allocation rule in Section 4.) For example, if $p = t = 3$, the most general symmetric design is

n_1	n_1	n_1	n_2	n_2	n_2	n_2	n_2	n_2	n_3	n_3	n_3	n_3	n_3	n_3	
	1	2	3	1	1	2	2	3	3	1	1	2	2	3	3
	1	2	3	1	1	2	2	3	3	2	3	1	3	1	2
	1	2	3	2	3	1	3	1	2	1	1	2	2	3	3
$d_3 :$															
	n_4	n_4	n_4	n_4	n_4	n_4	n_5	n_5	n_5	n_5	n_5	n_5	n_5	n_5	
	1	1	2	2	3	3	1	1	2	2	3	3			
	2	3	1	3	1	2	2	3	1	3	1	2			
	2	3	1	3	1	2	3	2	3	1	2	1			

where 1, 2, 3 denote treatments, n_1 is the common value of $n_{(111)}, n_{(222)}$ and $n_{(333)}$, n_2 is the common value of $n_{(112)}, n_{(113)}, n_{(221)}, n_{(223)}, n_{(331)}$ and $n_{(332)}$, and similarly for the other treatment sequences and their symmetry blocks. As indicated above, there are five symmetry blocks in design d_3 . Using its first treatment sequence to denote the symmetry block, they are:

- $\langle 111 \rangle$: sequences 1–3; $\langle 112 \rangle$: sequences 4–9; $\langle 121 \rangle$: sequences 10–15;
- $\langle 122 \rangle$: sequences 16–21; $\langle 123 \rangle$: sequences 22–27.

Thus $N_{(111)} = 3n_1$, $N_{(112)} = 6n_2$, $N_{(121)} = 6n_3$, $N_{(122)} = 6n_4$, $N_{(123)} = 6n_5$.

Kushner (1997b) gave a formula, which we now review, for the treatment effects information matrix of any fixed symmetric design. Let $q_\tau(x; \mathbf{V}^*) = a_\tau + b_\tau x + c_\tau x^2$, the non-negative quadratic function associated with treatment sequence τ , whose coefficients are

$$(a_\tau, b_\tau, c_\tau) = \left(\sum_{i=1}^p \sum_{j=1}^p \delta_{i,t_j} v_{ij}^*, 2 \sum_{i=1}^p \sum_{j=1}^{p-1} \delta_{i,t_j} v_{i,j+1}^*, \sum_{i=1}^{p-1} \sum_{j=1}^{p-1} \delta_{i,t_j} v_{i+1,j+1}^* - v_{11}^*/t \right), \tag{8}$$

where δ_{mn} is the Kronecker delta. The coefficients in (8) depend on \mathbf{V}^* , which in turn depends on the unknown covariance matrix, \mathbf{V} . (For models other than (2), the formulas for the quadratic may not require \mathbf{V}^* but only \mathbf{V}^{-1} (Kushner, 1997b, Section 4.4).) For any fixed symmetric design consisting of $N = \sum_{\tau} N_{\tau}$ subjects,

$$\mathbf{C}_d = (A - B^2/4C)(t\mathbf{I}_t - \mathbf{J}_t)/t(t - 1),$$

where A, B and C are linear functions of N_{τ} :

$$(A, B, C) = \left(\sum_{\tau} N_{\tau} a_{\tau}, \sum_{\tau} N_{\tau} b_{\tau}, \sum_{\tau} N_{\tau} c_{\tau} \right). \tag{9}$$

In (9), $\langle \tau \rangle$ runs through all distinct symmetry blocks in the design. In the notation used in design d_3 , $N = N_{(111)} + N_{(112)} + N_{(121)} + N_{(122)} + N_{(123)} = 3n_1 + 6(n_2 + n_3 + n_4 + n_5)$. From (8), the quantities A, B and C depend on \mathbf{V}^* .

3.2. The allocation rule for symmetric designs

Let $\mathcal{N} = (N_{\tau})$ denote the array specifying the number of subjects assigned by a given d to the symmetry blocks, and let $F(\mathcal{N}) = A - B^2/4C$, a function of all the N_{τ} . The quantity $\Delta F = F(\mathcal{N}') - F(\mathcal{N})$ measures the increment of treatment information in design d' where $\mathcal{N}' = (N'_{\tau})$ with $N'_{\tau} \geq N_{\tau}$. If the number of such designs with fixed total number of subjects, $\sum_{\tau} N'_{\tau}$, is not too large, one could find, by an exhaustive search, the designs that maximize ΔF . The following alternative approximate approach (and also that in Section 4) is intended for a situation in which such a search is not feasible.

By treating the N_{τ} as continuous variables, a direct calculation gives

$$\partial F / \partial N_{\tau} = q_{\tau}(\theta; \mathbf{V}^*),$$

where

$$\theta = -B/2C. \tag{10}$$

It follows that

$$\Delta F \approx \sum_{\tau} q_{\tau}(\theta; \mathbf{V}^*) \Delta N_{\tau}.$$

Since $q_{\tau}(\theta; \mathbf{V}^*) \geq 0$ for all τ , choosing $\Delta N_{\tau} \geq 0$ to maximize ΔF suggests the following allocation rule for symmetric designs:

$$\begin{aligned} & \text{Assign the new subjects to a symmetric block } \langle \tau \rangle \text{ for which } q_{\tau}(\theta; \mathbf{V}^*) \\ & \text{is maximized.} \end{aligned} \tag{11}$$

It is possible that the maximum in (11) is achieved at more than one τ , but the choice among them is arbitrary. If ΔN new subjects are added, so that N increases to $N + \Delta N$, the rule implies $\Delta N_{\tau} = \Delta N$ and $\Delta N_{\kappa} = 0$ for each $\kappa \neq \tau$, and $F(\mathcal{N})$ increases approximately by $\Delta N \cdot q_{\tau}(\theta; \mathbf{V}^*)$.

To preserve the symmetry of each design, we will assume that ΔN is a multiple of the number of sequences in the assigned symmetric block. For example, ΔN is a multiple of 2 when $t = 2$ and a multiple of 3 or 6 when $t = 3$. In a manner that preserves design symmetry, the new subjects are then randomly assigned to the treatment sequences in the symmetric block determined by (11). The method can be generalized to cyclic designs, permitting ΔN to be any fixed integer greater than one. If these conditions on ΔN cannot be satisfied throughout the trial, then the method in Section 4, which has no conditions on ΔN , can be used.

3.3. Use of the allocation rule

Having so far observed N subjects, one computes, as described in Appendix C.2, \mathbf{V}_N^* , the “current” ML estimate of \mathbf{V}^* . Also, θ_N , the current ML estimate of θ , is obtained by replacing \mathbf{V}^* by \mathbf{V}_N^* in (8)–(10). Finally, one computes $q_\tau(\theta_N; \mathbf{V}_N^*)$ an ML estimate of $q_\tau(\theta; \mathbf{V}^*)$ based on N observations, and substitutes it in (11) to determine the single symmetry block, $\langle \tau \rangle$, to which ΔN new subjects should be allocated. The form of (11) for use in an experiment is:

$$\text{Assign the new subjects to the symmetric block } \langle \tau \rangle \text{ that maximizes } q_\tau(\theta_N; \mathbf{V}_N^*), \text{ where } \theta_N \text{ and } \mathbf{V}_N^* \text{ are current estimates of } \theta \text{ and } \mathbf{V}^*. \quad (11)'$$

After $M = N + \Delta N$ subjects have been observed, updated estimates, \mathbf{V}_M^* and $q_\tau(\theta_M; \mathbf{V}_M^*)$, are computed for the next allocation.

