

# Universally Optimal Crossover Designs under Subject Dropout

BY WEI ZHENG

*Indiana University-Purdue University Indianapolis*

Subject dropout is very common in practical applications of crossover designs. However, there is very limited design literature taking this into account. Optimality results have not yet been well established due to the complexity of the problem. This paper establishes feasible, as well as necessary and sufficient conditions for a crossover design to be universally optimal in approximate design theory in the presence of subject dropout. These conditions are essentially linear equations with respect to proportions of all possible treatment sequences being applied to subjects and hence they can be easily solved. A general algorithm is proposed to derive exact designs which are shown to be efficient and robust.

**1. Introduction.** Crossover designs have been widely used in industry due to their cost effectiveness and statistical efficiency. They are applicable for experiments aiming to compare effects of different treatments by applying them to a number of subjects across several periods. The response observation is typically modeled by additive effects of subjects, periods, treatments, and the carryover effects of the treatment from the previous period. There has been tremendous amount of literature regarding the identification of optimal designs. See Hedayat and Afsarinejad (1978), Cheng and Wu (1980), Kunert (1984), Stufken (1991), Kushner (1997a, 1997b, 1998), Kunert and Martin (2000), Kunert and Stufken (2002), Hedayat and Yang (2003, 2004), and Hedayat and Zheng (2010) among others. For comprehensive reviews, see for instance Matthews (1988), Ratkowsky, Evans, and Alldredge (1992), Stufken (1996), Jones and Kenward (2003), Senn (2003), and Bose and Dey (2009).

An important issue regarding crossover designs is that subject may drop out of the study. As a result, the experiment will not be carried out as planned. Matthews (1988) commented this is one of the main concerns of crossover designs. Low, Lewis, and Prescott (1999) observed that “A dropout rate of between 5% and 10% is not uncommon and, in some areas, can be as high as 25%.” Meanwhile, a design, which is optimal or highly efficient in the absence of dropout, would become inefficient or even disconnected in the

---

*AMS 2000 subject classifications:* Primary 62K05; secondary 62J05

*Keywords and phrases:* Crossover Designs, Efficiency, Robustness, Subject Dropout, Universal Optimality

presence of subject dropout. Examples could be found in Godolphin (2004), Majumdar, Dean, and Lewis (2008) as well as Section 5 of this paper.

To conclude, it is very important to find optimal or efficient designs in the presence of subject dropout, yet there is very limited literature on this. Bose and Bagchi (2008) derived designs which are universally optimal for both direct and carryover effects for both the situation of no dropout and the situation that all subjects drop out after period  $q$  with  $q$  being judiciously chosen. Similar results are presented by Majumdar, Dean, and Lewis (2008). The latter restricted the comparison of designs within the subclass of uniformly balanced repeated measurement designs (UBRMDs), whose optimality property has been well recognized in literature for the situation of no dropout. For the second situation with any given  $q$ , they proposed type  $\mathcal{W}_q$  UBRMDs, which reduce the maximum loss of the information for parameters in terms of A-criterion as compared to general UBRMDs. Following the latter paper, Zhao and Majumdar (2012) further explored the special case when  $q$  is one less the number of periods and the numbers of treatments and periods are the same.

The previous three papers share two drawbacks: (i) The proposed designs exist only under very rare combinations of the numbers of subjects, periods, and treatments. See Section 5.1.2 for relevant discussions of the former paper. As for the other two papers, it is well known that the existence of UBRMDs is rare. (ii) The information regarding the mechanism of how subjects drop out was not taken into account.

To address the latter drawback, it is plausible to measure the performance of designs by taking the expectation of a regular optimality criterion with respect to the mechanism of subject dropout. Low, Lewis, and Prescott (1999) worked in this direction by using intensive computer programming. They concluded that when the Latin squares consisting of the design is more diverse, the resulting design performs better in terms of both efficiency and robustness. This argument is further supported by the comparison in Section 5. However, the case studies they provided fail to provide general guidance in identifying efficient designs. To serve this purpose, theoretical results are called for.

In this paper, we develop feasible equivalent conditions for a design to be universally optimal for direct treatment effects in approximate design theory under the same setup as that of Low, Lewis, and Prescott (1999). The equivalence holds for any probability distribution of subject dropout. The results can be easily modified to find optimal or highly efficient exact designs for any combination of the numbers of subjects, periods, and treatments. As a result, the two drawbacks are both addressed here.

The rest of the paper is organized as follows. Section 2 formulates the problem, introduces notations, and gives some preliminary results. Section 3 introduces necessary concepts in approximate design theory, proves the existence of universally optimal designs, and also gives necessary, sufficient, and equivalent conditions for universal optimality. Section 4 gives explicit and feasible forms of optimality conditions in terms of linear equations, which are built upon the preceding section. Section 5 further provides a general algorithm for deriving an optimal or efficient exact design for any combination of the numbers of subjects, periods, and treatments as well as any probability distribution of subject dropout. Besides, comparisons are made to designs in literature. Section 6 summarizes the results. Finally, some proofs are deferred to Section 7.

**2. Framework.** This section introduces the framework of the problem. Section 2.1 introduces the statistical model for the design problem and provides notations and assumptions necessary to the rest of the paper. Section 2.2 defines an ideal target function in finding a design, proposes a corresponding surrogate target function, and discusses the relationship between these two target functions. Section 2.3 provides some preliminary results as a preparation for the rest of the paper.

2.1. *Modeling and notations.* In a crossover design with  $p$  periods,  $t$  treatments, and  $n$  subjects the response is typically modeled as

$$(1) \quad Y_{dku} = \mu + \pi_k + \varsigma_u + \tau_{d(k,u)} + \gamma_{d(k-1,u)} + \varepsilon_{ku},$$

where  $\{\varepsilon_{ku}, 1 \leq k \leq p, 1 \leq u \leq n\}$  are independent with mean zero and variance  $\sigma^2$ . Here,  $Y_{dku}$  denotes the response from subject  $u$  in period  $k$  to which treatment  $d(k, u) \in \{1, 2, \dots, t\}$  was assigned by design  $d$ . Furthermore,  $\mu$  is the general mean,  $\pi_k$  is the  $k$ th period effect,  $\varsigma_u$  is the  $u$ th subject effect,  $\tau_{d(k,u)}$  is the (direct) treatment effect of treatment  $d(k, u)$ , and  $\gamma_{d(k-1,u)}$  is the carryover effect of treatment  $d(k-1, u)$  that subject  $u$  received in the previous period (by convention  $\gamma_{d(0,u)} = 0$ ).

Let  $G$  be a temporary object whose meaning differs from context to context. Then we define  $G'$  to represent the transpose of the matrix  $G$ ,  $G^-$  to represent a generalized inverse of the matrix  $G$ ,  $tr(G)$  to represent the trace of the matrix  $G$ , and  $pr^\perp$  to be a projection operator such that  $pr^\perp G = I - G(G'G)^-G'$ . For two square matrices of equal size,  $G_1$  and  $G_2$ ,  $G_1 \leq G_2$  means that  $G_2 - G_1$  is nonnegative definite. For a set  $G$ , the number of elements in the set is represented by  $|G|$ .

Besides,  $I_k$  is the  $k \times k$  identity matrix,  $1_k$  is the vector of length  $k$  with all its entries as 1,  $J_k = 1_k 1_k'$  is the square matrix with all its entries as 1.

We further define  $B_k = I_k - J_k/k$ ,  $B_{ij}^k$  to be the  $i \times j$  matrix with its upper left corner filled with the submatrix  $B_k$  while the remaining entries filled with 0, and  $B_i^k = B_{ii}^k$ . The notations of  $I_{ij}^k$  and  $I_i^k$  are defined in the same fashions as  $B_{ij}^k$  and  $B_i^k$ . Finally,  $\otimes$  represents the Kronecker product of two matrices. To make the problem resolvable, it is necessary to make two mild assumptions as follows.

**Assumption 1.** *Once a subject drops out of the study, the probability that the subject reenters the study is zero.*

By Assumption 1, we are able to define  $l_i$ ,  $1 \leq i \leq n$  to be the total number of periods that subject  $i$  stayed in the experiment. Further it is realistic in a large number of applications to assume

**Assumption 2.** *The dropping out mechanism is independent of the choice of design  $d$  as well as the outcome of the experiments. Moreover  $\{l_i, 1 \leq i \leq n\}$  are iid.*

By Assumption 2, we could define  $a_k$  to be the probability that  $l_i = k$ ,  $1 \leq k \leq p$ , and hence we are in place to define the following technical terms:

- $\vec{a} = (a_1, a_2, \dots, a_p)$
- $a_{jk} = \sum_{i=j}^k a_i$ ,  $1 \leq j \leq k \leq p$ . (Convention:  $a_{p+1,p} = 0$ )
- $m = \min\{k : a_k > 0\}$ .
- $\alpha_k = n^{-1} \left( (n+1)a_k + a_{1,k-1}^{n+1} - a_{1k}^{n+1} \right)$ ,  $1 \leq k \leq p$ .
- $\beta_k = a_k + a_{k+1,p}a_{1k}^n - a_{kp}a_{1,k-1}^n$ ,  $1 \leq k \leq p$ .
- $A = \sum_{k=1}^p \alpha_k B_p^k$ .
- $B = \sum_{k=1}^p \beta_k B_p^k$ .

**Definition 1.** *An experiment is said to be complete if there is no dropout.*

By definition the complete experiment is a special case in our framework and has been extensively studied in literature. Here, we aim to investigate desirable designs for any given dropout mechanism  $\vec{a}$ .

Notice that  $A$  and  $B$  are both nonnegative definite matrices. Since  $\beta_k \geq a_k + a_{k+1,p}a_{1k}^n - a_{kp}a_{1,k-1}^n = a_k(1 - a_{1k}^n) \geq 0$ , we have  $B \geq 0$ . By the mean value theorem one could show that  $\alpha_k \geq 0$  and hence  $A \geq 0$ . Note that  $a_k = 0$  implies  $\alpha_k = \beta_k = 0$ . Hence we have  $A = \sum_{k=m}^p \alpha_k B_p^k$  and  $B = \sum_{k=m}^p \beta_k B_p^k$ . The same representation will be adopted in the sequel whenever the summation over the period  $k$  is involved. Lastly, we should be aware of the differences and relationships among the matrices  $B_k$ ,  $B_p^k$ , and  $B$ .

