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), Chêng 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), for instance. For comprehensive reviews, see 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 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 notation 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 notation 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 notation.* In a crossover design with p periods, t treatments and n subjects, the response is typically modeled as

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

where $\{\varepsilon_{ku}, 1 \le k \le p, 1 \le u \le n\}$ are independent with mean zero and variance σ^2. Here, Y_{dku} denotes the response from subject u in period k to which treatment $d(k, u) \in \{1, 2, \ldots, t\}$ was assigned by design d. Furthermore, μ is the general mean, π_k is the kth period effect, ς_u is the uth 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, $\text{tr}(G)$ to represent the trace of the matrix G and $\text{pr}^\perp$ to be a projection operator such that $\text{pr}^\perp G = I - G(G'G)^- G'$. For two square matrices of equal size, G_1 and G_2, $G_1 \le G_2$ 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 notation 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, $1 \le i \le n$, 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 the following:

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

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

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

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 $\beta_k \ge a_k + a_{k+1,p}a_{1k}^n - a_{kp}a_{1k}^n = a_k(1 - a_{1k}^n) \ge 0$, we have $B \ge 0$. By the mean value theorem one could show that $\alpha_k \ge 0$ and hence $A \ge 0$. 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. Finally, 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 $Y_d = (Y_{d11}, Y_{d21}, \ldots, Y_{dp1}, Y_{d12}, \ldots, Y_{dpn})'$, model (1) can be written as

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

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

$$C_d(\tau, l) = (MT_d)' \mathrm{pr}^{\perp}(MZ|MU|MF_d)(MT_d)$$

$$= C_{d11}(l) - C_{d12}(l)[C_{d22}(l)]^{-} C_{d21}(l),$$

where

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

Under a complete experiment, Kiefer (1975) defined a design to be universally optimal if it maximizes $\Phi(C_d(\tau, p1_n))$ for any Φ satisfying:

(C.1) Φ 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 Φ 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 Φ 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 Φ and $\vec{a}$ for ϕ_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) Φ is a nonlinear function and hence the expectation would interact with the form of Φ. (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)
$$C_d = C_{d11} - C_{d12}C_{d22}^{-}C_{d21},$$
$$C_{dij} = \mathbb{E}C_{dij}(l), \qquad 1 \le i, j \le 2.$$

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 ϕ_i. Then define $e_i(d) = \phi_i(d)/\phi_i(d_i^*)$, $i = 0, 1$, to be the efficiency of d under ϕ_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 ϕ_1 instead of ϕ_0, the ϕ_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 $G \ge 0$ is a concave nondecreasing function of G.*

LEMMA 2. *For any Φ, $\vec{a}$ and design d, we have $\phi_0(d) \le \phi_1(d)$. Further we have*

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

In particular, for any ϕ_1-optimal design d, we have $e_0(d) \ge g(d)$.

PROOF. By Lemma 1 we have $\mathbb{E}C_d(\tau, l) \le C_d$. Then we have

$$\phi_0(d) = \mathbb{E}\Phi\big(C_d(\tau, l)\big)$$

(4) $$\le \Phi\big(\mathbb{E}C_d(\tau, l)\big)$$

(5) $$\le \Phi(C_d)$$

(6) $$= \phi_1(d).$$

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

By (6) we have $g(d) \le 1$, and hence $g(d_1^*) \le e_0(d_1^*)$. That means if we could find a ϕ_1-optimal design, then the value of the gap function g evaluated at this design serves as a lower bound of its ϕ_0-efficiency. Inequalities (4) and (5) are essentially Jensen-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 Φ 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 ϕ_1-universal optimality derived in Section 4 reduce to that of Kushner (1997b). Note that the matrix C_d in (3) is no longer an information matrix for any design, and 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 $A \ne B$ 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*

$$V = \sum_{k=m}^{p} \big(\alpha_k I_n - n^{-1}\beta_k J_n\big) \otimes B_p^k$$

(7)
$$= I_n \otimes A - n^{-1}J_n \otimes B.$$

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 $1 \le k \le p - 1$ 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 ϕ_1-optimal with Φ 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*