3.4. Simulated experiments for symmetric designs

In this section we assess, based on simulations, the performance of the allocation rule for symmetric designs. Restricting attention to $p = 3$, we will consider, in Sections 3.4.1–3.4.3, experimental situations for $t = 2$ with \mathbf{V} as in (1), and, for $2 \leq t \leq 3$, with \mathbf{V} the 3×3 autoregressive matrix,

$$\mathbf{V} = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}, \quad -1 < \rho < 1. \quad (12)$$

For these cases, $b(\mathbf{V}^*)$ is given in Appendix A. The graphs in Figs. 1 and 2 were computed under the assumption $\tau_1 = \tau_2 = 0$, $\gamma_1 = 1$, $\gamma_2 = -1$ and, in Fig. 3, under the assumption $\tau_1 = \tau_2 = \tau_3 = 0$, $\gamma_1 = 1$, $\gamma_2 = 2$, $\gamma_3 = -3$. The values of the period and subject effects are immaterial (see Appendix C) and so were set to zero.

In the simulations, the initial (N_i) and final (N_f) number of subjects were set in advance. The initial subjects were assigned arbitrarily to treatment sequences. (It is possible that the efficiencies of the adaptive designs can be improved if the initial subjects are assigned more systematically.) The n th subject, for $n > N_i$, was assigned to a treatment sequence according to Section 3.3 and its p -dimensional response was

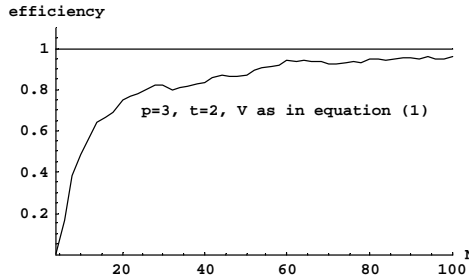


Fig. 1. Efficiency vs. number of subjects in adaptive dual-balanced designs obtained by allocation rule (11)' in 1000 simulated experiments.

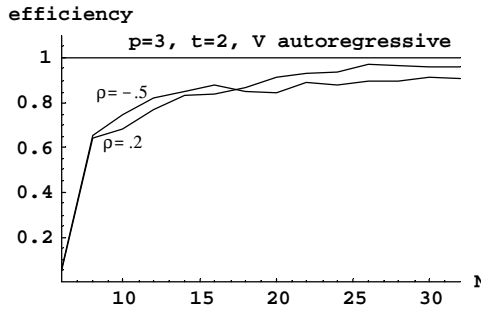


Fig. 2. Efficiency vs. number of subjects in adaptive dual-balanced designs obtained by allocation rule (11)' in 1000 simulated experiments.

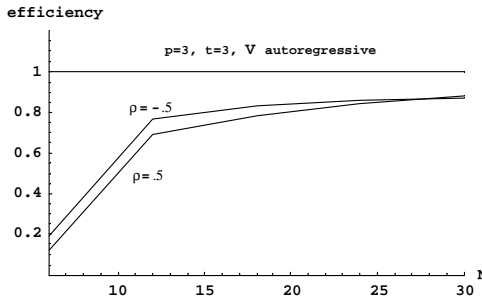


Fig. 3. Efficiency vs. number of subjects in adaptive symmetric designs obtained by allocation rule (11)' in 1000 simulated experiments.

then generated assuming a multivariate normal distribution having mean vector $(\tau_{d(n,j)} + \gamma_{d(n,j-1)})$, $1 \leq j \leq p$, and covariance matrix \mathbf{V} . In the k th simulated experiment, after the responses of N subjects were generated, the equations in Appendix C.2 yielded both \mathbf{V}_N^* , the current ML estimate of \mathbf{V}^* required for the next allocation of subjects, and $\tau_{i,N}(k)$, the current ML estimate of τ_i . After 1000 simulations, $\mathbf{MSE}_d(N)$, an estimate

of \mathbf{MSE}_d based on the first N subjects, was computed by

$$\mathbf{MSE}_d(N) = \left(\sum_{k=1}^{1000} (\tau_{i,N}(k) - \tau_i)(\tau_{j,N}(k) - \tau_j) / 1000 \right), \quad 1 \leq i, j \leq t.$$

$\mathbf{MSE}_d(N)$ was used in place of \mathbf{MSE}_d in (7a), yielding an estimate, $\text{eff}_A(d_N)$, of the A-efficiency of the allocation rule in an N -subject symmetric design, d_N . It is not necessary to compute $\text{eff}_D(d_N)$ from (7b) or $\text{eff}_E(d_N)$ from (7c) since, when d is a symmetric design, or when $t=2$, all measures of efficiency coincide (unlike the general design situation considered in Section 4). Figs. 1–3 plot the points $(N, \text{eff}(d_N))$, where $N \in [N_i, N_f]$, obtained from 1000 simulated experiments, and $\text{eff}(d_N)$ is the common value of $\text{eff}_A(d_N)$, $\text{eff}_D(d_N)$ and $\text{eff}_E(d_N)$.

3.4.1. $p=3, t=2, \mathbf{V}$ as in (1)

Here the most general dual-balanced design is

n_1	n_1	n_2	n_2	n_3	n_3	n_4	n_4
1	2	1	2	1	2	1	2
1	2	1	2	2	1	2	1
1	2	2	1	1	2	2	1

Thus $N_{(111)} = 2n_1$, $N_{(112)} = 2n_2$, $N_{(121)} = 2n_3$, $N_{(122)} = 2n_4$ and $N = 2(n_1 + n_2 + n_3 + n_4)$. The allocation rule (11)' was sequentially applied, starting with an initial number of $N_i = 4$ subjects with $n_1 = n_3 = 0$, $n_2 = n_4 = 1$, i.e., the initial design is d_2 with $N = 4$. The number of new subjects was always $\Delta N = 2$, until finally there was a total of $N_f = 100$ subjects. The N in each simulated experiment run through the even integers in $[4, 100]$. In each simulation, the allocation rule (11)' was used $48 = (90 - 4) / 2$ times.

Fig. 1 shows how the efficiency of initial design d_2 improves as N increases. The graph shows a rapid initial growth in efficiency and then a slower (perhaps asymptotic) growth tending to 1. For the 1000 simulated adaptive designs, the final efficiency, for $N_f = 100$, was 0.96 and the average of the final proportions, $(N_{(111)}, N_{(112)}, N_{(121)}, N_{(122)}) / N_f$, was $(0, 0.95988, 0, 0.04012, 0)$. Under (1), the optimal proportions in approximate fixed design theory are $(0, 0.99490, 0, 0.005010)$.

3.4.2. $p=3, t=2, \mathbf{V}$ autoregressive

The most general symmetric design is as in Section 3.4.1. The allocation rule (11)' was sequentially applied, starting with an initial number of $N_i = 6$ subjects with $n_1 = 0$, $n_2 = n_3 = n_4 = 1$. The number of new subjects was always $\Delta N = 2$, until there was a total of $N_f = 32$ subjects. The N in each simulation run through the even integers in $[6, 32]$. In each simulation, the allocation rule (11)' was used $13 = (32 - 6) / 2$ times.

Fig. 2 again shows how the efficiency of the design improves as N increases. Simulations were performed for each of the five covariance matrices (12) corresponding to $\rho \in \{-0.5, 0.2, 0.5, 0.6, 0.8\}$, values which belong to the five intervals appearing in (A.3). For simplicity, only the simulations for $\rho = -0.5$ and for $\rho = 0.2$ are presented (the graphs for the other values of ρ are similar). For $\rho = -0.5$ and 1000 simulated adaptive

Table 1
True parameters vs. average of their 1000 ML estimates from the $\rho = 0.2$ simulation

	True value	Average final ($N_f = 32$) ML estimate
τ_1	0	0.00151867
γ_1	1	0.995857
\mathbf{V}^*	$\begin{pmatrix} 0.744048 & -0.446429 & -0.297619 \\ -0.446429 & 0.892857 & -0.446429 \\ -0.297619 & -0.446429 & 0.744048 \end{pmatrix}$	$\begin{pmatrix} 0.848823 & -0.507437 & -0.431386 \\ -0.507437 & 1.03026 & -0.522824 \\ -0.431386 & -0.522824 & 0.864209 \end{pmatrix}$

designs, the final efficiency, for $N_f = 32$, was 0.96 and the average of the final proportions, $(N_{(111)}, N_{(112)}, N_{(121)}, N_{(122)})/N_f$, was (0.000188, 0.288313, 0.0625, 0.649). From Matthews (1987), the optimal proportions in approximate fixed design theory are (0, 0.4, 0, 0.6). For $\rho = 0.2$, the final efficiency was 0.91 and the average of the final proportions was (0.00081, 0.35331, 0.0625, 0.58338). The optimal proportions in approximate fixed design theory are (0, 0, 0.0078, 0.9922). Table 1 compares the true parameter values and their average final ML estimates.