2.2. *Optimality criteria.* Writing the  $np \times 1$  response vector as  $Y_d = (Y_{d11}, Y_{d21}, \dots, Y_{dp1}, Y_{d12}, \dots, Y_{dpn})'$ , Model (1) can be written as

$$(2) \quad Y_d = 1_{np}\mu + Z\boldsymbol{\pi} + U\boldsymbol{\varsigma} + T_d\boldsymbol{\tau} + F_d\boldsymbol{\gamma} + \varepsilon,$$

where  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)'$ ,  $\boldsymbol{\varsigma} = (\varsigma_1, \dots, \varsigma_n)'$ ,  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_t)'$ ,  $\boldsymbol{\gamma} = (\rho_1, \dots, \rho_t)'$ ,  $Z = 1_n \otimes I_p$ ,  $U = I_n \otimes 1_p$ , and  $T_d$  and  $F_d$  denote the treatment/subject and carryover/subject incidence matrices. Here  $\mathbb{E}\varepsilon = 0$  and  $\text{Var}(\varepsilon) = \sigma^2 I_{np}$ . For design  $d$  under a realization of experiment  $l = (l_1, \dots, l_n)'$ , the information matrix for the direct treatment effects  $\tau$  under Model (2) with  $\sigma^2 = 1$  is

$$\begin{aligned} C_d(\tau, l) &= (MT_d)' pr^\perp (MZ|MU|MF_d)(MT_d) \\ &= C_{d11}(l) - C_{d12}(l)[C_{d22}(l)]^{-1}C_{d21}(l), \end{aligned}$$

where

$$\begin{aligned} C_{d11}(l) &= T_d'OT_d, & C_{d12}(l) &= T_d'OF_d, \\ C_{d21}(l) &= C'_{d12}, & C_{d22}(l) &= F_d'OF_d, \\ M &= \text{diag}(I_{l_i, p}^{l_i}, i = 1, 2, \dots, n), \\ O &= M'pr^\perp(MZ|MU)M. \end{aligned}$$

Under a complete experiment, Kiefer (1975) defined a design to be universally optimal if it maximizes  $\Phi(C_d(\tau, p1_n))$  for any  $\Phi$  satisfying the following three conditions.

- (C.1)  $\Phi$  is concave.
- (C.2)  $\Phi(S'CS) = \Phi(C)$  for any permutation matrix  $S$ .
- (C.3)  $\Phi(bC)$  is nondecreasing in the scalar  $b > 0$ .

Optimality criteria defined by such a  $\Phi$  includes, but is not limited to, A, D, E, and T. See Kiefer (1975) and Yeh (1986) for instance. In the subject dropout setup there does not exist a design which maximizes  $\Phi(C_d(\tau, l))$  for all realizations of  $l$ . One reasonable target is to find a design which maximizes  $\phi_0(d|\Phi, \vec{a}) := \mathbb{E}_{\vec{a}}\Phi(C_d(\tau, l))$  for any  $\Phi$  satisfying the above three conditions. Here the expectation is taken over the probability space of  $l$  with parameter  $\vec{a}$ . For notational simplicity, we would omit the subscript  $\vec{a}$  for  $\mathbb{E}$  and the parameters  $\Phi$  and  $\vec{a}$  for  $\phi_0$  whenever it is clear from the context. So we have  $\phi_0(d) := \phi_0(d|\Phi, \vec{a}) = \mathbb{E}\Phi(C_d(\tau, l))$ .

There are two major difficulties in maximizing  $\phi_0(d)$  which make the problem intractable, if not impossible: (i)  $\Phi$  is a nonlinear function and hence the expectation would interact with the form of  $\Phi$ . (ii) Even when the dropout situation  $l$  is fixed, there is still a lack of tools to deal with the information matrix  $C_d(\tau, l)$  if subjects drop out at different periods under  $l$ .

In order to tackle these difficulties, we propose to replace the original target function of  $\phi_0(d)$  with the surrogate target function of  $\phi_1(d) = \Phi(C_d)$  where

$$(3) \quad \begin{aligned} C_d &= C_{d11} - C_{d12}C_{d22}^-C_{d21}, \\ C_{dij} &= \mathbb{E}C_{dij}(l), 1 \leq i, j \leq 2. \end{aligned}$$

It will be shown in Section 5 that this replacement is very successful in identifying highly efficient, if not optimal, designs for the criterion  $\phi_0(d)$ .

For  $i = 0$  or  $1$ , let  $d_i^*$  be an optimal design under  $\phi_i$ . Then define  $e_i(d) = \phi_i(d)/\phi_i(d_i^*)$  to be the efficiencies of  $d$  under  $\phi_i$ -criterion. Also we call  $g(d) = \phi_0(d)/\phi_1(d)$  to be the *gap* function between the two target functions for design  $d$ . Even though we are working on  $\phi_1$  instead of  $\phi_0$ , the  $\phi_0$ -efficiency  $e_0(d)$  could be bounded by  $e_1(d)g(d)$  as shown by Lemma 2.

**Lemma 1** (Pukelsheim (1993), pages 74-77). *The Schur complement of a matrix  $G \geq 0$  is a concave nondecreasing function of  $G$ .*

**Lemma 2.** *For any  $\Phi$ ,  $\vec{a}$ , and design  $d$ , we have  $\phi_0(d) \leq \phi_1(d)$ . Define  $e_i(d), i = 0, 1$  as the efficiency of design  $d$  under criterion  $\phi_i$ , we have*

$$e_0(d) \geq e_1(d)g(d).$$

*In particular, for any  $\phi_1$ -optimal design  $d$ , we have  $e_0(d) \geq g(d)$ .*

**Proof:** By Lemma 1 we have  $\mathbb{E}C_d(\tau, l) \leq C_d$ . Then we have

$$\begin{aligned} \phi_0(d) &= \mathbb{E}\Phi(C_d(\tau, l)) \\ (4) \quad &\leq \Phi(\mathbb{E}C_d(\tau, l)) \\ (5) \quad &\leq \Phi(C_d) \\ (6) \quad &= \phi_1(d). \end{aligned}$$

By (6) we have  $e_0(d) = \phi_0(d)/\phi_0(d_0^*) \geq \phi_0(d)/\phi_1(d_0^*) \geq \phi_0(d)/\phi_1(d_1^*) = e_1(d)g(d)$ .

◇

By (6) we have  $g(d) \leq 1$ , and hence  $g(d_1^*) \leq e_0(d_1^*)$ . That means if we could find a  $\phi_1$ -optimal design, then the value of the gap function  $g$  evaluated at this design serves as a lower bound of its  $\phi_0$ -efficiency. Inequalities (4) and (5) are essentially Jensen's type inequalities. The equalities therein both hold if the realization of subject dropout,  $l$ , is not random. When the variation in  $l$  is not very large, it would be plausible to work on the surrogate target

of maximizing  $\phi_1(C_d)$  instead of  $\phi_0(C_d)$  since the value of the gap function  $g$  would be close to unity. Note that a popular choice of  $\Phi$  is the trace of a matrix (T-criterion), for which the equality in (4) always holds.

When the experiment is complete, the necessary and sufficient conditions for  $\phi_1$ -universal optimality derived in Section 4 would be consistent with the results of Kushner (1997b). Note that the matrix  $C_d$  in (3) is no longer an information matrix for any design, as a result the ideas of proving the existence of universally optimal designs given by Theorem 3.4 of Kushner (1997b) are not applicable here. However, we found that similar results could be derived by direct manipulation on the matrix  $C_d$ . See Sections 3.2 and 3.3 for details. Moreover, since  $A \neq B$  in general, the arguments in deriving the linear equation as in proof of Theorem 5.3 of Kushner (1997b) are not applicable here either. For the approach of tackling this difficulty, see Section 4.1 for details.

### 2.3. Preliminary results.

**Lemma 3.** *Under Assumptions 1 and 2 we have  $C_{d11} = T_d'VT_d$ ,  $C_{d12} = T_d'VF_d$ , and  $C_{d22} = F_d'VF_d$  with*

$$(7) \quad \begin{aligned} V &= \sum_{k=m}^p (\alpha_k I_n - n^{-1} \beta_k J_n) \otimes B_p^k \\ &= I_n \otimes A - n^{-1} J_n \otimes B. \end{aligned}$$

Since  $B_p^1 = 0$  the  $m$  in (7) could be replaced by  $\max(m, 2)$ . A heuristic explanation for this observation is that when  $l_i = 1$  there is no information gained from this subject, because we rely on within subject comparison for treatments in crossover designs. When the experiment is complete we have  $\alpha_k = \beta_k = 0$  for all  $1 \leq k \leq p-1$  and  $\alpha_p = \beta_p = 1$ . In this case, we have the reduction of  $A = B = B_p$  and  $V = B_n \otimes B_p$ , for which the optimality problem has been extensively studied in literature.

**Corollary 1.** *Any design which is  $\phi_1$ -optimal with  $\Phi$  satisfying Conditions (C.1)–(C.3) under Model (1) is still optimal under the same criterion when the within subject covariance is of the form*

$$(8) \quad \Sigma = I_p + \eta 1_p' + 1_p \eta'.$$

*One special case is the compound symmetric covariance matrix, i.e.  $\Sigma = I_p + bJ_p$ . Here  $\eta$  is an arbitrary vector and  $b$  is an arbitrary real number.*

**Proof:** Let  $\Sigma_k$  be the  $k \times k$  upper left submatrix of  $\Sigma$  for  $1 \leq k \leq p$ . By direct calculation, we have

$$(9) \quad \Sigma_k^{-1} - \Sigma_k^{-1} J_k \Sigma_k^{-1} / 1'_k \Sigma_k^{-1} 1_k = B_k.$$

By following the same calculation as the proof of Lemma 3, the corollary is established in view of Equation (9). ◇

**Remark 1.** *The covariance matrix as in (8) is called a "type-H" matrix, see Huynh and Feldt (1970).*

**3.  $\phi_1$ -Universal Optimality.** This section explores the  $\phi_1$ -universal optimality in approximate design theory, where  $\phi_1$ -universal optimality is defined as follows.

**Definition 2.** *Given  $p, t, n$ , and a dropout mechanism  $\vec{a}$ , a design  $d$  is said to be  $\phi_1$ -universally optimal if  $d$  maximizes  $\phi_1(d)$  over all designs for any  $\Phi$  satisfying Conditions (C.1), (C.2), and (C.3).*

Section 3.1 introduces the ideas in approximate design theory as well as the concept of symmetric designs. Section 3.2 shows that a design would be  $\phi_1$ -universally optimal as long as its information matrix is of the form  $C_d = ny^* B_t / (t - 1)$  with  $y^*$  introduced by Equation (14). Section 3.3 shows that there always exists a symmetric design which satisfies this sufficient condition for  $\phi_1$ -universal optimality, and further by argument of Kiefer (1975) that this condition is also necessary for any design to be  $\phi_1$ -universally optimal. However, this condition is not immediately applicable for application. Section 4 gives an equivalent condition which is more readily applicable. Some relevant technical preparations are given in Sections 3.4 and 3.5.