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

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

PROOF. Let Σ_k be the $k \times k$ upper left submatrix of Σ for $1 \le k \le p$. By direct calculation, we have

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

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. ϕ_1-universal optimality.
This section explores the ϕ_1-universal optimality in approximate design theory, where ϕ_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 ϕ_1-universally optimal if d maximizes $\phi_1(d)$ over all designs for any Φ 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 ϕ_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 ϕ_1-universal optimality, and further by argument of Kiefer (1975) that this condition is also necessary for any design to be ϕ_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 σ be a permutation of symbols $\{1, 2, \ldots, t\}$. For a sequence $s = (t_1, \ldots, t_p)$, we define $\sigma s = (\sigma(t_1), \ldots, \sigma(t_p))$. Then the design σd is defined by $P_{\sigma d} = (p_{\sigma^{-1}s}, s \in \mathcal{S})$. The permutation matrix S_σ 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, that is, $|\mathcal{P}| = t!$. 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 $(p_{\langle s \rangle}, s \langle \in \rangle \mathcal{S})$, where $s \langle \in \rangle \mathcal{S}$ means that s runs through all distinct symmetric blocks contained in $\mathcal{S}$.

3.2. *A sufficient condition for ϕ_1-universal optimality.* Denote by T_u (resp., F_u) the $p \times t$ submatrix of T (resp., F) corresponding to the uth subject. Define $\overline{T} = n^{-1} \sum_{u=1}^n T_u$, $\hat{T}_u = T_u B_t$ and $\hat{\overline{T}} = \overline{T} B_t$. The notation $\overline{F}$, $\hat{F}_u$ and $\hat{\overline{F}}$ are defined in the same way corresponding to carryover effects. Let $\overline{C}_{dij} = \sum_{\sigma \in \mathcal{P}} S_\sigma' C_{dij} S_\sigma / |\mathcal{P}|$, $1 \leq i, j \leq 2$ and $\overline{C}_d = \sum_{\sigma \in \mathcal{P}} S_\sigma' C_d S_\sigma / |\mathcal{P}|$. Note that $\overline{C}_{dij}, 1 \leq i, j \leq 2$, are completely symmetric, also $\overline{C}_{d11}$ and $\overline{C}_{d12} = (\overline{C}_{d21})'$ have row and column sums as zero. Let $\mathbb{I}$ be the indicator function. By Proposition 1 of [Kunert and Martin (2000)](), we have

$$(10) \qquad \overline{C}_d \leq \tilde{C}_d$$

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

where

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

$$c_{dij} = \text{tr}(B_t \overline{C}_{dij} B_t) = \text{tr}(B_t C_{dij} B_t), \qquad 1 \leq i, j \leq 2.$$

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

$$(12) \qquad (B_t C_{dij} B_t)_{1 \leq i, j \leq 2} = \begin{pmatrix} \hat{C}_{d11} & \hat{C}_{d12} \\ \hat{C}_{d21} & \hat{C}_{d22} \end{pmatrix} - n \begin{pmatrix} \hat{\overline{T}}_d' \\ \hat{\overline{F}}_d' \end{pmatrix} B \begin{pmatrix} \hat{\overline{T}}_d & \hat{\overline{F}}_d \end{pmatrix}$$

$$\leq (\hat{C}_{dij})_{1 \leq i, j \leq 2}.$$

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

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

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

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

THEOREM 1. *If $C_d = ny^* B_t/(t-1)$ with y^* defined in (14), then the design d is ϕ_1-universally optimal.*

PROOF. By conditions (C.1) and (C.2) of Φ we have

$$(15) \qquad \Phi(C_d) \leq \Phi(\overline{C}_d),$$

where the equality holds if C_d is completely symmetric, that is, $C_d = \text{tr}(C_d) B_t/(t-1)$ 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 ϕ_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 ϕ_1 instead of ϕ_0 directly.