3.4.3. $p=t=3$, \mathbf{V} autoregressive

The most general symmetric design is displayed in d_3 . For initial $N_i=6$ designs with $n_1=n_2=n_3=n_5=0$ and $n_4=1$, Fig. 3 was obtained for $N \in [6, 30]$. Here $\Delta N=6$ always, i.e., the rule never assigned a subject to symmetry block $\langle 111 \rangle$. Hence the N in Fig. 3 run through the integers in $[6, 30]$ which are divisible by 6. In each simulation, the allocation rule $(11)'$ was used $4 = (30 - 6)/6$ times.

Fig. 3 also shows that the efficiency of the design improves as N increases. The simulations were performed for two covariance matrices (12) corresponding to $\rho \in \{-0.5, 0.5\}$, values that belong to the two intervals appearing in (A.4). For $\rho = -0.5$ and 1000 simulated adaptive designs, the final efficiency, for $N_f = 30$, was 0.87 and the average of the final proportions, $(N_{(111)}, N_{(112)}, N_{(121)}, N_{(122)}, N_{(123)})/N_f$, was (0.0002, 0.1192, 0, 0.3478, 0.5328). The optimal proportions in approximate fixed design theory are (0, 0, 0, 0.20833, 0.79167). For $\rho = 0.5$, the final efficiency was 0.88 and the average of the final proportions was (0.0004, 0.1082, 0, 0.2702, 0.6212). The optimal proportions in approximate fixed design theory are (0, 0, 0, 0, 1).

4. The general allocation rule

Here, as in Section 3, p and t are still arbitrary, as is the number of subjects, n_τ , assigned to treatment sequence τ . The quantity N_τ plays no role and therefore the divisibility restrictions in Section 3 do not apply. This section states the general allocation rule, which can be used for any design. It then gives and tests simplified versions of the rule for the A-, D-, and E-optimality criteria. The reasoning that underlies the

general rule is found in Appendix B. Let

$$\begin{pmatrix} \mathbf{C}_{d11} & \mathbf{C}_{d12} \\ \mathbf{C}_{d21} & \mathbf{C}_{d22} \end{pmatrix} (\mathbf{C}_{d12} = \mathbf{C}'_{d21})$$

denote the joint treatment/carryover effects information matrix. For any $t \times t$ matrix, \mathbf{X} , set

$$\mathbf{Q}_{d,\tau}(\mathbf{X}; \mathbf{V}^*) = (\mathbf{T}_\tau - \mathbf{M}_d + (\tilde{\mathbf{T}}_\tau - \tilde{\mathbf{M}}_d)\mathbf{X})'\mathbf{V}^*(\mathbf{T}_\tau - \mathbf{M}_d + (\tilde{\mathbf{T}}_\tau - \tilde{\mathbf{M}}_d)\mathbf{X}), \quad (13)$$

where \mathbf{T}_τ (respectively, $\tilde{\mathbf{T}}_\tau$) is the $p \times t$ treatment (respectively, carryover) effects incidence matrix corresponding to treatment sequence τ ,

$$(\mathbf{M}_d, \tilde{\mathbf{M}}_d) = \left(\sum_{\tau} n_{\tau} \mathbf{T}_{\tau} / N, \sum_{\tau} n_{\tau} \tilde{\mathbf{T}}_{\tau} / N \right) \quad (14)$$

and

$$N = \sum_{\tau} n_{\tau}. \quad (15)$$

In (14)–(15), the τ run over all t^p treatment sequences. If d is a symmetric design, then $\mathbf{M}_d = \mathbf{J}_{p,t}/t$ where $\mathbf{J}_{p,t}$ is the $p \times t$ matrix of ones; and $\tilde{\mathbf{M}}_d = \mathbf{J}_{p,t}/t$, except that the first row of $\tilde{\mathbf{M}}_d$ is zero. Therefore, for symmetric designs, $\mathbf{Q}_{d,\tau}(\mathbf{X}; \mathbf{V}^*)$ is a matrix analog of the quadratic $q_{\tau}(x; \mathbf{V}^*)$ in Section 3. Also, from $(\mathbf{T}_\tau - \mathbf{M}_d)\mathbf{J}_t = \mathbf{0}$, $\mathbf{C}_{21}\mathbf{J}_t = \mathbf{0}$ and (13),

$$\mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22}^{-} \mathbf{C}_{d21}; \mathbf{V}^*)\mathbf{J}_t = \mathbf{0}. \quad (16)$$

Let \mathcal{U}_0 denote a neighborhood of \mathcal{U} . Assume also that Φ can be extended to a function, defined on \mathcal{U}_0 , whose partial derivatives are continuous. (We do not require that Φ be concave on all of \mathcal{U}_0 .) Set $\Phi_{ij}(\mathbf{U}) = \partial\Phi/\partial u_{ij}$, $1 \leq i, j \leq t$, $\mathbf{U} = (u_{ij}) \in \mathcal{U}_0$ and

$$\mathbf{D}_{\Phi}(\mathbf{U}) = [(1 + \delta_{ij})\Phi_{ij}(\mathbf{U})], \quad 1 \leq i, j \leq t, \quad \mathbf{U} \in \mathcal{U}.$$

4.1. The general allocation rule

For an arbitrary but small number of new subjects, ΔN , the general allocation rule is:

Given design d , assign all new subjects to a treatment sequence τ that maximizes

$$\text{Tr}[\mathbf{D}_{\Phi}(\mathbf{C}_d) \cdot \mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22}^{-} \mathbf{C}_{d21}; \mathbf{V}^*)]. \quad (17)$$

Any generalized inverse, \mathbf{C}_{d22}^{-} , can be used in (17). In the simulations in Section 4.3, new subjects are added one or two at a time, i.e., $\Delta N = 1$ or 2.

4.2. The general allocation rule when Φ is the A-, D-, or E-optimality criterion

This section presents formulas for the matrix $\mathbf{D}_{\Phi}(\mathbf{C}_d)$ when Φ is the A-, D-, or E-optimality criterion, and evaluates (17) for these cases. Let $0 = \lambda_t(\mathbf{U}) \leq \dots \leq \lambda_1(\mathbf{U})$ be the t ordered eigenvalues of \mathbf{U} .

(i) *The A-allocation rule:* Let $\Phi_A(\mathbf{U}) = \text{Tr}_{t-1}(\mathbf{U})/\text{Tr}_{t-2}(\mathbf{U}), \mathbf{U} \in \mathcal{U}_0$, where $\mathcal{U}_0 = \{\mathbf{U} : \text{Tr}_{t-2}(\mathbf{U}) > 0, \mathbf{U} \geq 0\}$. Then $\Phi_A(\mathbf{U}) = (\sum_{i=1}^{t-1} (\lambda_i(\mathbf{U}))^{-1})^{-1}, \mathbf{U} \in \mathcal{U}$, the information function for A-optimality analysis when $\mathbf{C}_d \in \mathcal{U}$. One can show that

$$\mathbf{D}_{\Phi_A}(\mathbf{U})/2 = (\Phi_A(\mathbf{U})\mathbf{U}^+)^2 + h\mathbf{J}_t/t, \quad \mathbf{U} \in \mathcal{U}, \tag{18}$$

where $h = 1 - \text{Tr}_{t-3}(\mathbf{U}) \cdot \text{Tr}_{t-1}(\mathbf{U})/(\text{Tr}_{t-2}(\mathbf{U}))^2$ for $t \geq 3$ and $h = 1$ for $t = 2$. From (16)–(18), the “A-allocation rule” is:

Given design d , assign all new subjects to the treatment sequence τ that maximizes

$$\text{Tr}[(\mathbf{C}_d^+)^2 \cdot \mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22}^- \mathbf{C}_{d21}; \mathbf{V}^*)]. \tag{19}$$