**3.1. Approximate design theory and symmetric designs.** A design  $d$  with  $p$  periods,  $t$  treatments, and  $n$  subjects could be considered as the result of selecting  $n$  sequences with replacement from the collection of all possible  $t^p$  sequences, and this collection is denoted by  $\mathcal{S}$ . Let  $n_s$  be the number of replications of sequence  $s$  in the design, and define  $P_d = (p_s, s \in \mathcal{S})$  with  $p_s = n_s/n$ . When we ignore the ordering of the  $n$  sequences in the design, we have the one to one correspondence of  $d \leftrightarrow (n, P_d)$  with the restrictions of (i)  $\sum_{s \in \mathcal{S}} p_s = 1$ , (ii)  $p_s \geq 0$ , and (iii)  $np_s$  being an integer for all  $s$ . In approximate design theory, we only keep the first two restrictions and allow  $np_s$  not to be an integer.



Let  $\sigma$  be a permutation of symbols  $\{1, 2, \dots, t\}$ . For a sequence  $s = (t_1, \dots, t_p)$ , we define  $\sigma s = (\sigma(t_1), \dots, \sigma(t_p))$ . Then the design  $\sigma d$  is defined by  $P_{\sigma d} = (p_{\sigma^{-1}s}, s \in \mathcal{S})$ . The permutation matrix  $S_\sigma$  is the unique matrix satisfying  $T_{\sigma s} = T_s S_\sigma$  for all  $s \in \mathcal{S}$ . In the sequel we replace the subject index  $u$  by sequence index  $s$  whenever it is necessary.

A design  $d$  is said to be symmetric if  $P_d = P_{\sigma d}$ . Also we define symmetric blocks as  $\langle s \rangle = \{\sigma s, \sigma \in \mathcal{P}\}$  where  $\mathcal{P}$  is the collection of all possible  $t!$  permutations, i.e.  $|\mathcal{P}| = t!$ . We further define  $p_{\langle s \rangle} = \sum_{\tilde{s} \in \langle s \rangle} p_{\tilde{s}}$ . For a symmetric design, we have  $p_{\tilde{s}} = p_{\langle s \rangle} / |\langle s \rangle|$  for any  $\tilde{s} \in \langle s \rangle$ . Given  $p, t, n$ , a symmetric design  $d$  is uniquely determined by  $(p_{\langle s \rangle}, s \in \mathcal{S})$ , where  $s \in \mathcal{S}$  means that  $s$  runs through all distinct symmetric blocks contained in  $\mathcal{S}$ .

**3.2. A sufficient condition for  $\phi_1$ -universal optimality.** Denote by  $T_u$  (resp.  $F_u$ ) the  $p \times t$  submatrix of  $T$  (resp.  $F$ ) corresponding to the  $u$ th subject. Define  $\bar{T} = n^{-1} \sum_{u=1}^n T_u$ ,  $\hat{T}_u = T_u B_t$ , and  $\hat{T} = \bar{T} B_t$ . The notations  $\bar{F}$ ,  $\hat{F}_u$ , and  $\hat{F}$  are defined in the same way corresponding to carryover effects. Let  $\bar{C}_{dij} = \sum_{\sigma \in \mathcal{P}} S'_\sigma C_{dij} S_\sigma / |\mathcal{P}|$ ,  $1 \leq i, j \leq 2$ , and  $\bar{C}_d = \sum_{\sigma \in \mathcal{P}} S'_\sigma C_d S_\sigma / |\mathcal{P}|$ . Note that  $\bar{C}_{dij}$ ,  $1 \leq i, j \leq 2$ , are completely symmetric, also  $\bar{C}_{d11}$  and  $\bar{C}_{d12} = (\bar{C}_{d21})'$  have row and column sums as zero. Let  $\mathbb{I}$  be the indicator function. By Proposition 1 of Kunert and Martin (2000), we have

$$(10) \quad \bar{C}_d \leq \tilde{C}_d$$

$$(11) \quad = \left( c_{d11} - \frac{c_{d12}^2}{c_{d22}} \mathbb{I}_{[c_{d22} > 0]} \right) \frac{B_t}{t-1},$$

where

$$\tilde{C}_d = \bar{C}_{d11} - \bar{C}_{d12} (\bar{C}_{d22})^{-1} \bar{C}_{d21},$$

$$c_{dij} = \text{tr}(B_t \bar{C}_{dij} B_t)$$

$$= \text{tr}(B_t C_{dij} B_t), 1 \leq i, j \leq 2.$$

Define  $\hat{C}_{dij} = \sum_{u=1}^n \hat{C}_{uij}$ , where  $\hat{C}_{uij} = G'_i A G_j$  with  $G_1 = \hat{T}_u$  and  $G_2 = \hat{F}_u$ . Since  $B \geq 0$ , we have

$$(B_t C_{dij} B_t)_{1 \leq i, j \leq 2} = \begin{pmatrix} \hat{C}_{d11} & \hat{C}_{d12} \\ \hat{C}_{d21} & \hat{C}_{d22} \end{pmatrix} - n \begin{pmatrix} \hat{T}'_d \\ \hat{F}'_d \end{pmatrix} B \begin{pmatrix} \hat{T}_d & \hat{F}_d \end{pmatrix}$$

$$(12) \quad \leq \begin{pmatrix} \hat{C}_{dij} \end{pmatrix}_{1 \leq i, j \leq 2},$$

Define  $q_{dij} = \text{tr}(\hat{C}_{dij})$  and  $q_{uij} = \text{tr}(\hat{C}_{uij})$ . Then we have  $q_{dij} = \sum_{u=1}^n q_{uij}$ . It is easy to see that  $q_{u22} > 0$  and hence  $q_{d22} > 0$ , which allow us to define

$$q_d^* = q_{d11} - \frac{q_{d12}^2}{q_{d22}}.$$

By (12) we have

$$\begin{pmatrix} c_{d11} & c_{d12} \\ c_{d21} & c_{d22} \end{pmatrix} \leq \begin{pmatrix} q_{d11} & q_{d12} \\ q_{d21} & q_{d22} \end{pmatrix},$$

and then by Lemma 1 we have

$$(13) \quad c_{d11} - \frac{c_{d12}^2}{c_{d22}} \mathbb{I}_{[c_{d22} > 0]} \leq q_d^*,$$

with the equality holds when  $\hat{T}_d = \hat{F}_d = 0$ . The latter is achieved by designs which are uniform on periods. To introduce the following theorem, we define

$$(14) \quad y^* = \frac{1}{n} \max_d q_d^*.$$

**Theorem 1.** *If  $C_d = ny^*B_t/(t-1)$  with  $y^*$  defined in (14), then the design  $d$  is  $\phi_1$ -universally optimal.*

**Proof:** By Conditions (C.1) and (C.2) of  $\Phi$  we have

$$(15) \quad \Phi(C_d) \leq \Phi(\bar{C}_d),$$

with the equality holds if  $C_d$  is completely symmetric, i.e.  $C_d = \text{tr}(C_d)B_t/(t-1)$  since  $C_d$  has row and column sums as zero. The theorem is proved in view of (10), (11), (13), (14), (15), and Condition (C.3).

◇

**3.3. Existence and equivalence.** Theorem 1 provides a sufficient condition for a design to be  $\phi_1$ -universally optimal. A natural question is the following: does there exist such a design? This section gives a positive answer as well as its corresponding implications.

**Theorem 2.** *For any symmetric design we have*

1.  $C_d$  is completely symmetric
2.  $\text{tr}(C_d) = q_d^*$
3. *Given any design  $d$  there always exist a corresponding symmetric design which has the same value of  $q_d^*$ .*

**Remark 2.** *Note that Theorem 2 does not hold if we replace  $C_d$  therein by  $C_d(\tau, l)$ . Hence the argument cannot be applied to  $\Phi(C_d(\tau, l))$  directly. This is why we work on  $\phi_1$  instead of  $\phi_0$  directly.*

**Corollary 2.** (i) *There exists a symmetric  $\phi_1$ -universally optimal design  $d$  with*

$$(16) \quad C_d = \frac{ny^*B_t}{t-1}.$$

(ii) *If a design  $d$  is  $\phi_1$ -universally optimal (or  $\phi_1$ -optimal with  $\Phi$  strictly concave or increasing) then we have (16).*

**Proof:** (i) is proved by Theorems 1 and 2. (ii) is proved by (i) and the Remark of Kiefer's (1975) Proposition 1 .

◇

3.4. *A necessary condition for  $\phi_1$ -universal optimality.* In this section we give a necessary condition for a design to be  $\phi_1$ -universally optimal and define quantities that will be useful for presenting the necessary and sufficient conditions for  $\phi_1$ -universal optimality in Section 4. Now define the function  $q_s(x) = q_{s11} + 2q_{s12}x + q_{s22}x^2$  and  $q_d(x) = q_{d11} + 2q_{d12}x + q_{d22}x^2$ . Since  $q_{dij} = \sum_{u=1}^n q_{uij} = n \sum_{s \in \mathcal{S}} p_s q_{sij}$  we have

$$(17) \quad q_d(x) = n \sum_{s \in \mathcal{S}} p_s q_s(x).$$

Since  $q_{d22} > 0$ , by direct calculation we have

$$(18) \quad \begin{aligned} q_d^* &= \min_x q_d(x) \\ &= n \min_x \sum_{s \in \mathcal{S}} p_s q_s(x). \end{aligned}$$

By (14) and (18) we have

$$(19) \quad y^* = \max_P \min_x \sum_{s \in \mathcal{S}} p_s q_s(x).$$

Let  $d^*$  be a design which maximizes  $q_d^*$ . By (17), (18), and (19) we have  $\min_x q_{d^*}(x) = q_{d^*}^* = ny^*$ . Since  $q_{d22} > 0$  the equation  $q_{d^*}(x) = ny^*$  has a unique solution which is denoted by  $x^*$ . Define

$$\mathcal{T} = \{s \in \mathcal{S} : y^* = q_s(x^*)\}.$$

Lemma 4 shows that any universally optimal design is supported on  $\mathcal{T}$ .

**Lemma 4.** *If a design  $d$  is  $\phi_1$ -universally optimal (or  $\phi_1$ -optimal with  $\Phi$  strictly concave or increasing) then we have*

$$p_s = 0, \quad s \notin \mathcal{T}.$$

**Proof:** By Corollary 2, we have  $tr(C_d) = ny^*$  and  $C_d = \overline{C}_d$ . By (10), (11), and (13) we have  $tr(\overline{C}_d) \leq q_d^*$ . The theorem is proved in view of (14) and Section 4.4 of Kushner (1997b). ◇

3.5. *Determination of  $x^*$ ,  $y^*$ , and  $\mathcal{T}$ .* For a sequence  $s = (t_1, t_2, \dots, t_p)$ , define  $s_k = (t_1, \dots, t_k)$  to be the first  $k$  periods of  $s$ . Particularly, we have  $s = s_p$ . For  $1 \leq k \leq p$  and  $1 \leq i \leq t$ , we define the treatment/sequence index  $f_{s_k, i} = \sum_{j=1}^k 1_{t_j=i}$ . To introduce the following theorem, we define two special symmetric blocks. The symmetry block  $\langle di \rangle$  consists of all sequences having distinct treatments in the  $p$  periods. The symmetry block  $\langle re \rangle$  consists of all sequences having distinct treatments in the first  $p-1$  periods, with the treatment in period  $p-1$  repeating in period  $p$ .