COROLLARY 2.   (i) *There exists a symmetric ϕ_1-universally optimal design d with*

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

(ii) *If a design d is ϕ_1-universally optimal (or ϕ_1-optimal with Φ 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 in Kiefer's (1975) Proposition 1.   □

3.4. *A necessary condition for ϕ_1-universal optimality.*   In this section we give a necessary condition for a design to be ϕ_1-universally optimal and define quantities that will be useful for presenting the necessary and sufficient conditions for ϕ_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

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

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

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

By (14) and (18) we have

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

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 ϕ_1-universally optimal (or ϕ_1-optimal with Φ strictly concave or increasing) then we have*

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

PROOF.   By Corollary 2, we have $\mathrm{tr}(C_d) = ny^*$ and $C_d = \overline{C}_d$. By (10), (11) and (13) we have $\mathrm{tr}(\overline{C}_d) \leq q_d^*$. 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 $s = (t_1, t_2, \ldots, t_p)$, define $s_k = (t_1, \ldots, t_k)$ 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*

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

then

$$x^* = 0,$$

$$y^* = \sum_{k=m}^{p} \alpha_k \left[k(1 - 1/t) - r_k(t - r_k)/pt \right],$$

$$\mathcal{T} = \{ s : f_{s_k, i} = z_k \text{ or } z_k + 1, 1 \leq i \leq t, m \leq k \leq p \}.$$

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

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

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

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

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 ϕ_1-universal optimality. Built upon the results of Section 3, this section provides feasible equivalent conditions in approximate design theory for ϕ_1-universal optimality.

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

$$(23) \qquad\qquad \check{C}_d = \check{C}_{d11} - \check{C}_{d12}\check{C}_{d22}^{-}\check{C}_{d21},$$

$$\check{C}_{dij} = \sum_{u=1}^{n} \check{C}_{uij}, \qquad 1 \leq i, j \leq 2,$$

where

$$\check{C}_{u11} = T_u'(A-B)T_u + \hat{T}_u'B\hat{T}_u, \qquad \check{C}_{u12} = T_u'(A-B)F_u + \hat{T}_u'B\hat{F}_u,$$

$$\check{C}_{u21} = \check{C}_{u12}', \qquad \check{C}_{u22} = F_u'(A-B)F_u + \hat{F}_u'B\hat{F}_u.$$

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

$$(24) \qquad \check{C}_{dij} = C_{dij} + nG_i'BG_j, \qquad 1 \le i, j \le 2,$$

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

LEMMA 5. *If d is ϕ_1-universally optimal (or ϕ_1-optimal with Φ strictly concave or increasing), we have $C_d = \check{C}_d = ny^* B_t/(t-1)$.*

PROOF. By (24) and Lemma 1 we have

$$(25) \qquad C_d \le \check{C}_d.$$

By Corollary 2(ii) we have

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

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

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

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

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

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

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

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

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

Combining (25)–(28) and (29), we have

$$\frac{ny^*}{t-1} B_t = C_d = \overline{C}_d$$
$$\le \sum_{\sigma \in \mathcal{P}} S_\sigma' \check{C}_d S_\sigma / |\mathcal{P}|$$
$$\le C_{\bar{d}}.$$

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

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

The lemma is now proved in view of (25) and (30). $\square$

THEOREM 4. *A design d is ϕ_1-universally optimal (or ϕ_1-optimal with Φ strictly concave or increasing) if and only if*

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

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

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

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

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

Based on Theorem 1 and Corollary 2, (16) is also a necessary and sufficient condition for ϕ_1-universal optimality. However, (16) is not directly applicable for identifying designs. Note that the conditions in Theorem 4 are merely linear equation 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, that is,