(ii) *The D-allocation rule:* Let $\Phi_D(\mathbf{U}) = (\text{Tr}_{t-1}(\mathbf{U}))^{1/(t-1)}, \mathbf{U} \in \mathcal{U}_0$, where $\mathcal{U}_0 = \{\mathbf{U} : \text{Tr}_{t-1}(\mathbf{U}) > 0, \mathbf{U} \geq 0\}$. Then $\Phi_D(\mathbf{U}) = (\prod_{i=1}^{t-1} \lambda_i(\mathbf{U}))^{1/(t-1)}, \mathbf{U} \in \mathcal{U}$, the information function for D-optimality analysis when $\mathbf{C}_d \in \mathcal{U}$. One can show that

$$(t-1) \cdot \mathbf{D}_{\Phi_D}(\mathbf{U})/2 = \Phi_D(\mathbf{U})\mathbf{U}^+ + h\mathbf{J}_t/t, \quad \mathbf{U} \in \mathcal{U}, \tag{20}$$

where $h = \text{Tr}_{t-2}(\mathbf{U})/(\text{Tr}_{t-1}(\mathbf{U}))^{(t-2)/(t-1)}$. From (16), (17) and (20), the “D-allocation rule” is:

Given design d , assign all new subjects to the treatment sequence τ that maximizes

$$\text{Tr}[\mathbf{C}_d^+ \cdot \mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22}^- \mathbf{C}_{d21}; \mathbf{V}^*)]. \tag{21}$$

(iii) *The E-allocation rule:* For $\Phi_E(\mathbf{U}) = \lambda_{t-1}(\mathbf{U}), \mathbf{U} \in \mathcal{U}_0$, where $\mathcal{U}_0 = \{\mathbf{U} : \mathbf{U} \geq 0\}$, we get $\Phi_E(\mathbf{U}) = \lambda_{\min}(\mathbf{U}), \mathbf{U} \in \mathcal{U}$, the minimum non-zero eigenvalue of \mathbf{U} , which is the information function for E-optimality analysis when $\mathbf{C}_d \in \mathcal{U}$. One can show that the partial derivatives of $\Phi_E(\mathbf{U})$ exist for all \mathbf{U} (even if multiplicity $[\lambda_{t-1}(\mathbf{U})] > 1$) and that, if multiplicity $[\lambda_{\min}(\mathbf{U})] = 1$, then

$$\mathbf{D}_{\Phi_E}(\mathbf{U})/2 = \omega_{\min}(\mathbf{U})\omega'_{\min}(\mathbf{U}), \quad \mathbf{U} \in \mathcal{U}, \tag{22}$$

where $\omega_{\min}(\mathbf{U})$ is the eigenvector of length 1 corresponding to $\lambda_{\min}(\mathbf{U})$. We omit the formula for $\mathbf{D}_{\Phi_E}(\mathbf{U})$ when multiplicity $[\lambda_{\min}(\mathbf{U})] > 1$, since this case is unlikely to occur in practice. From (17) and (22), the “E-allocation rule” is:

Given design d such that multiplicity $[\lambda_{\min}(\mathbf{C}_d)] = 1$, assign all new subjects to the treatment sequence τ that maximizes

$$\omega'_{\min}(\mathbf{C}_d)\mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22}^- \mathbf{C}_{d21}; \mathbf{V}^*)\omega_{\min}(\mathbf{C}_d). \tag{23}$$

4.3. Use of the A-, D-, or E-general allocation rules

The procedure is similar to that in Section 3.3: From \mathbf{V}_N^* (see Appendix C), the current ML estimate of \mathbf{V}^* , one computes from (B.1)–(B.2) the matrices $\mathbf{C}_{d21,N}, \mathbf{C}_{d22,N}$ and $\mathbf{C}_{d,N}$, the current estimates of $\mathbf{C}_{d21}, \mathbf{C}_{d22}$ and \mathbf{C}_d , and substitutes them in (19), (21) and (23) to obtain the A-, D- and E-general allocation rules for use in experiments:

A-allocation rule: Given design d with N subjects, assign all new subjects to the treatment sequence τ that maximizes

$$\text{Tr}[(\mathbf{C}_{d,N}^+)^2 \cdot \mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22,N}^- \mathbf{C}_{d21,N}; \mathbf{V}_N^*)]. \tag{19}'$$

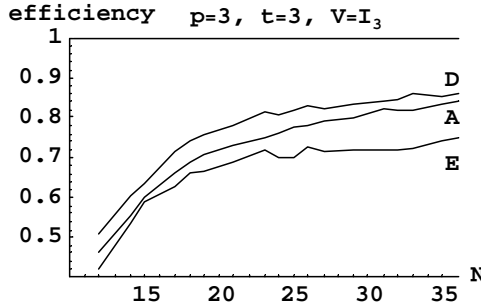


Fig. 4. A-, D-, and E-efficiencies vs. number of subjects in adaptive designs obtained by general allocation rules (19)', (21)' and (23)' in 1000 simulated experiments.

D-allocation rule: *Given design d with N subjects, assign all new subjects to the treatment sequence τ that maximizes*

$$\text{Tr}[\mathbf{C}_{d,N}^+ \cdot \mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22,N}^- \mathbf{C}_{d21,N}; \mathbf{V}_N^*)]. \tag{21}'$$

E-allocation rule: *Given design d with N subjects such that multiplicity $[\lambda_{\min}(\mathbf{C}_{d,N})] = 1$, assign all new subjects to the treatment sequence τ that maximizes*

$$\omega'_{\min}(\mathbf{C}_{d,N}) \mathbf{Q}_{d,\tau}(-\mathbf{C}_{d22,N}^- \mathbf{C}_{d21,N}; \mathbf{V}_N^*) \omega_{\min}(\mathbf{C}_{d,N}). \tag{23}'$$

4.4. Simulated experiments for the A-, D-, and E-general allocation rules

As in Section 3.4, we examine, by simulations, the performance of the A-, D- and E-general allocation rules. We consider the case $p = t = 3$. The most general design has the 27 treatment sequences in d_3 , but with any n_τ on sequence τ . Suppose that $\mathbf{V} = \mathbf{I}_3$. From (A.4) with $\rho = 0$, $b((\mathbf{I}_3)^*) = \frac{29}{18}$. For an initial design with $n_\tau = 1$ for $\tau \in \{(113), (122), (131), (133), (211), (223), (233), (311), (313), (333)\}$, Fig. 4 was obtained for $N \in [10, 36]$. In all of the simulated experiments, a common sequence of ΔN was fixed in advance ($\Delta N = 1$, 8 times; $\Delta N = 2$, 9 times). Consequently, in each simulated experiment, N runs through the same values between 10 and 36 and one of the allocation rules, (19)', (21)' or (23)', was used 17=8+9 times. The simulations in Fig. 4 were conducted with $\tau_1 = \tau_2 = \tau_3 = 0$, $\gamma_1 = 1$, $\gamma_2 = 2$, $\gamma_3 = -3$. For a general design, the A-, D- and E-efficiencies do not coincide (cf. Section 3.4). When $\text{MSE}_d(N)$ — computed as in Section 3.4 — is used in place of MSE_d in (7a)–(7c), one obtains $\text{eff}_A(d_N)$, $\text{eff}_D(d_N)$ and $\text{eff}_E(d_N)$, which are plotted in Fig. 4.

The (rounded) average proportions for the D-optimality simulation for $N = 36$ were, in the order of the 27 sequences in d_3 : (0, 0, 0.03; 0.01, 0.03, 0.01, 0.03, 0, 0; 0, 0.03, 0, 0, 0.03, 0; 0.05, 0.04, 0.07, 0.05, 0.04, 0.07; 0.09, 0.10, 0.06, 0.09, 0.09, 0.09). From Kushner (1998), in an optimal fixed design there are no subjects in the first 15 sequences but the total number of subjects in sequences 16–21 is one-fifth of that in sequences 22–27. For $p = 3$, $\mathbf{V} = \mathbf{I}_3$ and $2 \leq t \leq 5$, Table 2 gives the final ($N_f = 36$) efficiencies from the above $t = 3$ simulation and additional simulations.