**Theorem 3.** *For any integer  $k > t$ , define  $z_k$  and  $r_k$  to be integers satisfying  $k = z_k t + r_k$  and  $0 < r_k \leq t$ .*

(i) *If  $m > t$  and*

$$(20) \quad \sum_{k=m}^p \alpha_k [k(mt - t^2 + 1 - k) + t - r_k(t - r_k + 1)] \geq 0,$$

*then*

$$\begin{aligned} x^* &= 0, \\ y^* &= \sum_{k=m}^p \alpha_k [k(1 - 1/t) - r_k(t - r_k)/pt], \\ \mathcal{T} &= \{s : f_{s_k, i} = z_k \text{ or } z_k + 1, 1 \leq i \leq t, m \leq k \leq p\}. \end{aligned}$$

(ii) *If  $p \leq t$  and*

$$(21) \quad \sum_{k=m}^{p-1} \alpha_k (k-1)(p + 1/t - k) \leq \alpha_p [(p-1)^2 - (1 + 1/t)p + 1/t],$$

*then*

$$x^* = 1/(p-1),$$

$$y^* = \sum_{k=m}^p \alpha_k (k-1) \left( 1 - \frac{2p-1-k+1/t}{k(p-1)^2} \right),$$

$$\mathcal{T} = \langle re \rangle \cup \langle di \rangle.$$

When the two sides of (21) are equal, we have  $\mathcal{T} = \langle re \rangle$ .

(iii) Let

$$x_0 = \frac{\sum_{k=m}^{p-1} \alpha_k (k-1)}{\sum_{k=m}^p \alpha_k (k-1)(k-1-1/t)}.$$

If  $(p-1)^{-1} < x_0 < (p-2)^{-1}$  then

$$x^* = x_0,$$

$$y^* = \sum_{k=m}^p \alpha_k (k-1) (1 - 1/k - 1/kt) x_0^2 - 2 \sum_{k=m}^{p-1} \alpha_k (1 - 1/k) x_0$$

$$+ \sum_{k=m}^p \alpha_k (k-1) - 2/p,$$

$$\mathcal{T} = \langle re \rangle.$$

**Remark 3.** Under complete experiment, Theorem 3 (i) applies to the case  $p > t$  and Theorem 3 (ii) applies to the case  $p \leq t$ . Actually Theorem 3 (i), (ii) reduce to Theorem 1 of Kushner (1998). One can extrapolate by continuity that Theorem 3 (i), (ii) cover the cases when the dropout issue is not very serious.

**Remark 4.** When  $m = p-1$ , we would also discuss parts (i) and (ii) of Theorem 3. For (i), a sufficient condition for (20) is  $p > t+3$ . For (ii), Inequality (21) simplifies to

$$(22) \quad \alpha_p \geq \frac{(p-2)(1+1/t)}{(p-1)^2 - 2 - 1/t}.$$

The right hand side of (22) mainly depends on  $p$  and it will become very small for large  $p$ . Particularly, a sufficient condition for (22) is

$$a_p \geq \frac{n}{n+1} \frac{(p-2)(1+1/t)}{(p-1)^2 - 2 - 1/t} + \frac{1}{n+1}.$$

**4. Linear Equations for  $\phi_1$ -universal Optimality.** Built upon the results of Section 3, this section provides feasible equivalent conditions in approximate design theory for  $\phi_1$ -universal optimality.

4.1. *Equations for general designs.* Recall that  $\hat{T}_u = T_u B_t$  and  $\hat{F}_u = F_u B_t$ , then we define

$$(23) \quad \begin{aligned} \check{C}_d &= \check{C}_{d11} - \check{C}_{d12} \check{C}_{d22}^- \check{C}_{d21}, \\ \check{C}_{dij} &= \sum_{u=1}^n \check{C}_{uij}, 1 \leq i, j \leq 2, \end{aligned}$$

where

$$\begin{aligned} \check{C}_{u11} &= T'_u(A-B)T_u + \hat{T}'_u B \hat{T}_u, & \check{C}_{u12} &= T'_u(A-B)F_u + \hat{T}'_u B \hat{F}_u, \\ \check{C}_{u21} &= \check{C}'_{u12}, & \check{C}_{u22} &= F'_u(A-B)F_u + \hat{F}'_u B \hat{F}_u. \end{aligned}$$

We shall replace  $\check{C}_{uij}$  with  $\check{C}_{sij}$  in emphasizing Sequence  $s$  instead of Subject  $u$  of a design. By direct calculation we have

$$(24) \quad \check{C}_{dij} = C_{dij} + nG'_i B G_j, 1 \leq i, j \leq 2,$$

where  $G_1 = \hat{T}_d$  and  $G_2 = \hat{F}_d$ . The following lemma is crucial for the proof of Theorem 4.

**Lemma 5.** *If  $d$  is  $\phi_1$ -universally optimal (or  $\phi_1$ -optimal with  $\Phi$  strictly concave or increasing), we have  $C_d = \check{C}_d = ny^* B_t / (t-1)$ .*

**Proof:** By (24) and Lemma 1 we have

$$(25) \quad C_d \leq \check{C}_d.$$

By Corollary 2 (ii) we have

$$(26) \quad C_d = ny^* B_t / (t-1).$$

Let  $\bar{d}$  be the symmetrized version of design  $d$  as defined by (48), then by (23) we have

$$(27) \quad \sum_{\sigma \in \mathcal{P}} S'_\sigma \check{C}_{dij} S_\sigma / |\mathcal{P}| = \check{C}_{\bar{d}ij}.$$

Again by (24) we have  $\check{C}_{dij} = G'_i \Lambda G_j$  with  $G_1 = (\hat{T}'_d, T'_d)'$ ,  $G_2 = (\hat{F}'_d, F'_d)'$ , and

$$\Lambda = \begin{pmatrix} nB & 0 \\ 0 & V \end{pmatrix}.$$

Since  $\Lambda \geq 0$  we have by Proposition 1 of Kunert and Martin (2000) that

$$(28) \quad \sum_{\sigma \in \mathcal{P}} S'_\sigma \check{C}_d S_\sigma / |\mathcal{P}| \leq \check{C}_{\bar{d}},$$

in view of (27). Since  $\hat{T}_{\bar{d}} = \hat{F}_{\bar{d}} = 0$  for the symmetric design  $\bar{d}$ , we have

$$(29) \quad \check{C}_{\bar{d}} = C_{\bar{d}}.$$

Combining (25), (26), (27), (28), and (29), we have

$$\begin{aligned} \frac{ny^*}{t-1} B_t = C_d &= \bar{C}_d \\ &\leq \sum_{\sigma \in \mathcal{P}} S'_\sigma \check{C}_d S_\sigma / |\mathcal{P}| \\ &\leq C_{\bar{d}}. \end{aligned}$$

Hence we have  $C_{\bar{d}} = ny^* B_t / (t-1)$  in view of Corollary 2 and thus

$$\sum_{\sigma \in \mathcal{P}} S'_\sigma \check{C}_d S_\sigma / |\mathcal{P}| = ny^* B_t / (t-1),$$

which in turn yields

$$(30) \quad \text{tr}(\check{C}_d) = ny^*.$$

The lemma is now proved in view of (25) and (30). ◇

**Theorem 4.** *A design  $d$  is  $\phi_1$ -universally optimal (or  $\phi_1$ -optimal with  $\Phi$  strictly concave or increasing) if and only if*

$$(31) \quad \sum_{s \in \mathcal{T}} p_s [\check{C}_{s11} + x^* \check{C}_{s12} B_t] = \frac{y^*}{t-1} B_t,$$

$$(32) \quad \sum_{s \in \mathcal{T}} p_s [\check{C}_{s21} + x^* \check{C}_{s22} B_t] = 0,$$

$$(33) \quad \sum_{s \in \mathcal{T}} p_s B(\hat{T}_s + x^* \hat{F}_s) = 0,$$

$$(34) \quad \sum_{s \in \mathcal{T}} p_s = 1,$$

$$(35) \quad p_s = 0, \quad s \notin \mathcal{T}.$$

Based on Theorem 1 and Corollary 2, (16) is also a necessary and sufficient condition for  $\phi_1$ -universal optimality. However, (16) is not directly applicable for identifying designs. Note that the conditions in Theorem 4 are merely linear equations systems for  $p_s$ , and hence can be easily implemented to derive exact designs. See Section 5.

4.2. *Equations for symmetric designs.* Note that  $q_s(x)$  is invariant to treatment permutation, i.e.

$$(36) \quad q_s(x) = q_{\sigma s}(x).$$

Combining Theorem 4.5 of Kushner (1997b), Theorem 2, Corollary 2, Lemma 4, and Equation (36), we have

**Theorem 5.** *A symmetric design is  $\phi_1$ -universally optimal if*

$$\begin{aligned} \sum_{s \in \mathcal{T}} p_{\langle s \rangle} q'_s(x^*) &= 0, \\ \sum_{s \in \mathcal{T}} p_{\langle s \rangle} &= 1, \\ p_s &= 0, \quad s \notin \mathcal{T}, \end{aligned}$$

where  $q'_s(x)$  is the derivative of  $q_s(x)$  with respect to  $x$ .

**5. Exact Designs.** This section gives algorithms to identify efficient exact designs based on the optimality equations in Section 4. Results are compared to designs proposed in literature. For the matrix  $C_d(\tau, l)$ , denote its eigenvalues by  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_t$ . We define the criteria of A, D, E, and T as

- $\Phi_A(C_d(\tau, l)) = (t-1)(n \sum_{i=2}^t \lambda_i^{-1})^{-1}$ . ( $\lambda_2 = 0$  implies  $\Phi_A(C_d(\tau, l)) = 0$ )
- $\Phi_D(C_d(\tau, l)) = n^{-1}(\prod_{i=2}^t \lambda_i)^{1/(t-1)}$
- $\Phi_E(C_d(\tau, l)) = n^{-1} \lambda_2$
- $\Phi_T(C_d(\tau, l)) = [n(t-1)]^{-1}(\sum_{i=2}^t \lambda_i)$

Section 5.1 provides an algorithm to derive exact designs for general configurations of  $p, t, n$ . Section 5.2 illustrates how to derive symmetric designs by straightforward calculations. In utilizing Lemma 2,  $e_1(d)$  is further bounded by  $\tilde{e}_1 = \phi_1(d)/\phi_1(\tilde{d})$ , where  $\tilde{d}$  is a  $\phi_1$ -optimal design in asymptotic design theory which may not necessarily exist as an exact design. Thus the function  $\ell(d) = \tilde{e}_1(d)g(d)$  serves as a feasible lower bound of  $e_0(d)$ .