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

Combining Theorem 4.5 of Kushner (1997b), Theorem 2, Corollary 2, Lemma 4 and equation (36), we have the following:

THEOREM 5. *A symmetric design is ϕ_1-universally optimal if*

$$\sum_{s \langle \in \rangle \mathcal{T}} p_{\langle s \rangle} q_s'(x^*) = 0,$$

$$\sum_{s \langle \in \rangle \mathcal{T}} p_{\langle s \rangle} = 1,$$

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

where $q_s'(x)$ is the derivative of $q_s(x)$ with respective 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 $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_t$. 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 ϕ_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 compares 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 the following:

COROLLARY 3. *A design d is ϕ_1-universally optimal (or ϕ_1-optimal with Φ strictly concave or increasing) if and only if*

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

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

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

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

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

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 $N_T = \{n_s, s \in T\}'$, and then equations (37)–(39) could be written in a matrix form as

$$X_T N_T = Y_T,$$

with X_T and Y_T uniquely determined by equations (37)–(41) and the ordering of the n_s in the vector N_T. To find an efficient design for an arbitrarily given n, one could choose a design which

Minimizes

(42) $$\|X_T N_T - Y_T\|,$$

subject to

$$1'_{|T|} N_T = n.$$

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 straightforward by utilizing integer optimization packages/softwares. Note that the computational complexity of the above minimization problem depends on $|T|$, 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 $V_\Phi(d) = \mathrm{Var}(\Phi(C_d(\tau, l)))$ 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 Figure 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.

$$d_2: \quad \begin{matrix} 2\,1\,2\,3\,3\,4\,3\,2\,1\,1\,1\,2\,4\,4\,4\,3 \\ 4\,4\,3\,4\,1\,1\,2\,1\,2\,2\,3\,4\,3\,3\,2\,1 \\ 3\,2\,1\,1\,2\,3\,4\,4\,3\,3\,4\,1\,2\,2\,1\,4 \\ 3\,2\,1\,1\,2\,3\,4\,4\,4\,4\,2\,3\,1\,1\,3\,2 \end{matrix}.$$

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 ϕ_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 function g for T-criterion are always the largest among all criteria, which is due to the linearity of T-criterion.
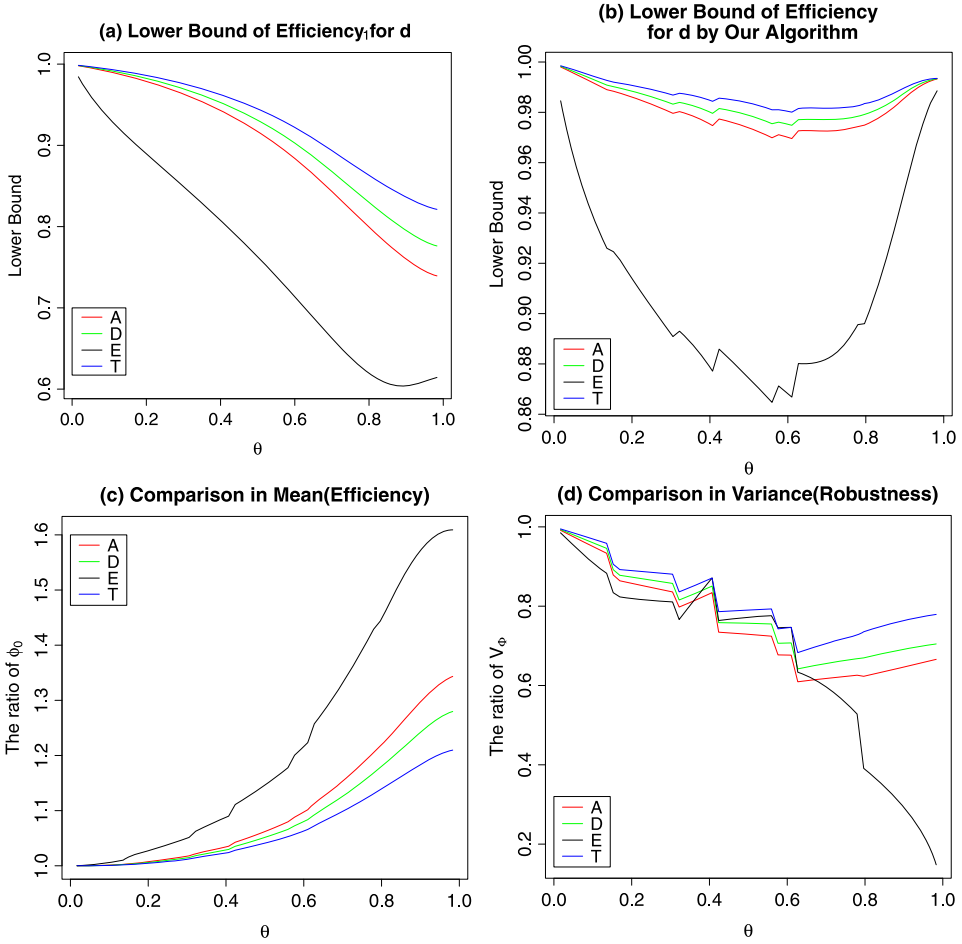
(a) Lower Bound of Efficiency₁for d