Table 2

For $p=3$, $\mathbf{V}=\mathbf{I}_3$, final efficiencies for $2 \leq t \leq 5$ and various criteria in adaptive designs obtained by general allocation rules (19)', (21)' and (23)' in 1000 simulated experiments

t	2	3	3	3	4	4	4	5	5	5
Criterion	Any	A	D	E	A	D	E	A	D	E
Final ($N_f = 36$) efficiency	0.91	0.84	0.86	0.75	0.85	0.84	0.70	0.85	0.84	0.79

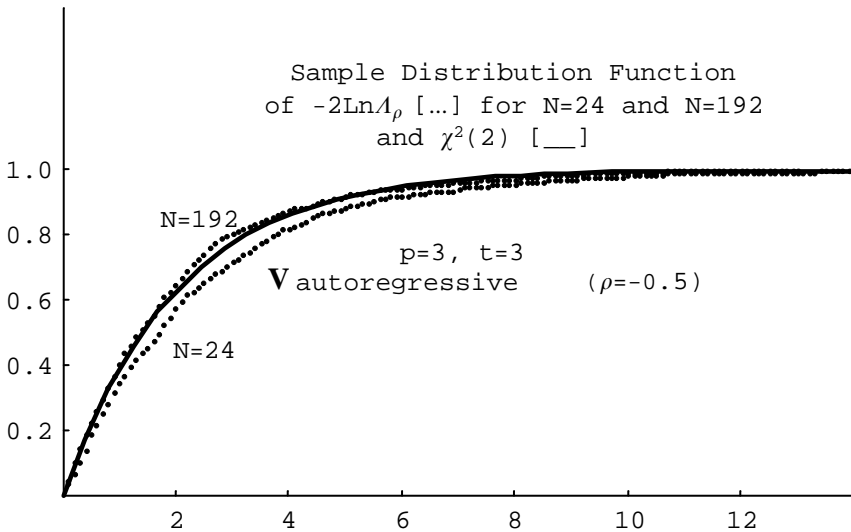


Fig. 5. The sample ($N = 24$ and 192) distribution function of $-2\text{Ln}A_\rho$ in the situation of Fig. 3 (\mathbf{V} autoregressive with $\rho = -0.5$), and the distribution function of $\chi^2(2)$.

5. Testing the hypothesis of identical treatment effects

In Appendix C, we propose a statistic, (C.9), for testing $H: \tau_1 = \dots = \tau_t$, the hypothesis of no treatment effects differences. Appendix C.1, by Laska and Meisner, proves that this test, developed in Appendices C.2–C.4, is an LR test. The asymptotic distribution of the test statistic, therefore, is $\chi^2(t - 1)$. The non-central distribution of the statistic depends only on the true values of τ_i , $1 \leq i \leq t$, and on \mathbf{V}^* , but not on the nuisance parameters — the subject, period, and carryover effects — in model (2). To investigate, by simulations, a specific instance of the chi-square approximation to the distribution of the statistic, suppose that \mathbf{V} is given by (12), and let $-2\text{Ln}A_\rho$ denote the random variable given in (C.9). Fig. 5 plots the sample distribution of the statistic obtained from simulations in the situation of Fig. 3 for $\rho = -0.5$ with $N = 24$ and $N = 192$, and also graphs the distribution of $\chi^2(2)$.

In the experimental situation considered in this section, Fig. 5 indicates the closeness of the distribution of the test statistic to $\chi^2(2)$.

6. Summary

This paper gives response-adaptive rules for allocation of new subjects to treatment sequences in repeated measurements designs. The simulations in Figs. 1–4 and the final efficiencies in Table 2 indicate how the efficiency of the resulting designs grows with N , averting the low efficiency that can result if designs are specified when the covariance matrix of responses is unknown. An LR test of identical treatment effects is presented.

Acknowledgements

The formulation of the optimality problem considered here is due to E. Laska. I am indebted to him and to M. Meisner for Section C.1 and for extensive conversations concerning the methods and results of the paper. I am grateful to J. Wanderling for his assistance in preparing the graphs. The research was partially supported by National Institute of Mental Health Grants #MH 42959 and #P50MH51359.

Appendix A

The general definition of $b(\mathbf{V}^*)$ and its values when \mathbf{V} is as in (1) and $t = 2$ or as in (12) and $2 \leq t \leq 3$

A.1. The general definition of $b(\mathbf{V}^*)$

With \mathbf{V}^* as in (3), $b(\mathbf{V}^*)$ is the scalar (denoted by b in Kushner, 1997b):

$$b(\mathbf{V}^*) = \min_{-\infty < x < \infty} \max_{\tau} q_{\tau}(x; \mathbf{V}^*), \tag{A.1}$$

where $q_{\tau}(x; \mathbf{V}^*)$ is as in Section 3.1. When \mathbf{V} is known, $b(\mathbf{V}^*)$ can be computed via

$$b(\mathbf{V}^*) = \max_{\tau, \kappa} y_{\tau, \kappa},$$

where the point $(x_{\tau, \kappa}, y_{\tau, \kappa})$ in the real plane (the “admissible intersection” of the quadratics $q_{\tau}(x; \mathbf{V}^*)$ and $q_{\kappa}(x; \mathbf{V}^*)$, if it exists) is defined by

$$y_{\tau, \kappa} = q_{\tau}(x_{\tau, \kappa}; \mathbf{V}^*) = q_{\kappa}(x_{\tau, \kappa}; \mathbf{V}^*) \quad \text{with} \quad q'_{\tau}(x_{\tau, \kappa}; \mathbf{V}^*) \cdot q'_{\kappa}(x_{\tau, \kappa}; \mathbf{V}^*) \leq 0.$$

A.2. The values of $b(\mathbf{V}^*)$ when \mathbf{V} is as in (1) and $t = 2$ or as in (12) and $2 \leq t \leq 3$

When $p = 3$, $t = 2$ and \mathbf{V} is as in (1), then from (A.1),

$$b(\mathbf{V}^*) = 21.1092. \tag{A.2}$$

When $p = 3, t = 2$, and \mathbf{V} is as in (12), it follows from Matthews (1987) that

$$(1 - \rho^2)(3 - \rho)b(\mathbf{V}^*) = \begin{cases} 4, & -1 < \rho \leq 0, \\ 4(9 + 6\rho - 2\rho^2)/(3 + \rho)^2, & 0 < \rho \leq 0.464, \\ (7 - 4\rho + 11\rho^2 + 6\rho^3)/(1 + \rho)^2, & 0.464 < \rho \leq 0.562, \\ (3 - \rho)(1 + \rho), & 0.562 < \rho \leq 0.675, \\ (2(2 + 5\rho + 6\rho^2 + 2\rho^3) - (2\rho + \rho^2) \\ (1 + 2\rho + 2\rho^2)^{1/2})/(1 + \rho)^2, & 0.675 < \rho < 1. \end{cases} \quad (\text{A.3})$$

When $p = t = 3$, and \mathbf{V} is as in (12), it follows from Kushner (1997a) that

$$(1 - \rho^2)(3 - \rho)b(\mathbf{V}^*) = \begin{cases} (29 + \rho - \rho^2 + 3\rho^3)/6, & -1 < \rho \leq \frac{1}{3}, \\ 8(3 + 2\rho)/(5 + 3\rho), & 1/3 < \rho < 1. \end{cases} \quad (\text{A.4})$$

Appendix B

The reasoning that supports the general allocation rule (17)

The $t \times t$ matrices $\mathbf{C}_{d11}, \mathbf{C}_{d21}$ and \mathbf{C}_{d22} are

$$\mathbf{C}_{d11} = \sum_{\tau} n_{\tau}(\mathbf{T}_{\tau} - \mathbf{M}_d)' \mathbf{V}^*(\mathbf{T}_{\tau} - \mathbf{M}_d), \quad (\text{B.1a})$$

$$\mathbf{C}_{d21} = \sum_{\tau} n_{\tau}(\tilde{\mathbf{T}}_{\tau} - \tilde{\mathbf{M}}_d)' \mathbf{V}^*(\mathbf{T}_{\tau} - \mathbf{M}_d) \quad (\text{B.1b})$$

and

$$\mathbf{C}_{d22} = \sum_{\tau} n_{\tau}(\tilde{\mathbf{T}}_{\tau} - \tilde{\mathbf{M}}_d)' \mathbf{V}^*(\tilde{\mathbf{T}}_{\tau} - \tilde{\mathbf{M}}_d). \quad (\text{B.1c})$$