5.1. *General exact designs.* This section gives an algorithm to derive efficient exact optimal designs for any given configuration of  $p, t, n$  and compare them to designs in literature. Note that the latter designs are proposed for judiciously chosen  $p, t, n$  while our algorithm works for any configuration of  $p, t, n$ . Even under these chosen circumstances our designs are still shown to be more efficient and robust. By Theorem 4 we have

**Corollary 3.** *A design  $d$  is  $\phi_1$ -universally optimal (or  $\phi_1$ -optimal with  $\Phi$  strictly concave or increasing) if and only if*

$$(37) \quad \sum_{s \in \mathcal{T}} n_s [\check{C}_{s11} + x^* \check{C}_{s12} B_t] = \frac{ny^*}{t-1} B_t,$$

$$(38) \quad \sum_{s \in \mathcal{T}} n_s [\check{C}_{s21} + x^* \check{C}_{s22} B_t] = 0,$$

$$(39) \quad \sum_{s \in \mathcal{T}} n_s B(\hat{T}_s + x^* \hat{F}_s) = 0,$$

$$(40) \quad \sum_{s \in \mathcal{T}} n_s = n,$$

$$(41) \quad n_s = 0, \quad s \notin \mathcal{T}.$$

Note that an exact design satisfying Equations (37)-(41) does not necessarily exist due to the discrete nature of the problem, especially when the dropout mechanism is arbitrary. However, as shown by the following examples, it is plausible to find a design which is as *close* to satisfying Equations (37)-(41) as possible. Specifically, let  $N_{\mathcal{T}} = \{n_s, s \in \mathcal{T}\}'$ , then Equations (37)-(39) could be written in a matrix form as

$$X_{\mathcal{T}} N_{\mathcal{T}} = Y_{\mathcal{T}},$$

with  $X_{\mathcal{T}}$  and  $Y_{\mathcal{T}}$  uniquely determined by Equations (37)-(41) and the ordering of the  $n_s$  in the vector  $N_{\mathcal{T}}$ . To find an efficient design for an arbitrarily given  $n$ , one could choose a design which

Minimizes

$$(42) \quad \|X_{\mathcal{T}} N_{\mathcal{T}} - Y_{\mathcal{T}}\|,$$

subject to

$$1'_{|\mathcal{T}|} N_{\mathcal{T}} = n.$$

Here  $\|\cdot\|$  is a norm for a vector. For all subsequent examples in this section, we take  $\|\cdot\|$  to be the Euclidean norm. Then solving for (42) is straight forward by utilizing integer optimization packages/software. Note that the

$\Phi$	$\phi_0(d_1)$	$V_\Phi(d_1)$	$\tilde{e}_1(d_1)$	$g(d_1)$	$\ell(d_1)$
A	0.6646345	0.07223834	0.9558432	0.9592851	0.9169261
D	0.6747419	0.06776632	0.9603310	0.9693223	0.9308702
E	0.5528575	0.09039916	0.8960473	0.8512042	0.7627192
T	0.6848634	0.06334558	0.9650531	0.9790485	0.9448338

TABLE 1

Performance of  $d_1$  under  $\vec{a} = (0, 0, 1/2, 1/2)$ .

$\Phi$	$\phi_0(d_2)$	$V_\Phi(d_2)$	$\tilde{e}_1(d_2)$	$g(d_2)$	$\ell(d_2)$
A	0.7058735	0.05266523	0.9989759	0.9748175	0.9738192
D	0.7094851	0.05129209	0.9991830	0.9796020	0.9788017
E	0.6337475	0.06979073	0.9848636	0.8877519	0.8743145
T	0.7130567	0.05005383	0.9993922	0.9843273	0.9837291

TABLE 2

Performance of  $d_2$  under  $\vec{a} = (0, 0, 1/2, 1/2)$ .

computational complexity of the above minimization problem depends on  $|\mathcal{T}|$ , which in turn depends on  $p$  and  $t$  only.

Besides maximizing the expectation  $\phi_0(d) = \mathbb{E}\Phi(C_d(\tau, l))$ , one might also be interested in minimizing the variance  $V_\Phi(d) = \text{Var}(\Phi(C_d(\tau, l)))$  to achieve robustness. To compare two designs under these two functions, we define  $\phi_0(d, d') = \phi_0(d)/\phi_0(d')$  and  $V_\Phi(d, d') = V_\Phi(d)/V_\Phi(d')$ .

5.1.1. *Comparisons to designs of Low, Lewis, and Prescott (1999)*. The setup and target of Low, Lewis, and Prescott (1999) are the same as in this paper. However, they searched all combinations of Latin squares for the special cases of  $p = t = 4, n = 16$  and  $p = t = 4, n = 24$  only.

When  $p = t = 4$  and  $n = 16$ , they proposed a design as shown by Fig 1(b) therein, which is said to be  $d_1$  here. By Algorithm (42), the dropout mechanism  $\vec{a} = (0, 0, 1/2, 1/2)$  yields  $d_2$ :

$$d_2 : \begin{array}{cccccccccccccccc} 2 & 1 & 2 & 3 & 3 & 4 & 3 & 2 & 1 & 1 & 1 & 2 & 4 & 4 & 4 & 3 \\ 4 & 4 & 3 & 4 & 1 & 1 & 2 & 1 & 2 & 2 & 3 & 4 & 3 & 3 & 2 & 1 \\ 3 & 2 & 1 & 1 & 2 & 3 & 4 & 4 & 3 & 3 & 4 & 1 & 2 & 2 & 1 & 4 \\ 3 & 2 & 1 & 1 & 2 & 3 & 4 & 4 & 4 & 4 & 2 & 3 & 1 & 1 & 3 & 2 \end{array} .$$

Tables 1 and 2 summarize the performances of designs  $d_1$  and  $d_2$  under criteria of  $A$ ,  $D$ ,  $E$ , and  $T$ . Since  $e_0(d) \geq \ell(d) = \tilde{e}_1(d)g(d)$ , a design  $d$  would be  $\phi_0$ -efficient if both  $\tilde{e}_1(d)$  and  $g(d)$  are close to unity. Algorithm (42) focuses on  $e_1(d)$  and provides a satisfactory solution in view of the column of  $\tilde{e}_1$  in Table 2. We observe that the values of  $g$  in both of these tables are very close to unity except for E-criterion. Notice that the values of gap

$\Phi$	$\phi_0(d_4)$	$V_\Phi(d_4)$	$\tilde{e}_1(d_4)$	$g(d_4)$	$\ell(d_4)$	$\phi_0(d_4, d_3)$	$V_\Phi(d_4, d_3)$
A	0.6791	0.0526	0.999983	0.9777	0.9777	1.0112	0.9705
D	0.6822	0.0516	0.999983	0.9822	0.9821	1.0115	0.9631
E	0.6118	0.0648	0.999979	0.8809	0.8809	1.0089	1.0386
T	0.6852	0.0506	0.999983	0.9866	0.9869	1.0118	0.9562

TABLE 3

Performance of  $d_3$  and  $d_4$  under  $\vec{a} = (0, 1/10, 2/5, 1/2)$ .

function  $g$  for T-criterion are always the largest among all criteria, which is due to the linearity of T-criterion.

In comparison,  $d_2$  is more efficient and robust than  $d_1$  under all criteria in view of the columns of  $\phi_0$  and  $V_\Phi$  respectively. A lesson from the latter is that a design with a more diverse composition of sequences is generally more robust. Here in  $d_2$ , only the sequences of 1234 and 4321 appear twice while each of the remaining sequences appears only once. Low, Lewis, and Prescott (1999) had similar observations.

We now consider the performance of a design obtained by Algorithm (42) for dropout mechanisms of the form  $\vec{a} = (0, 0, \theta, 1 - \theta)$ ,  $0 < \theta < 1$ . By heuristic arguments in Section 2.2, the value of gap function  $g$  would be smaller if there is larger variability in  $l$ . This is supported by the U-shape curve of  $\ell(d)$  in Figure 1(b). From Figure 1(a), we see that the efficiency of  $d_1$  has a reverse relationship with the value of  $\theta$ . Figure 1(c) shows that the advantage of our algorithm against  $d_1$  is more obvious when there is large chance of dropout. This means that our algorithm succeeded in adapting the choice of designs to different dropout mechanisms. Figure 1(d) shows that the design by our algorithm is also more robust than  $d_1$  against the randomness of subject dropout.

When  $p = t = 4$  and  $n = 24$ , Low, Lewis, and Prescott (1999) proposed a design which consists of two copies of three distinct  $4 \times 4$  Latin squares, which is denoted by  $d_3$  here. When  $\vec{a} = (0, 1/10, 2/5, 1/2)$  our algorithm yields  $d_4$  which consists of one copy of the first twelve sequences and two copies of the last six sequences of (43). According to the last two columns of Table 3,  $d_4$  outperforms  $d_3$  in terms of both efficiency and robustness with the exception for the robustness under E-criterion.

$$(43) d_4 : \begin{array}{cccccccccccc|cccccc} 2 & 2 & 2 & 3 & 4 & 4 & 2 & 3 & 3 & 4 & 4 & 3 & 3 & 4 & 2 & 1 & 1 & 1 \\ 3 & 4 & 3 & 4 & 3 & 2 & 3 & 4 & 4 & 2 & 2 & 2 & 1 & 1 & 1 & 2 & 4 & 3 \\ 4 & 3 & 1 & 1 & 2 & 1 & 1 & 1 & 2 & 1 & 3 & 4 & 2 & 3 & 4 & 4 & 3 & 2 \\ 4 & 3 & 1 & 1 & 2 & 1 & 4 & 2 & 1 & 3 & 1 & 1 & 2 & 3 & 4 & 3 & 2 & 4 \end{array} \times 2.$$