(b) Lower Bound of Efficiency for d by Our Algorithm

(c) Comparison in Mean(Efficiency)

(d) Comparison in Variance(Robustness)

FIG. 1. *The letters A, D, E and T represent the choice of criteria function Φ.* (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* (42). *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*: $V_\Phi(d)/V_\Phi(d_1)$.

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

Φ	$\phi_0(d_1)$	$V_\Phi(d_1)$	$\tilde{e}_1(d_1)$	$g(d_1)$	$\ell(d_1)$
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 2
Performance of d_2 under $\vec{a} = (0, 0, 1/2, 1/2)$

Φ	$\phi_0(d_2)$	$V_\Phi(d_2)$	$\tilde{e}_1(d_2)$	$g(d_2)$	$\ell(d_2)$
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

In comparison, d_2 is more efficient and robust than d_1 under all criteria in view of the columns of ϕ_0 and V_Φ, 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 θ. 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 4×4 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

TABLE 3
Performance of d_3 and d_4 under $\vec{a} = (0, 1/10, 2/5, 1/2)$

Φ	$\phi_0(d_4)$	$V_\Phi(d_4)$	$\tilde{e}_1(d_4)$	$g(d_4)$	$\ell(d_4)$	$\phi_0(d_4, d_3)$	$V_\Phi(d_4, d_3)$
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

exception for the robustness under E-criterion.

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

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) have 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, that is, $a_3 = 1$. 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:

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

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, that is, $\Phi_E(C_d(\tau, l)) = 0$.

Since the magnitude of the differences between d_5 and d_6 are small in terms of both efficiency and robustness, we conclude that the designs of Bose and Bagchi (2008) successfully defended the loss of information due to subject dropout. The same conclusion applies to Majumdar, Dean and Lewis (2008) and Zhao and Majumdar (2012) since they use similar ideas.

TABLE 4
Performance of d_5 and d_6 under $\vec{a} = (0, 1/20, 3/20, 1/5, 3/5)$

Φ	$\phi_0(d_6)$	$V_\Phi(d_6)$	$\tilde{e}_1(d_6)$	$g(d_6)$	$\ell(d_6)$	$\phi_0(d_6, d_5)$	$V_\Phi(d_6, d_5)$
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

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),

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

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 $q'_s(x^*) = 2q_{s12}$. 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

TABLE 5
Performance of d_7 and d_8 under $\vec{a} = (0, 0, 1/3, 1/3, 1/3)$

Φ	$\phi_0(d_8)$	$V_\Phi(d_8)$	$\tilde{e}_1(d_8)$	$g(d_8)$	$\ell(d_8)$	$\phi_0(d_8, d_7)$	$V_\Phi(d_8, d_7)$
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 6
Performance of d_9 under $\vec{a} = (0, 0, 0, 0, 2/5, 3/5)$