The τ in (B.1) run over all t^P treatment sequences. The treatment effects information matrix, \mathbf{C}_d , is

$$\mathbf{C}_d = \mathbf{C}_{d11} - \mathbf{C}'_{d21} \mathbf{C}^-_{d22} \mathbf{C}_{d21}. \quad (\text{B.2})$$

Viewed as a function of a particular non-negative integer, n_{τ} , both \mathbf{C}_d and $\Phi(\mathbf{C}_d)$ are non-decreasing functions. We now consider the analogs of these facts when n_{τ} is a continuous variable. For any $t \times t$ matrix, \mathbf{X} , let

$$\mathbf{Q}_d(\mathbf{X}; \mathbf{V}^*) = \sum_{\tau} n_{\tau} \mathbf{Q}_{d,\tau}(\mathbf{X}; \mathbf{V}^*) = \mathbf{C}_{d11} + \mathbf{C}'_{d21} \mathbf{X} + \mathbf{X}' \mathbf{C}_{d21} + \mathbf{X}' \mathbf{C}_{d22} \mathbf{X}.$$

Both $\mathbf{Q}_{d,\tau}(\mathbf{X}; \mathbf{V}^*)$, given in (13), and $\mathbf{Q}_d(\mathbf{X}; \mathbf{V}^*)$ are non-negative $t \times t$ matrices: $\mathbf{Q}_d(\mathbf{X}; \mathbf{V}^*) \geq 0$ and $\mathbf{Q}_{d,\tau}(\mathbf{X}; \mathbf{V}^*) \geq 0$, all \mathbf{X} . From Pukelsheim (1993, p. 75), the matrices \mathbf{C}_d and $\mathbf{Q}_d(\mathbf{X}; \mathbf{V}^*)$ are related by

$$\mathbf{Q}_d(\mathbf{X}; \mathbf{V}^*) = (\mathbf{X} + \mathbf{C}^-_{d22} \mathbf{C}_{d21})' \mathbf{C}_{d22} (\mathbf{X} + \mathbf{C}^-_{d22} \mathbf{C}_{d21}) + \mathbf{C}_d. \quad (\text{B.3})$$

Suppose that $\mathbf{C}_{d11}, \mathbf{C}_{d21}$ and \mathbf{C}_{d22} are functions of some scalar variable, v . Then, viewing $\mathbf{Q}_d(\mathbf{X}; \mathbf{V}^*)$ and \mathbf{C}_d as functions of v , it follows from (B.3) that

$$\partial \mathbf{C}_d / \partial v = \partial \mathbf{Q}_d(-\mathbf{C}^-_{d22} \mathbf{C}_{d21}; \mathbf{V}^*) / \partial v. \quad (\text{B.4})$$

To apply formula (B.4) when $v = n_\tau$, observe that

$$\partial \mathbf{Q}_d(\mathbf{X}; \mathbf{V}^*) / \partial n_\tau = \partial \mathbf{C}_{d11} / \partial n_\tau + (\partial \mathbf{C}'_{d21} / \partial n_\tau) \mathbf{X} + \mathbf{X}' (\partial \mathbf{C}_{d21} / \partial n_\tau) + \mathbf{X}' (\partial \mathbf{C}_{d22} / \partial n_\tau) \mathbf{X},$$

$$\partial \mathbf{C}_{d11} / \partial n_\tau = (\mathbf{T}_\tau - \mathbf{M}_d)' \mathbf{V}^* (\mathbf{T}_\tau - \mathbf{M}_d),$$

and similar formulas for $\partial \mathbf{C}_{d21} / \partial n_\tau$ and $\partial \mathbf{C}_{d22} / \partial n_\tau$. From these formulas and (B.4),

$$\partial \mathbf{C}_d / \partial n_\tau = \mathbf{Q}_{d,\tau} (-\mathbf{C}_{d22}^- \mathbf{C}_{d21}; \mathbf{V}^*). \tag{B.5}$$

Eq. (B.5) shows that \mathbf{C}_d is a non-decreasing function of each continuous variable, n_τ . Since $\Phi(\mathbf{U})$ is a non-decreasing function of $\mathbf{U} \in \mathcal{U}$, it follows that $\Phi(\mathbf{C}_d)$ is also a non-decreasing function of each continuous variable, n_τ :

$$\partial \Phi(\mathbf{C}_d) / \partial n_\tau \geq 0. \tag{B.6}$$

The increment, $\Delta \Phi$, in treatment information is approximately given by

$$\Delta \Phi \approx \sum_{\tau} (\partial \Phi(\mathbf{C}_d) / \partial n_\tau) \Delta n_\tau. \tag{B.7}$$

From the chain rule,

$$\begin{aligned} \partial \Phi / \partial n_\tau &= \sum_{i \geq j \geq 1}^t \Phi_{ij}(\mathbf{C}_d) \cdot \partial (\mathbf{C}_d)_{ij} / \partial n_\tau \\ &= (\frac{1}{2}) \text{Tr}[\mathbf{D}_\Phi(\mathbf{C}_d) \cdot \mathbf{Q}_{d,\tau} (-\mathbf{C}_{d22}^- \mathbf{C}_{d21}; \mathbf{V})]. \end{aligned} \tag{B.8}$$

If $\Delta n_\tau \geq 0$ for all τ , then from (B.6)–(B.8), maximizing $\Delta \Phi$ suggests the general allocation rule in Section 4.1.

Appendix C

C.1. Maximum likelihood estimators in adaptive and fixed designs

Schwabe (1995) notes that a classical design’s least-squares estimator for a fixed effect is the ML estimator both in a fixed and adaptive design. The ML estimators in our situation can be found, too, from the following result, due to Laska and Meisner.

A fixed design’s ML estimator of any unknown parameter is also its ML estimator in an adaptive design.

Proof. It suffices to show that the likelihood function of an adaptive design is identical, except for a factor which does not depend on unknown parameters, to the likelihood obtained by treating the design as fixed. Given design d , suppose that $f(\mathbf{y} | d; \Theta)$ is the density of \mathbf{Y} , a vector random variable, where Θ is a vector of unknown parameters. (In our situation, \mathbf{Y} is the p -dimensional vector of subject responses, Θ includes all the fixed effects and covariance matrix in model (2), and d can be specified by a design matrix.) Let $\mathbf{Y}_n, 1 \leq n \leq N$, denote the independent vector random variables of responses of the subjects and let \mathbf{y}_n denote an observed value of \mathbf{Y}_n . We now

calculate the likelihood function in an adaptive design. For the observation \mathbf{y}_1 the likelihood function is $r \cdot f(\mathbf{y}_1 | d_1; \Theta)$, where d_1 (the design consisting of the first treatment sequence) is randomly chosen with probability r that does not depend on Θ . The next design, d_2 (the design consisting of the treatment sequence assigned to the second subject by an adaptive allocation rule), is *fully determined* by \mathbf{y}_1 , and is the first adaptive design. Thus the probability, given $\mathbf{Y}_1 = \mathbf{y}_1$, of observing d_2 is 1, and the likelihood function of the pair $(\mathbf{y}_1, \mathbf{y}_2)$ is $r \cdot f(\mathbf{y}_1 | d_1; \Theta) \cdot f(\mathbf{y}_2 | d_2; \Theta)$. Similarly, the likelihood function of $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ in an adaptive design is $L_{\text{adaptive}} = r \prod_{n=1}^N f(\mathbf{y}_n | d_n; \Theta)$, where $d_n, 1 \leq n \leq N$, are the designs (treatment sequences), fully determined, for $n \geq 2$, by $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n-1})$, the preceding $(n-1)$ responses. On the other hand, in a fixed design in which $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ are observed and d_n is chosen (when the n th subject is randomized to a treatment sequence) with probability r that does not depend on Θ , the likelihood function of $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ is $L_{\text{fixed}} = r^N \prod_{n=1}^N f(\mathbf{y}_n | d_n; \Theta)$. Clearly, maximizing either L yields the same ML estimators of Θ .