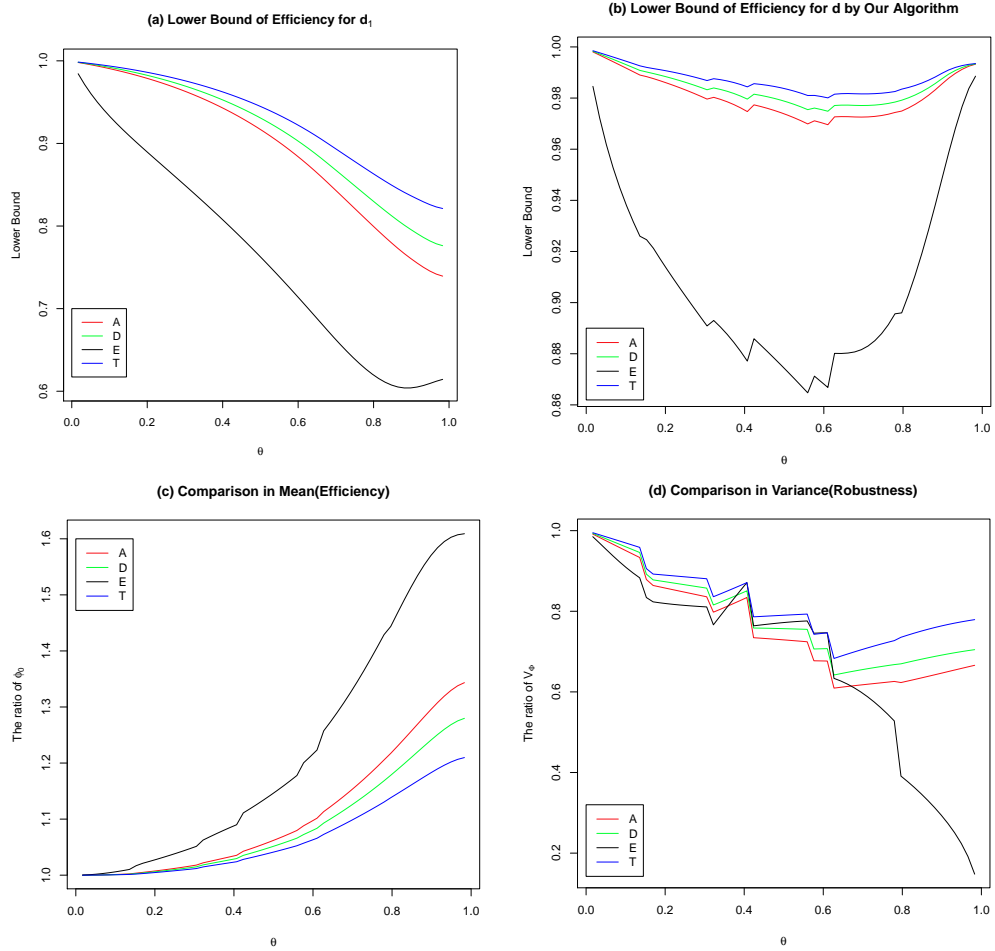


FIG 1. The letters A, D, E, and T represent the choice of criteria function  $\Phi$ . (a) The lower bound of efficiency  $\ell(d_1)$  for  $a = (0, 0, \theta, 1 - \theta)$  with  $\theta \in (0, 1)$ . (b) The lower bound of efficiency  $\ell(d)$  with  $d$  obtained by Algorithm (4.2). Particularly  $\theta = 1/2$  implies  $d = d_2$ . (c) The ratio of mean:  $\phi_0(d)/\phi_0(d_1)$ . (d) The ratio of variance:  $V_\Phi(d)/V_\Phi(d_1)$ .

$\Phi$	$\phi_0(d_6)$	$V_\Phi(d_6)$	$\bar{e}_1(d_6)$	$g(d_6)$	$\ell(d_6)$	$\phi_0(d_6, d_5)$	$V_\Phi(d_6, d_5)$
A	0.7555	0.05944	0.99888	0.97848	0.97738	1.00117	0.98842
D	0.7589	0.05827	0.99891	0.98277	0.98170	1.00172	0.98333
E	0.6712	0.07399	0.99091	0.87621	0.86825	0.99449	1.03435
T	0.7621	0.05719	0.99894	0.98700	0.98595	1.00224	0.97877

TABLE 4

Performance of  $d_5$  and  $d_6$  under  $\vec{a} = (0, 1/20, 3/20, 1/5, 3/5)$ .

5.1.2. *Comparison to designs of Bose and Bagchi (2008), Majumdar, Dean, and Lewis (2008), and Zhao and Majumdar (2012).* When the realization of subject dropout  $l$  is not random, we have  $\phi_0 \equiv \phi_1$ . In this case, Bose and Bagchi (2008) has the following results:

1. When  $p = t \geq 3$  is a prime or primer power and  $n = t(t - 1)$ , a design is found to be universally optimal whenever  $a_q = 1$  for any  $3 \leq q \leq p$ .
2. When  $p = t \geq 3$  is a prime or primer power,  $t \equiv 3 \pmod{4}$ , and  $n = 2t$ , a design is found to be universally optimal whenever  $a_q = 1$  with  $q = (p + 1)/2$  or  $p$ .
3. When  $p = t \geq 3$  is a prime or primer power,  $t \equiv 1 \pmod{4}$ , and  $n = 4t$ , a design is found to be universally optimal whenever  $a_q = 1$  with  $q = (p + 1)/2$  or  $p$ .

For example, when  $t = p = 5$  the smallest  $n$  should be  $4t = 20$ . In this case the design proposed by them is universally optimal, either when the experiment is complete or when all subjects immediately drop out after period 3 with probability 1, i.e.  $a_3 = 1$ . We denote this design by  $d_5$  which is given by Example 3 of Bose and Bagchi (2008). When  $\vec{a} = (0, 1/20, 3/20, 1/5, 3/5)$  Algorithm (42) yields  $d_6$  as follows:

$$d_6 : \begin{array}{cccccccccccccccccccc} 1 & 2 & 4 & 4 & 3 & 2 & 1 & 2 & 1 & 1 & 3 & 2 & 4 & 5 & 5 & 5 & 5 & 3 & 4 & 3 \\ 2 & 5 & 1 & 2 & 1 & 3 & 2 & 3 & 5 & 4 & 4 & 5 & 5 & 4 & 4 & 2 & 3 & 1 & 3 & 1 \\ 3 & 4 & 3 & 5 & 5 & 4 & 4 & 5 & 3 & 3 & 1 & 1 & 1 & 2 & 2 & 1 & 2 & 2 & 5 & 4 \\ 4 & 1 & 5 & 3 & 2 & 5 & 3 & 1 & 2 & 2 & 5 & 3 & 3 & 1 & 1 & 4 & 4 & 4 & 2 & 5 \\ 4 & 1 & 5 & 3 & 2 & 1 & 5 & 4 & 4 & 5 & 2 & 4 & 2 & 3 & 3 & 3 & 1 & 5 & 1 & 2 \end{array}$$

Table 4 shows that  $d_6$  is more efficient and robust than  $d_5$  under criteria of A, D, and T, while the result is reversed under the criterion of E. The reason for the latter is that  $d_5$  did a better job in avoiding disconnected designs under subject dropout, i.e.  $\Phi_E(C_d(\tau, l)) = 0$ .

Since the magnitude of the differences between  $d_5$  and  $d_6$  are small in terms of both efficiency and robustness, we conclude that Bose and Bagchi (2008)'s designs successfully defended the loss of information due to subject

$\Phi$	$\phi_0(d_8)$	$V_\Phi(d_8)$	$\tilde{e}_1(d_8)$	$g(d_8)$	$\ell(d_8)$	$\phi_0(d_8, d_7)$	$V_\Phi(d_8, d_7)$
A	1.2340	0.053908	1	0.99591	0.99591	1.11018	0.57705
D	1.2347	0.053736	1	0.99643	0.99643	1.10598	0.59362
E	1.2004	0.059782	1	0.96877	0.96877	1.16339	0.51397
T	1.2353	0.053573	1	0.99696	0.99696	1.10177	0.60992

TABLE 5

Performance of  $d_7$  and  $d_8$  under  $\vec{a} = (0, 0, 1/3, 1/3, 1/3)$ .

dropout. The same conclusion applies to Majumdar, Dean, and Lewis (2008) and Zhao and Majumdar (2012) since they used similar ideas.

5.1.3. *Comparisons to designs of Kushner (1998)*. Kushner (1998) derived conditions for universal optimality as a special case of ours under complete experiment. Particularly, when  $t = 3$ ,  $p = 5$ , and  $n = 30$ , Example 4 of Kushner (1998) gives a design satisfying the optimality equations therein, which is denoted  $d_7$  here. When  $\vec{a} = (0, 0, 1/3, 1/3, 1/3)$  our algorithm gives  $d_8$  which consist of five copies of (44):

$$(44) \quad d_8 : \begin{array}{cccccc} 1 & 2 & 3 & 3 & 1 & 2 \\ 3 & 3 & 2 & 1 & 2 & 1 \\ 2 & 1 & 1 & 2 & 3 & 3 \\ 2 & 1 & 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 3 & 2 & 1 \end{array} \times 5.$$

Based on Table 5  $d_8$  outperforms  $d_7$  in terms of both efficiency and robustness even though  $d_7$  is universally optimal under complete experiment.

5.2. *Symmetric exact designs*. This section illustrates the usage of Theorem 5 in deriving efficient symmetric exact designs. By Remark 4 in Section 3.5, when  $t = 2$ ,  $p = 6$ , and  $m = p - 1 = 5$ , Inequality (20) in Theorem 3 always holds regardless of the value of  $\vec{a}$ . By applying Theorem 3(i), we have  $x^* = 0$  and hence  $q'_s(x^*) = 2q_{s_{12}}$ . Moreover, it is easy to see that the support  $\mathcal{T}$  essentially contains all sequences which assign a subject to each of the two treatments for 3 out of the total of 6 periods, and hence  $|\mathcal{T}| = 20$ . Within each symmetric block, there are two sequences since  $t = 2$ . Hence there are 10 symmetric blocks. However, it is not necessary to include all these symmetric blocks in the design. Particularly when  $\vec{a} = (0, 0, 0, 0, 2/5, 3/5)$  we have  $q'_{s_1}(x^*)/q'_{s_2}(x^*) = q_{s_1_{12}}/q_{s_2_{12}} = -6.01$  for  $s_1 = 122121$  and  $s_2 = 122211$ . In the spirit of Theorem 5 we propose a small size design,  $d_9$ , which consists of one copy of sequences 122121 and 211212 and six copies of the sequences

$\Phi$	$\phi_0(d_9)$	$V_\Phi(d_9)$	$\tilde{e}_1(d_9)$	$g(d_9)$	$\ell(d_9)$
A,D,E,T and etc.	2.7368	0.09152	0.99511	0.997823	0.99295

TABLE 6

*Performance of  $d_9$  under  $\vec{a} = (0, 0, 0, 0, 2/5, 3/5)$ .*

122211 and 211122. So we have  $n = 14$  for  $d_9$ . The point is that we have the freedom of selecting different subclasses of  $\mathcal{T}$ . The performance of  $d_9$  is given in table 6. It shows the high efficiency and robustness of  $d_9$ . Note that when  $t = 2$  all criteria are equivalent.

**6. Discussions.** Subject dropout is a very important issue in planning a crossover design. It is shown by Table 5 and other examples in literature that an optimal design under complete experiment is no longer optimal and possibly even disconnected when there is subject dropout. However, the problem has received very limited attention in literature so far and the majority of the research assumes that there is no subject dropout. Bose and Bagchi (2008), Majumdar, Dean, and Lewis (2008), Zhao and Majumdar (2012) all considered the nested structure such that a design together with its subdesign obtained by taking only the first  $q (< p)$  periods are both optimal or efficient. Naturally such designs would still be efficient when all subjects drop out at periods between  $p$  and  $q$ . The issue with this approach is that we lose adaptation to different dropout mechanisms. Furthermore, their methods only apply to special configurations of  $p, t, n$ .