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

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 sized design, d_9, which consists of one copy of sequences 122121 and 211212 and six copies of the sequences 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 ϕ_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) have followed the same approach. However, they only provided two case studies, and there were 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 ϕ_1 in place of ϕ_0. It turns out that this replacement is very successful. Examples in Section 5 show that ϕ_1-optimal (or highly efficient) designs are also highly efficient under ϕ_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 ϕ_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 $MU = \mathrm{diag}(1_{l_1}, 1_{l_2}, \ldots, 1_{l_n})$ and $MZ = (I_{l_1 p}^{l_1'}, I_{l_2 p}^{l_2'}, \ldots, I_{l_n p}^{l_n'})'$. Then we have

$$\mathrm{pr}^\perp(MU) = \mathrm{diag}(B_{l_1}, \ldots, B_{l_n}),$$

$$\mathrm{pr}^\perp(MU)MZ = \left(B_{l_1 p}^{l_1'}, B_{l_2 p}^{l_2'}, \ldots, B_{l_n p}^{l_n'}\right)',$$

$$Z'M'\,\mathrm{pr}^\perp(MU)MZ = \sum_{u=1}^{n} B_p^{l_u} = \sum_{i=1}^{p} h_i B_p^i.$$

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

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

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

with $h_k = \sum_{i=1}^{n} 1_{l_i = k}, 1 \le k \le p$, and $r_i = \sum_{k=i}^{p} h_i$ denotes the number of subjects remaining at period i, $1 \le i \le p$. 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 following arguments and thus the lemma would still hold. Now we have

$$\text{pr}^\perp(MZ|MU) = \text{pr}^\perp(MU) - \text{pr}(\text{pr}^\perp(MU)MZ)$$

$$= \text{diag}(B_{l_1}, \ldots, B_{l_n}) - \Delta,$$

$$\Delta = \left(\sum_{k=1}^{p} g_k B_{l_i l_j}^{\min(k,l_i,l_j)}\right)_{i,j=1,2,\ldots,n}.$$

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

$$O_{ii} = B_p^{l_i} - \sum_{k=1}^{p} g_k B_p^{\min(k,l_i)},$$

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

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

$$\sum_{k=1}^{p} g_k B_p^{\min(k,l_i)} = \sum_{k=1}^{l_i-1} g_k B_p^k + \sum_{k=l_i}^{p} g_k B_p^{l_i}$$

$$= \sum_{k=1}^{l_i-1} \left(\frac{1}{r_k} - \frac{1}{r_{k+1}}\right) B_p^k + \frac{1}{r_{l_i}} B_p^{l_i}.$$

When $k \le l_i$ 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

$$\mathbb{E}(r_k^{-1}|l_i, k \le l_i) = \sum_{j=0}^{n-1} \frac{1}{j+1} \frac{(n-1)!}{(n-1-j)!j!} a_{kp}^j (1 - a_{kp})^{n-1-j}$$

$$= \frac{1 - (1 - a_{kp})^n}{n a_{kp}} := b_k.$$

Hence we have

$$\mathbb{E}\left(\sum_{k=1}^{p} g_k B_p^{\min(k,l_i)} \Big| l_i, 1 \le i \le n\right) = \sum_{k=1}^{l_i-1} (b_k - b_{k+1}) B_p^k + b_{l_i} B_p^{l_i}$$

$$= \sum_{k=1}^{p} \left[(b_k - b_{k+1}) 1_{[k<l_i]} + b_k 1_{[k=l_i]}\right] B_p^k.$$