C.2. The equations from which \mathbf{V}_N^* is computed

For a fixed design under a linear model, the equations satisfied by the ML estimators of the covariance matrix and fixed effects are given in Ratkowsky et al. (1993, p. 378). Our model (2), however, contains N fixed subject effects, which causes the number of unknowns in those equations to be very large. We adapt their approach to obtain equations satisfied by the ML estimators of \mathbf{V}^* and of the treatment and carryover effects. The ML estimate, \mathbf{V}_N^* , obtained from the equations is used in the simulations. Let

$$U_{nj} = Y_{nj} - \left(\sum_{k=1}^p Y_{nk} \right) / p - \left(\sum_{m=1}^N Y_{mj} \right) / N + \left(\sum_{m=1}^N \sum_{k=1}^p Y_{mk} \right) / pN, \quad 1 \leq n \leq N, \quad 1 \leq j \leq p. \tag{C.1}$$

Set $\mathbf{B}_p = \mathbf{I}_p - \mathbf{J}_p/p$ and $\mathbf{U}_n = (U_{n1}, \dots, U_{np})'$, a p -dimensional vector which is multivariate normal. From Martin and Eccleston (1998) and Mardia et al. (1979, p. 41), the joint (singular) density of all N vectors, $\mathbf{U}_n, 1 \leq n \leq N$, is, with $c = (2\pi)^{-(p-1)(N-1)/2}$,

$$c(\text{Tr}_{p-1}(\mathbf{V}^*))^{(N-1)/2} \exp \left(- \sum_{m,n=1}^N (\mathbf{u}_m - E\mathbf{U}_m)' \mathbf{V}^* (\mathbf{u}_n - E\mathbf{U}_n) (\delta_{mn} - 1/N) / 2 \right) \\ = c(\text{Tr}_{p-1}(\mathbf{V}^*))^{(N-1)/2} \exp \left(- \sum_{n=1}^N (\mathbf{u}_n - E\mathbf{U}_n)' \mathbf{V}^* (\mathbf{u}_n - E\mathbf{U}_n) / 2 \right). \tag{C.2}$$

Set $\mathbf{Y}_n = (Y_{n1}, \dots, Y_{np})'$, $\bar{\mathbf{Y}} = (\sum_{n=1}^N \mathbf{Y}_n) / N$, $\mathbf{X}_n = [\mathbf{T}_n; \tilde{\mathbf{T}}_n]$, $\bar{\mathbf{X}} = [\bar{\mathbf{T}}; \tilde{\bar{\mathbf{T}}}]$, where \mathbf{T}_n (respectively, $\tilde{\mathbf{T}}_n$) is the n th subject's $p \times t$ design matrix of treatment effects (respectively, carryover effects), $\bar{\mathbf{T}} = (\sum_{n=1}^N \mathbf{T}_n) / N$ and $\tilde{\bar{\mathbf{T}}} = (\sum_{n=1}^N \tilde{\mathbf{T}}_n) / N$. (In the notation of (14), $\mathbf{M}_d = \bar{\mathbf{T}}, \tilde{\mathbf{M}}_d = \tilde{\bar{\mathbf{T}}}$.) Since $\mathbf{U}_n = \mathbf{B}_p(\mathbf{Y}_n - \bar{\mathbf{Y}})$ and $E\mathbf{U}_n = \mathbf{B}_p(\mathbf{X}_n - \bar{\mathbf{X}})\boldsymbol{\beta}$, where

$\boldsymbol{\beta} = (\tau_1, \dots, \tau_t, \gamma_1, \dots, \gamma_t)'$, we obtain that the density (C.2) is

$$c(\text{Tr}_{p-1}(\mathbf{V}^*))^{(N-1)/2} \exp\left(-\sum_{n=1}^N (\mathbf{y}_n - \bar{\mathbf{y}} - (\mathbf{X}_n - \bar{\mathbf{X}})\boldsymbol{\beta})' \mathbf{V}^* (\mathbf{y}_n - \bar{\mathbf{y}} - (\mathbf{X}_n - \bar{\mathbf{X}})\boldsymbol{\beta})/2\right). \quad (\text{C.3})$$

Maximizing (C.3) over all $2t$ -dimensional vectors $\boldsymbol{\beta}$ and over all non-negative $p \times p$ matrices \mathbf{V}^* such that $\mathbf{V}^* \mathbf{1}_p = 0$ and $\text{rank}(\mathbf{V}^*) = p - 1$ leads to the following equations for $\hat{\mathbf{V}}^*$ and $\hat{\boldsymbol{\beta}}$, the ML estimators of \mathbf{V}^* and $\boldsymbol{\beta}$:

$$\sum_{n=1}^N (\mathbf{X}_n - \bar{\mathbf{X}})' \hat{\mathbf{V}}^* (\mathbf{X}_n - \bar{\mathbf{X}}) \hat{\boldsymbol{\beta}} = \sum_{n=1}^N (\mathbf{X}_n - \bar{\mathbf{X}})' \hat{\mathbf{V}}^* (\mathbf{y}_n - \bar{\mathbf{y}}) \quad (\text{C.4})$$

$$\mathbf{B}_p \left(\sum_{n=1}^N (\mathbf{y}_n - \bar{\mathbf{y}} - (\mathbf{X}_n - \bar{\mathbf{X}})\hat{\boldsymbol{\beta}})(\mathbf{y}_n - \bar{\mathbf{y}} - (\mathbf{X}_n - \bar{\mathbf{X}})\hat{\boldsymbol{\beta}})' \right) \mathbf{B}_p = (N - 1)(\hat{\mathbf{V}}^*)^+. \quad (\text{C.5})$$

Since $\mathbf{C}_d(\tau, \gamma)$, the joint information matrix of treatment and carryover effects, is given by $\mathbf{C}_d(\tau, \gamma) = \sum_{n=1}^N (\mathbf{X}_n - \bar{\mathbf{X}})' \mathbf{V}^* (\mathbf{X}_n - \bar{\mathbf{X}})$, Eq. (C.4) can be written as

$$\hat{\mathbf{C}}_d(\tau, \gamma) \hat{\boldsymbol{\beta}} = \sum_{n=1}^N (\mathbf{X}_n - \bar{\mathbf{X}})' \hat{\mathbf{V}}^* (\mathbf{X}_n - \bar{\mathbf{X}}), \quad (\text{C.6})$$

where $\hat{\mathbf{C}}_d(\tau, \gamma)$ denotes the ML estimator of $\mathbf{C}_d(\tau, \gamma)$. For a $\boldsymbol{\beta}$ for which

$$0 = \sum_{i=1}^t \tau_i = \sum_{i=1}^t \gamma_i, \quad (\text{C.7})$$

we then obtain

$$\hat{\boldsymbol{\beta}} = (\hat{\mathbf{C}}_d(\tau, \gamma))^+ \sum_{n=1}^N (\mathbf{X}_n - \bar{\mathbf{X}})' \hat{\mathbf{V}}^* (\mathbf{y}_n - \bar{\mathbf{y}}). \quad (\text{C.8})$$

Eqs. (C.5) and (C.8) determine the ML estimators based on N observations, $\hat{\mathbf{V}}_N^*$ and $\hat{\boldsymbol{\beta}}_N$, of \mathbf{V}^* and of a $\boldsymbol{\beta}$ satisfying (C.7). When the \mathbf{y}_i are numerically specified, these equations can be solved iteratively, and yield the current ML estimates, \mathbf{V}_N^* and $\boldsymbol{\beta}_N$, for use in the simulations. In the k th simulation, the first t components of $\boldsymbol{\beta}_N$ are the quantities $\tau_{i,N}(k)$, $1 \leq i \leq t$, mentioned in Section 3.4.