In order to take into account the dropout mechanism, one has to make assumptions to formulate the dropout mechanism. This paper adopts two mild assumptions and works on the target function,  $\phi_0$ , which is given by taking the expectation of a regular optimality criterion with respect to a given dropout mechanism. Actually Low, Lewis, and Prescott (1999) has followed the same approach. However, they only provided two case studies and there was no theoretical results regarding how to identify an efficient design in general. The latter problem is itself intractable. To tackle it, we propose to use the surrogate target function of  $\phi_1$  in place of  $\phi_0$ . It turns out that this replacement is very successful. Examples in Section 5 show that  $\phi_1$ -optimal (or highly efficient) designs are also highly efficient under  $\phi_0$ . Moreover, these designs are also shown to be very robust against the randomness of subject dropout due to the substantial diversity in the composition of treatment sequences.

Theoretically, we derive feasible equivalent conditions for a design to be  $\phi_1$ -universally optimal in asymptotic design theory. These conditions are essentially linear equations with respect to proportions of treatment sequences

from  $\mathcal{T}$ , a subclass of all possible treatment sequences. A solution for the equations, which yields an exact design, does not necessarily exist due to the discrete nature of the problem. However, one can follow the spirit of the conditions and easily propose an applicable algorithm to derive an efficient exact design for any criterion and any configuration of  $p, t, n$ . In this paper, we adopt Algorithm (42) for general designs as well as the approach in Section 5.2 for symmetric designs.

The problem of identifying exact designs for large values of  $p$  and  $t$  remains as an open problem. The critical difficulty is that as  $p$  and  $t$  grow the size of the support for admissible sequences,  $|\mathcal{T}|$ , increases very fast. Typically  $\mathcal{T}$  contains two distinct symmetric blocks, in which case  $p = t = 6$  usually yields  $|\mathcal{T}| = 2 \times 6! = 1440$ . That means the majority of the sequences in  $\mathcal{T}$  would not appear in the design for a moderate value of  $n$ . The same issue has appeared in Kushner (1997b). If we adopt the approach of symmetric designs as in Section 5.2 we would need  $n$  to be as of the same magnitude as  $|\mathcal{T}|$ . On the other hand, Algorithm (42) is essentially an integer programming problem and the number of the integer variables is equal to  $|\mathcal{T}|$ . Hence it would be infeasible for a computer to handle when  $|\mathcal{T}|$  is too large. For this problem, one possible solution is to reduce the size of  $\mathcal{T}$  through the study of intrinsic relationships among treatment sequences. Another approach is to resort to algorithm improvement.

**7. Proofs. Proof of Lemma 3:** It would be enough to show that  $V = \mathbb{E}O$ . First, it is easy to show that  $B_{ij}^{m_1} B_{jk}^{m_2} = B_{ik}^{\min(m_1, m_2)}$ . We have  $MU = \text{diag}(1_{l_1}, 1_{l_2}, \dots, 1_{l_n})$  and  $MP = (I_{l_1 p}^{l_1}, I_{l_2 p}^{l_2}, \dots, I_{l_n p}^{l_n})'$ . Then we have

$$\begin{aligned} pr^\perp(MU) &= \text{diag}(B_{l_1}, \dots, B_{l_n}), \\ pr^\perp(MU)MP &= (B_{l_1 p}^{l_1}, B_{l_2 p}^{l_2}, \dots, B_{l_n p}^{l_n})', \\ P'M'pr^\perp(MU)MP &= \sum_{u=1}^n B_p^{l_u} = \sum_{i=1}^p h_i B_p^i. \end{aligned}$$

Without loss of generality, we could assume  $h_p > 0$ . Then one choice of the g-inverse of  $P'M'pr^\perp(MU)MP$  is  $\sum_{i=1}^p g_i B_p^i$  where

$$(45) \quad g_i = r_i^{-1} - r_{i+1}^{-1}, \quad 1 \leq i \leq p-1,$$

$$(46) \quad g_p = h_p^{-1},$$

with  $h_k = \sum_{i=1}^n 1_{l_i=k}$ ,  $1 \leq k \leq p$ , and  $r_i = \sum_{k=i}^p h_k$  denotes the number of subjects remaining at period  $i$ ,  $1 \leq i \leq p$ . Note that if  $h_p = 0$ , the value of  $p$  in (45) and (46) should be replaced by  $\tilde{p} = \max\{k : h_k > 0\}$  and for



$k > \tilde{p}$  we let  $h_k = 0$ . It is easily seen that the follow arguments and thus the lemma would still hold. Now we have

$$\begin{aligned} pr^\perp(MP|MU) &= pr^\perp(MU) - pr(pr^\perp(MU)MP) \\ &= \text{diag}(B_{l_1}, \dots, B_{l_n}) - \Delta, \\ \Delta &= \left( \sum_{k=1}^p g_k B_{l_i l_j}^{\min(k, l_i, l_j)} \right)_{i,j=1,2,\dots,n}. \end{aligned}$$

Let  $O = (O_{ij})_{1 \leq i, j \leq n} = M' pr^\perp(MP|MU)M$ , then we have

$$\begin{aligned} O_{ii} &= B_p^{l_i} - \sum_{k=1}^p g_k B_p^{\min(k, l_i)}, \\ O_{ij} &= - \sum_{k=1}^p g_k B_p^{\min(k, l_i, l_j)}. \end{aligned}$$

We will derive the expectation of  $\sum_{k=1}^p g_k B_p^{\min(k, l_i)}$  and other components could be dealt with by similar arguments. First we have the decomposition

$$\begin{aligned} \sum_{k=1}^p g_k B_p^{\min(k, l_i)} &= \sum_{k=1}^{l_i-1} g_k B_p^k + \sum_{k=l_i}^p g_k B_p^{l_i} \\ &= \sum_{k=1}^{l_i-1} \left( \frac{1}{r_k} - \frac{1}{r_{k+1}} \right) B_p^k + \frac{1}{r_{l_i}} B_p^{l_i}. \end{aligned}$$

When  $k \leq l_i$  and  $l_i$  is given, we know that  $r_k - 1$  follows the binomial distribution with parameters  $n - 1$  and  $a_{kp}$ . Hence we have

$$\begin{aligned} \mathbb{E}(r_k^{-1} | l_i, k \leq l_i) &= \sum_{j=0}^{n-1} \frac{1}{j+1} \frac{(n-1)!}{(n-1-j)!j!} a_{kp}^j (1-a_{kp})^{n-1-j} \\ &= \frac{1 - (1-a_{kp})^n}{na_{kp}} := b_k. \end{aligned}$$

Hence we have

$$\begin{aligned} \mathbb{E} \left( \sum_{k=1}^p g_k B_p^{\min(k, l_i)} | l_i, 1 \leq i \leq n \right) &= \sum_{k=1}^{l_i-1} (b_k - b_{k+1}) B_p^k + b_{l_i} B_p^{l_i} \\ &= \sum_{k=1}^p [(b_k - b_{k+1}) 1_{[k < l_i]} + b_k 1_{[k = l_i]}] B_p^k. \end{aligned}$$

Here we have the convention of  $b_{p+1} = 0$  for notational convenience. Hence

$$\begin{aligned} \mathbb{E} \left( \sum_{k=1}^p g_k B_p^{\min(k, l_i)} \right) &= \sum_{k=1}^p [(b_k - b_{k+1})a_{k+1, p} + b_k(a_{kp} - a_{k+1, p})] B_p^k \\ &= \sum_{k=1}^p (a_{kp} b_k - a_{k+1, p} b_{k+1}) B_p^k \\ &= \frac{1}{n} \sum_{k=1}^p (a_{1k}^n - a_{1, k-1}^n) B_p^k. \end{aligned}$$

Following this strategy, it is easy to show that

$$\begin{aligned} \mathbb{E} O_{ii} &= \sum_{k=1}^p [a_k - n^{-1}(a_{1k}^n - a_{1, k-1}^n)] B_p^k, \\ \mathbb{E} O_{ij} &= -\frac{1}{n} \sum_{k=1}^p (a_k + a_{k+1, p} a_{1k}^n - a_{kp} a_{1, k-1}^n) B_p^k. \end{aligned}$$

Then we have  $V = \mathbb{E}O$ .

◇

**Proof of Theorem 2:** By definition of symmetric designs we have

$$(47) \quad \begin{aligned} T_{\sigma d} &= T_d S_\sigma \\ &= (\tilde{S}_{\sigma, d} \otimes I_p) T_d, \\ F_{\sigma d} &= (\tilde{S}_{\sigma, d} \otimes I_p) F_d, \end{aligned}$$

where  $\tilde{S}_{\sigma, d}$  is a permutation matrix for subjects induced by  $\sigma$  and (symmetric)  $d$ . Note that we have  $(\tilde{S}_{\sigma, d} \otimes I_p)' V (\tilde{S}_{\sigma, d} \otimes I_p) = V$ . So  $C_{dij}$ ,  $1 \leq i, j \leq 2$  are completely symmetric and hence  $C_d$  is completely symmetric for a symmetric design  $d$ . This yields

$$C_d = \bar{C}_d,$$

and the equality in (10). By (47) we have  $\bar{T} = \bar{T} S_\sigma$  for any  $\sigma \in \mathcal{B}$  and hence  $\bar{T} = n^{-1} \mathbf{1}_p \mathbf{1}'_t$ . Hence we have  $\hat{T} = \bar{T} B_t = 0$ . By the same argument we have  $\hat{F} = 0$ . Then the equality in (12) holds and so does the equality in (13). Hence we proved  $\text{tr}(C_d) = q_d^*$ .

Given any design  $d$  with corresponding  $P = (p_s, s \in \mathcal{S})$ , we could define a new design  $\bar{d} \leftrightarrow P_{\bar{d}} = (\bar{p}_s, s \in \mathcal{S})$  by

$$(48) \quad P_{\bar{d}} = \frac{\sum_{\sigma \in \mathcal{P}} P_{\sigma d}}{t!}.$$

Then we have  $\sum_{s \in \mathcal{S}} p_s q_s(x) = \sum_{s \in \mathcal{S}} \bar{p}_s q_s(x)$  in view of (36) and  $q_d^* = q_{\bar{d}}^*$ .