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

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

$$= \sum_{k=1}^{p} (a_{kp}b_k - a_{k+1,p}b_{k+1})B_p^k$$

$$= \frac{1}{n} \sum_{k=1}^{p} (a_{1k}^n - a_{1,k-1}^n)B_p^k.$$

Following this strategy, it is easy to show that

$$\mathbb{E}O_{ii} = \sum_{k=1}^{p} [a_k - n^{-1}(a_{1k}^n - a_{1,k-1}^n)]B_p^k,$$

$$\mathbb{E}O_{ij} = -\frac{1}{n} \sum_{k=1}^{p} (a_k + a_{k+1,p}a_{1k}^n - a_{kp}a_{1,k-1}^n)B_p^k.$$

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

PROOF OF THEOREM 2. By definition of symmetric designs we have

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

where $\tilde{S}_{\sigma,d}$ is a permutation matrix for subjects induced by σ 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}, $1 \le i, j \le 2$, are completely symmetric and hence C_d is completely symmetric for a symmetric design d. This yields

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

and the equality in (10). By (47) we have $\overline{T} = \overline{T}S_\sigma$ for any $\sigma \in \mathcal{B}$ and hence $\overline{T} = n^{-1}1_p1_t'$. Hence we have $\hat{\overline{T}} = \overline{T}B_t = 0$. By the same argument we have $\hat{\overline{F}} = 0$. 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)
$$P_{\bar{d}} = \frac{\sum_{\sigma \in \mathcal{P}} P_{\sigma d}}{t!}.$$

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 $q_{\bar{d}}^* = q_{\hat{d}}^*$. $\quad\square$

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

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

where $\xi_{sk} = \sum_{i=1}^t (f_{s_k, i})^2$ and $\rho_{sk} = \sum_{j=1}^{k-1} 1_{t_j = t_{j+1}}$. For notational simplicity we define $\xi_k = \xi_{sk}$, $\rho_k = \rho_{sk}$ and $f_k = f_{s_k, t_k}$. Also let Ξ_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)
$$q_s^k(x) \propto_k -\xi_k(x+1)^2 + 2f_k(x + x^2) + 2k\rho_k x + \Xi_{k,t,x}$$
$$= -(\xi_k - 2f_k)(x+1)^2 + 2k(\rho_k - f_k)x$$
(50)
$$+ 2f_k[(k-1)x - 1] + \Xi_{k,t,x}.$$

From (49), for any $x > 0$, the sequence which maximizes $q_s^k(x)$ has to be of the form $(1 * 1_{f_{s_k,1}} | 2 * 1_{f_{s_k,2}} | \ldots \ldots | (t-1) * 1_{f_{s_k,t-1}} | t * 1_{f_{s_k,t}})$ with the restrictions of $f_{s_k, i+1} \geq f_{s_k, i}$, $i = 1, 2, \ldots, t-1$ 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 $(1, 2, \ldots, k-h, t * 1_h)$. 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 $k = m, \ldots, p$. Since all the sequences in the class of $\langle re \rangle$ have 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. $\square$

PROOF OF THEOREM 4. By Lemma 4, equations (31)–(35) is equivalent to

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

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

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

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

(54)
$$\check{C}_g \geq \check{C}_d/2 + \check{C}_f/2$$
$$= \frac{ny^*}{t-1} B_t.$$

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

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

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

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

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

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

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

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

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

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

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

$$C_{d11} + x^*C_{d12}B_t = \frac{ny^*}{t-1}B_t,$$

$$C_{d21} + x^*C_{d22}B_t = 0,$$

which in turn yields

$$C_d = C_{11} + x^*C_{d12}C_{d22}^{-}C_{d22}B_t$$

$$= \frac{ny^*}{t-1}B_t. \qquad \square$$

Acknowledgements. We are grateful to the referees and the Associate Editor for their constructive comments on earlier versions of this manuscript.