C.3. The Likelihood Ratio test of identical treatment effects

At the end of the experiment, the null hypothesis H (see Section 5) is tested by the LR statistic,

$$-2 \text{Ln } \mathcal{A} = -(N - 1) \text{Ln}(\text{Tr}_{p-1}[\hat{\mathbf{V}}_{H,N}^*] / \text{Tr}_{p-1}[\hat{\mathbf{V}}_N^*]), \quad (\text{C.9})$$

where $N = N_f$ and $\hat{\mathbf{V}}_{H,N}^*$ is obtained by solving the above equations for the hypothesis H with $\hat{\boldsymbol{\beta}}_H$, a $2t$ -dimensional vector whose first t components are zero, in place of $\hat{\boldsymbol{\beta}}$.

(Or, equivalently, one can solve the above equations with $\tilde{\mathbf{T}}_i$, $\bar{\mathbf{T}}$, $\hat{\mathbf{C}}_{d22}$ and $\hat{\gamma}$ (the ML estimator of the t -dimensional vector of carryover effects) in place, respectively, of \mathbf{X}_i , $\bar{\mathbf{X}}$, $\hat{\mathbf{C}}_d(\tau, \gamma)$ and $\hat{\boldsymbol{\beta}}$.) From Section C.1, Eq. (C.9) gives an LR test of H , for both fixed and adaptive designs. Since all the principal $(p-1) \times (p-1)$ determinants of \mathbf{V}^* are equal, $\text{Tr}_{p-1}[\mathbf{V}^*]/p = |\mathbf{V}_{p-1}^*|$, the determinant of the upper left $(p-1) \times (p-1)$ submatrix of \mathbf{V}^* , and we can write (C.9) in the simpler equivalent form,

$$-2\text{Ln } A = -(N-1) \text{Ln}(|\hat{\mathbf{V}}_{H,N,p-1}^*|/|\hat{\mathbf{V}}_{N,p-1}^*|). \tag{C.9}'$$

*C.4. Non-centrality parameters in the distribution of the LR statistic. Parameters in the distributions of the ML estimators of treatment effects, carryover effects and of \mathbf{V}^**

Under the standard Multivariate Linear Model (Muirhead, 1982, Chapter 10), the non-central distribution of the statistic used in an LR test of a linear hypothesis does not depend on the nuisance parameters in the model, but only on the true values of the parameters of interest, and on the true value of the covariance matrix, $\boldsymbol{\Sigma}$. Also, the distribution of the ML estimator of $\boldsymbol{\Sigma}$ depends only on $\boldsymbol{\Sigma}$; and the distribution of the ML estimators of the model's fixed effects depends only on the true values of the fixed effects and on $\boldsymbol{\Sigma}$. We now show that similar results are true for model (2) — which does not conform to the standard Multivariate Linear Model.

The non-central distribution of A , our LR statistic in (C.9), and the distributions of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{V}}^*$, clearly do not depend on the values of any of the N subject effects or on any of the p period effects, since these do not even appear in the density (C.3). Denote the density (C.3), A , the ML estimator of \mathbf{V}^* , and the ML estimators of treatment and carryover effects (normalized by (C.7)), by, respectively, $f(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N; \boldsymbol{\beta}, \mathbf{V}^*)$, $A(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$, $\hat{\mathbf{V}}^*(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ and $\hat{\boldsymbol{\beta}}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$. Then

$$f(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\alpha}, \mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\alpha}, \dots, \mathbf{y}_N - \mathbf{X}_N\boldsymbol{\alpha}; \boldsymbol{\beta}, \mathbf{V}^*) = f(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N; \boldsymbol{\alpha} + \boldsymbol{\beta}, \mathbf{V}^*), \tag{C.10a}$$

$$A(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\alpha}_H, \mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\alpha}_H, \dots, \mathbf{y}_N - \mathbf{X}_N\boldsymbol{\alpha}_H) = A(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N), \tag{C.10b}$$

$$\hat{\mathbf{V}}^*(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\alpha}, \mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\alpha}, \dots, \mathbf{y}_N - \mathbf{X}_N\boldsymbol{\alpha}) = \hat{\mathbf{V}}^*(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N), \tag{C.10c}$$

$$\hat{\boldsymbol{\beta}}(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\alpha}, \mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\alpha}, \dots, \mathbf{y}_N - \mathbf{X}_N\boldsymbol{\alpha}) = \hat{\boldsymbol{\beta}}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) + \boldsymbol{\alpha}, \tag{C.10d}$$

where $\boldsymbol{\alpha}$ is any $2t$ -dimensional vector, and $\boldsymbol{\alpha}_H = (\mathbf{0}_t; \boldsymbol{\delta})$, with $\mathbf{0}_t$ the t -dimensional vector of zeroes and $\boldsymbol{\delta}$ any t -dimensional vector. It follows from (C.10a)–(C.10c) (cf. Eaton, 1983, Chapter 7 or Muirhead, 1982, Chapter 6) that the non-central distribution of A (respectively, the distribution of $\hat{\mathbf{V}}^*$) does not depend on the carryover effects, but only on treatment effects, and on \mathbf{V}^* (respectively, only on \mathbf{V}^*). A short calculation, using (C.10a) and (C.10d), shows that the distribution of the ML estimator of a treatment or carryover effect depends only on the corresponding true treatment or carryover effect, and on \mathbf{V}^* .

References

- Armitage, P., 1975. *Sequential Medical Trials*. Blackwell, Oxford.
- Gebhardt, R., Heckendorff, H., 1983. Zur sequentiellen Versuchsplanung für das Regressionsproblem. *Math. Operationsforsch. Statist. Ser. Statist.* 14, 355–3662.
- Eaton, M.L., 1983. *Multivariate Statistics*. Wiley, New York.
- Kunert, J., 1985. Optimal repeated measurements designs for correlated observations and analysis by weighted least squares. *Biometrika* 72, 375–389.
- Kushner, H.B., 1997a. Optimality and efficiency of two-treatment repeated measurements designs. *Biometrika* 84, 455–468.
- Kushner, H.B., 1997b. 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.
- Laska, E., Meisner, M., 1985. A variational approach to optimal two-treatment crossover designs: application to carryover-effect models. *J. Amer. Statist. Assoc.* 80, 704–710.
- Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. *Multivariate Analysis*. Academic Press, New York.
- Martin, R.J., Eccleston, J.A., 1998. Variance-balanced change-over designs for dependent observations. *Biometrika* 85, 883–888.
- Matthews, J.N.S., 1987. Optimal crossover designs for the comparison of two treatments in the presence of carryover effects and auto-correlated errors. *Biometrika* 74, 311–320.
- Muirhead, R.J., 1982. *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- Pukelsheim, F., 1993. *Optimal Design of Experiments*. Wiley, New York.
- Ratkowsky, D.A., Evans, M.A., Alldredge, J.R., 1993. *Cross-over Experiments*. Marcel Dekker, Inc., New York.
- Rosenberger, W.F., Lachin, J.M., 1993. The use of response-adaptive designs in clinical trials. *Controlled Clin. Trials* 14, 471–484.
- Schwabe, R., 1987. On an adaptive design in regression. *Statistics* 18, 521–525.
- Schwabe, R., 1990. Adaptive designs for linear problems. In: Kasprzak, W., Weron, A. (Eds.), *Stochastic Methods in Experimental Sciences*. Scientific, Singapore, pp. 407–420.
- Schwabe, R., 1995. Adaptation in the design of experiments for linear models. In: Flournoy, N., Rosenberger, W.F. (Eds.), *Adaptive Designs. Selected Proceedings of a 1992 Joint AMS-IMS-SIAM Summer Conference, IMS Lectures Notes-Monograph Series, Vol. 25*, IMS, Hayward, CA, pp. 276–288.
- Silvey, S.D., 1980. *Optimal Design. Monographs on Applied Probability and Statistics*. Chapman & Hall, London.
- Wei, L.J., Durham, S., 1978. The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* 73, 840–843.
- Zelen, M., 1969. Play the winner rule and the controlled clinical trial. *J. Amer. Statist. Assoc.* 64, 131–146.