◇

**Proof of Theorem 3:** In the following, we would apply Lemma 3.1 of Kushner (1997a) to prove (iii). The proof of (i) and (ii) follows from similar arguments. Given any sequence  $s$ , we have  $q_s(x) = \sum_{k=m}^p \alpha_k q_s^k(x)$  where  $q_s^k(x) = q_{s11}^k + 2q_{s12}^k x + q_{s22}^k x^2$  and  $q_{sij}^k = \text{tr}(G_i' B_p^k G_j)$  with  $G_1 = \hat{T}_u$  and  $G_2 = \hat{F}_u$ . By direct calculation we have

$$\begin{aligned} q_{s11}^k &= k - \xi_{s_k}/k, \\ q_{s12}^k &= (k\rho_{s_k} + f_{s_k, t_k} - \xi_{s_k})/k, \\ q_{s22}^k &= (kt - 1)(k - 1)/kt - (\xi_{s_k} - 2f_{s_k, t_k} + 1)/k, \end{aligned}$$

where  $\xi_{s_k} = \sum_{i=1}^t (f_{s_k, i})^2$  and  $\rho_{s_k} = \sum_{j=1}^{k-1} 1_{t_j = t_{j+1}}$ . For notational simplicity we define  $\xi_k = \xi_{s_k}$ ,  $\rho_k = \rho_{s_k}$ , and  $f_k = f_{s_k, t_k}$ . Also let  $\Xi_A$ ,  $A \subset \{x, k, p, t\}$ , denote a quantity that depends on the elements of  $A$ , and  $a \propto_k b$  means that  $a/b$  is a quantity that only depend on  $k$ . Then

$$(49) \quad \begin{aligned} q_s^k(x) &\propto_k -\xi_k(x+1)^2 + 2f_k(x+x^2) + 2k\rho_k x + \Xi_{k,t,x} \\ &= -(\xi_k - 2f_k)(x+1)^2 + 2k(\rho_k - f_k)x \end{aligned}$$

$$(50) \quad +2f_k[(k-1)x - 1] + \Xi_{k,t,x}.$$

From (49), for any  $x > 0$ , the sequence which maximizes  $q_s^k(x)$  has to be of the form  $(1 * 1_{f_{s_k,1}} | 2 * 1_{f_{s_k,2}} | \dots | (t-1) * 1_{f_{s_k,t-1}} | t * 1_{f_{s_k,t}})$  with the restrictions of  $f_{s_k, i+1} \geq f_{s_k, i}$ ,  $i = 1, 2, \dots, t-1$  and  $f_{s_k, t-1} - f_{s_k, 1} \leq 1$ . For the special case of  $k \leq t$ , the sequence reduces to the form of  $(1, 2, \dots, k-h, t * 1_h)$ . By (50) the sequence of  $\langle re \rangle$  maximizes  $q_s(x)$  for any  $x \in ((p-1)^{-1}, (p-2)^{-1}]$  since this sequence maximizes  $q_s^k(x)$  for all  $k = m, \dots, p$ . Since all the sequences in the class of  $\langle re \rangle$  has the same value of  $d(q_s(x))/dx$ , we need to choose  $x$  so that the derivative is zero and hence (iii) is proven.

◇

**Proof of Theorem 4:** By Lemma 4, Equations (31)-(35) is equivalent to

$$(51) \quad \check{C}_{d11} + x^* \check{C}_{d12} B_t = \frac{ny^*}{t-1} B_t,$$

$$(52) \quad \check{C}_{d21} + x^* \check{C}_{d22} B_t = 0,$$

$$(53) \quad B(\hat{T} + x^* \hat{F}) = 0.$$

First we show the *necessity*. Let  $f$  be a symmetric optimal design and  $g$  be a new design with  $P_g = P_d/2 + P_f/2$ . Then by Lemmas 1 and 5 we have

$$\check{C}_g \geq \check{C}_d/2 + \check{C}_f/2$$

$$(54) \quad = \frac{ny^*}{t-1}B_t.$$

Let  $\bar{g}$  be the symmetrized version of design  $g$  as defined by (48). Following the same argument as in Lemma 1 we have

$$(55) \quad \sum_{\sigma \in \mathcal{P}} S'_\sigma \check{C}_g S_\sigma / |\mathcal{P}| \leq C_{\bar{g}}.$$

Combining (54) and (55) we have

$$C_{\bar{g}} = \frac{ny^*}{t-1}B_t,$$

in view of Corollary 2 (ii). Then we have  $\text{tr}(\check{C}_d) = ny^*$  which together with (54) yields

$$\check{C}_g = \frac{ny^*}{t-1}B_t.$$

Following similar arguments as in Theorem 5.3 of Kushner (1997b) we have

$$(56) \quad \check{C}_{f22} \check{C}_{g22}^+ \check{C}_{g21} = \check{C}_{f21},$$

$$(57) \quad \check{C}_{d22} \check{C}_{g22}^+ \check{C}_{g21} = \check{C}_{d21},$$

where  $G^+$  denotes the Moore-Penrose inverse of  $G$ . Since  $f$  is a symmetric design, we have  $\check{C}_{f21} = q_{f12}B_t/(t-1)$  and  $\check{C}_{f22} = q_{f22}B_t/(t-1) + (1'_t \check{C}_{f22} 1_t)J_t/t^2$ . So we have  $\check{C}_{f22}^+ = (t-1)B_t/q_{f22} + J_t/(1'_t \check{C}_{f22} 1_t)$ . By left multiplying both sides of (56) we have

$$(58) \quad \begin{aligned} \check{C}_{g22}^+ \check{C}_{g21} &= \check{C}_{f22}^+ \check{C}_{f21} \\ &= -x^* B_t. \end{aligned}$$

By plugging (58) into (57) we have (52). Then we have

$$\begin{aligned} \frac{ny^*}{t-1}B_t = \check{C}_d &= \check{C}_{d11} - \check{C}_{d12} \check{C}_{d22}^- \check{C}_{d21} \\ &= \check{C}_{d11} + x^* \check{C}_{d12} \check{C}_{d22}^- \check{C}_{d22} B_t \\ &= \check{C}_{d11} + x^* \check{C}_{d12} B_t. \end{aligned}$$

Hence (51) is derived. From (10) and (5.3) of Kushner (1997b) we have

$$(59) \quad \begin{aligned} ny^* &= \text{tr}(\bar{C}_d) \\ &\leq \text{tr}(\check{C}_d) \\ &\leq \text{tr}(\bar{C}_{d11} + 2x\bar{C}_{d12} + x^2\bar{C}_{d22}) \\ &= q_d(x) - n \text{tr}[(\hat{T}_d + x\hat{F}_d)' B(\hat{T}_d + x\hat{F}_d)]. \end{aligned}$$

Setting  $x = x^*$  in (59) gives  $y^* \leq y^* - \text{tr}[(\hat{T}_d + x^*\hat{F}_d)'B(\hat{T}_d + x^*\hat{F}_d)]$  which yields (53) due to Pukelsheim[(1993), page 15].

Now we show the *sufficiency*. By utilizing (51), (52), and (53) we have

$$\begin{aligned} C_{d11} + x^*C_{d12}B_t &= \frac{ny^*}{t-1}B_t, \\ C_{d21} + x^*C_{d22}B_t &= 0, \end{aligned}$$

which in turn yields

$$\begin{aligned} C_d &= C_{11} + x^*C_{d12}C_{d22}^-C_{d22}B_t \\ &= \frac{ny^*}{t-1}B_t. \end{aligned}$$

◇

## References

- Bose, M. and Bagchi, S. (2008). Optimal crossover designs under premature stopping. *Utilitas Mathematica* **75** 273–285.
- Bose, M. and Dey, A. (2009). Optimal Crossover Designs. World Scientific Publishing Company.
- Cheng, C. S., and Wu, C. F. (1980). Balanced repeated measurements designs. *Annals of Statistics* **8** 1272–1283.
- Godolphin, J. D. (2004). Simple pilot procedures for the avoidance of disconnected experimental designs. *Journal of the Royal Statistical Society Series C* **53** 133–147.
- Hedayat, A. S., and Afarinejad, K. (1978). Repeated measurements designs, II. *Annals of Statistics* **6** 619–628.
- Hedayat, A. S., and Yang, M. (2003). Universal optimality of balanced uniform crossover designs. *Annals of Statistics* **31** 978–983.
- Hedayat, A. S., and Yang, M. (2004). Universal optimality for selected crossover designs. *Journal of The American Statistical Association* **99** 461–466.
- Hedayat, A. S., and Zheng, W. (2010). Optimal and efficient crossover designs for test-control study when subject effects are random. *Journal of the American Statistical Association* **105** 1581–1592.
- Huynh, H. and Feldt, L.S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association* **65** 1582–1589.
- Jones, B. and Kenward, M.G. (2003). Design and analysis of cross-over trials, 2nd ed. London: Chapman and Hall.

- Low, J.L., Lewis, S.M. and Prescott, P. (1999). Assessing robustness of crossover designs to subjects dropping out. *Statistics and Computing* **9** 219-227.
- Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In *A survey of statistical design and linear models*, J.N. Srivastava, ed., 333-353, North-Holland, Amsterdam.
- Kunert, J. (1984). Optimality of balanced uniform repeated measurements designs. *Annals of Statistics* **12** 1006-1017.
- Kunert, J., and Martin, R. J. (2000). On the determination of optimal designs for an interference model. *Annals of Statistics* **28** 1728-1742.
- Kunert, J., and Stufken, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. *Journal of The American Statistical Association* **97** 898-906.
- 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. *Annals of Statistics* **25** 2328-2344.
- Kushner, H. B. (1998). Optimal and efficient repeated-measurements designs for uncorrelated observations. *Journal of The American Statistical Association* **93** 1176-1187.
- Majumdar, D., Dean, A.M., and Lewis, S. M. (2008). Uniformly balanced repeated measurements designs in the presence of subject dropout. *Statist. Sinica* **18** 235-253.
- Matthews, J.N.S. (1988). Recent development in crossover designs. *Internat. Statist. Rev.* **56** 117-127.
- Pukelsheim, F. (1993). Optimal design of experiments. Wiley, New York.
- Ratkowsky, D.A., Evans, M.A., and Alldredge, J.R. (1992). Cross-over experiments: Design, analysis, and application. New York: Marcel Dekker.
- Senn, S. (2003). Cross-over trials in clinical research, 2nd ed. Chichester, England: Wiley.
- Stufken, J. (1991). Some families of optimal and efficient repeated measurements designs. *Journal of Statistical Planning and Inference* **27** 75-83.
- Stufken, J. (1996). Optimal crossover designs, Handbook of statistics 13, S. Ghosh and C. R. Rao, eds. *Design and Analysis of Experiments, North-Holland, Amsterdam* 63-90.
- Yeh, C.M. (1986). Conditions for universal optimality of block designs. *Biometrika* **73** 701-706.

DEPARTMENT OF MATHEMATICAL SCIENCES  
INDIANA UNIVERSITY-PURDUE UNIVERSITY INDIANAPOLIS  
INDIANAPOLIS, IN, 46202-3216  
E-MAIL: weizheng@iupui.edu