REFERENCES

BOSE, M. and BAGCHI, S. (2008). Optimal crossover designs under premature stopping. *Util. Math.* **75** 273–285. MR2392763

BOSE, M. and DEY, A. (2009). *Optimal Crossover Designs*. World Scientific, Hackensack, NJ. MR2524180

CHÊNG, C. S. and WU, C.-F. (1980). Balanced repeated measurements designs. *Ann. Statist.* **8** 1272–1283. MR0594644

GODOLPHIN, J. D. (2004). Simple pilot procedures for the avoidance of disconnected experimental designs. *J. Roy. Statist. Soc. Ser. C* **53** 133–147. MR2043764

HEDAYAT, A. and AFSARINEJAD, K. (1978). Repeated measurements designs. II. *Ann. Statist.* **6** 619–628. MR0488527

HEDAYAT, A. S. and YANG, M. (2003). Universal optimality of balanced uniform crossover designs. *Ann. Statist.* **31** 978–983. MR1994737

HEDAYAT, A. S. and YANG, M. (2004). Universal optimality for selected crossover designs. *J. Amer. Statist. Assoc.* **99** 461–466. MR2062831

HEDAYAT, A. S. and ZHENG, W. (2010). Optimal and efficient crossover designs for test-control study when subject effects are random. *J. Amer. Statist. Assoc.* **105** 1581–1592. MR2796573

HUYNH, H. and FELDT, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *J. Amer. Statist. Assoc.* **65** 1582–1589.

JONES, B. and KENWARD, M. G. (2003). *Design and Analysis of Cross-Over Trials*, 2nd ed. Chapman & Hall, London.

KIEFER, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models* (*Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo.*, 1973) (J. N. Srivastava, ed.) 333–353. North-Holland, Amsterdam. MR0395079

KUNERT, J. (1984). Optimality of balanced uniform repeated measurements designs. *Ann. Statist.* **12** 1006–1017. MR0751288

KUNERT, J. and MARTIN, R. J. (2000). On the determination of optimal designs for an interference model. *Ann. Statist.* **28** 1728–1742. MR1835039

KUNERT, J. and STUFKEN, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. *J. Amer. Statist. Assoc.* **97** 898–906. MR1941418

KUSHNER, H. B. (1997a). Optimality and efficiency of two-treatment repeated measurements designs. *Biometrika* **84** 455–468. MR1467060

KUSHNER, H. B. (1997b). Optimal repeated measurements designs: The linear optimality equations. *Ann. Statist.* **25** 2328–2344. MR1604457

KUSHNER, H. B. (1998). Optimal and efficient repeated-measurements designs for uncorrelated observations. *J. Amer. Statist. Assoc.* **93** 1176–1187. MR1649211

LOW, J. L., LEWIS, S. M. and PRESCOTT, P. (1999). Assessing robustness of crossover designs to subjects dropping out. *Statist. Comput.* **9** 219–227.

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. MR2384987

MATTHEWS, J. N. S. (1988). Recent developments in crossover designs. *Internat. Statist. Rev.* **56** 117–127. MR0963525

PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. Wiley, New York. MR1211416

RATKOWSKY, D. A., EVANS, M. A. and ALLDREDGE, J. R. (1992). *Cross-Over Experiments*: *Design*, *Analysis*, *and Application*. Dekker, New York.

SENN, S. (2003). *Cross-over Trials in Clinical Research*, 2nd ed. Wiley, Chichester.

STUFKEN, J. (1991). Some families of optimal and efficient repeated measurements designs. *J. Statist. Plann. Inference* **27** 75–83. MR1089354

STUFKEN, J. (1996). Optimal crossover designs. In *Design and Analysis of Experiments* (S. Ghosh and C. R. Rao, eds.). *Handbook of Statist.* **13** 63–90. North-Holland, Amsterdam. MR1492565

YEH, C.-M. (1986). Conditions for universal optimality of block designs. *Biometrika* **73** 701–706. MR0897862

ZHAO, S. and MAJUMDAR, D. (2012). On uniformly balanced crossover designs efficient under subject dropout. *J. Stat. Theory Pract.* **6** 178–189.

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