A General Equivalence Theorem for Crossover Designs under Generalized Linear Models

Jeevan Jankar
University of Georgia, Athens, Georgia
Jie Yang
University of Illinois at Chicago, Chicago, USA
Abhyuday Mandal
University of Georgia, Athens, Georgia

Abstract

With the help of Generalized Estimating Equations, we identify locally *D*-optimal crossover designs for generalized linear models. We adopt the variance of parameters of interest as the objective function, which is minimized using constrained optimization to obtain optimal crossover designs. In this case, the traditional general equivalence theorem could not be used directly to check the optimality of obtained designs. In this manuscript, we derive a corresponding general equivalence theorem for crossover designs under generalized linear models.

AMS (2000) subject classification. 05B15, 62J12, 62K05, 62L05.
Keywords and phrases. Approximate design, crossover design, *D*-optimality, generalized estimating equations, general equivalence theorem.

1 Introduction

Crossover designs, also known as change-over or repeated measurement designs, are widely used in many industrial, medical, and agricultural research. In crossover experiments, the effect of treatments is carried over in the periods following the period of their direct application. Crossover designs have been extensively studied in literature (Bailey and Kunert, 2006, Hedayat and Yang, 2005, Jones and Kenward, 2014).

Crossover designs have recently been used to address problems outside of medical and agriculture research. In recent years, most corporate offices

---

and organizations have adopted open office spaces over traditional cubical office spaces. Booking.com conducted an experiment to assess different office spacing efficiency and the case study was first reported in Pitchforth et al. (2020). In the absence of literature on optimal cross-over designs for generalized linear models (GLMs), traditional uniform designs are used. Uniform design is optimal under a linear model but they are no longer a good choice for non-Gaussian responses.

Over the years optimal crossover designs for normal responses have been widely studied in the literature, however, there are several examples in real life where responses are not normal and described by GLMs. Recently, Jankar et al. (2020) provided an algorithm to search locally *D*-optimal crossover designs in case of non-normal response, and showed that optimal designs obtained for normal responses can be quite inefficient in the case of GLMs. But, there was no guarantee that the designs obtained by their algorithm were indeed optimal. In this manuscript, we derive a general equivalence theorem specifically for crossover designs under GLMs, which can be used to verify the optimality of proposed designs. Moreover, it provides an alternative that is faster and numerically more stable than the general algorithm proposed in Jankar et al. (2020).

The General Equivalence Theorem is an important tool in optimum experimental designs, which has been widely used for checking for the optimality of designs in terms of the Fisher information matrix (Atkinson et al., 2007, Fedorov, 1971, 1972, Fedorov and Leonov, 2014, Fedorov and Malyutov, 1972, Kiefer and Wolfowitz, 1960, Whittle and Malyutov, 1973). Nevertheless, the traditional equivalence theorem does not apply to check the optimality of obtained crossover designs. The optimal crossover designs under GLMs discussed (Jankar et al. 2020) are identified using generalized estimating equations (GEEs) and are based on the variance matrix of the parameters of interest. Since the variance matrix is asymptotically connected with the inverse of the Fisher information matrix, it is natural to derive a condition that can be used to check the optimality of designs (see Remark 1 for more details).

For illustration purposes, we consider two real-life motivating examples. First, we consider an experiment conducted at Booking.com to determine the optimal office design. In the supplementary material, we discuss another motivating example, an experiment conducted to investigate the effects of various dietary starch levels on milk production. Kenward and Jones (1992) discussed this dietary example along with the data set used for analysis

(for more details see Jones and Kenward (2014)). The design used in both these examples is a 4×4 Latin Square design with four periods and four treatments.

The paper is organized as follows. To set ideas we describe notation and definitions for crossover designs in Section 2. In Section 3 we propose and derive two different versions of the general equivalence theorem for crossover designs. More specifically, in Section 3.1 we use the variance of all parameter estimates as an objective function, and in Section 3.2 we use the variance of treatment effects as an objective function to derive two versions of the theorem. We present illustration in Section 3.3 and real-life motivating examples in Section 4.

2 Notation and Preliminaries

Consider a crossover trial with t treatments, n subjects, and p periods. The responses obtained from these n subjects are denoted as $\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_n}$, where the response from the jth subject is $\boldsymbol{Y_j} = (Y_{1j}, \ldots, Y_{pj})'$. Let μ_{ij} denote the mean of a response Y_{ij}. To fix ideas, first, consider the following model (see, Eq. (4.1) in Wu and Hamada (2009) and Bose and Dey (2009) for linear model), which models the marginal mean μ_{ij} for crossover trial as

$$\mathrm{g}(\mu_{ij}) = \eta_{ij} = \lambda + \beta_i + \tau_{d(i,j)} + \rho_{d(i-1,j)}, \tag{1}$$

where $i = 1, \ldots, p; j = 1, \ldots, n$; λ is the overall mean, β_i represents the effect of the ith period, τ_s is the direct effect due to treatment s and ρ_s is the carryover effect due to treatment s, where $s = 1, \ldots, t$ and g is the link function.

In matrix notation, under baseline constraints $\beta_1 = \tau_1 = \rho_1 = 0$ we have $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_p)'$, $\boldsymbol{\tau} = (\tau_2, \ldots, \tau_t)'$ and $\boldsymbol{\rho} = (\rho_2, \ldots, \rho_t)'$, which defines the parameter vector $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}', \boldsymbol{\tau}', \boldsymbol{\rho}')'$. The linear predictor corresponding to the jth subject, $\boldsymbol{\eta_j} = (\eta_{1j}, \ldots, \eta_{pj})'$, can be written as

$$\boldsymbol{\eta_j} = \boldsymbol{X_j}\boldsymbol{\theta}.$$

The corresponding design matrix $\boldsymbol{X_j}$ can be written as $\boldsymbol{X_j} = [\boldsymbol{1_p}, \boldsymbol{P_j}, \boldsymbol{T_j}, \boldsymbol{F_j}]$, where $\boldsymbol{P_j}$ is $p \times (p-1)$ such that $\boldsymbol{P_j} = [\boldsymbol{0}_{(p-1)1}, \boldsymbol{I}_{p-1}]'$; $\boldsymbol{T_j}$ is a $p \times (t-1)$ matrix with its $(i, s-1)^{th}$ entry equal to 1 if subject j receives the direct effect of the treatment $s \ (\geq 2)$ in the ith period and zero otherwise; $\boldsymbol{F_j}$ is a

$p \times (t-1)$ matrix with its $(i, s-1)^{th}$ entry equal to 1 if subject j receives the carryover effect of the treatment s (≥ 2) in the ith period and zero otherwise.

If the number of subjects n and the number of periods p are fixed, then the goal is to determine the number of subjects assigned to different treatment sequences through some optimality criterion. As the number of periods p is fixed, each treatment sequence will be of length p and a typical sequence can be written as $\omega = (t_1, \ldots, t_p)'$ where $t_i \in \{1, \ldots, t\}$. Now let $\boldsymbol{\Omega}$ be the set of all such sequences and n_ω denote the number of subjects assigned to sequence ω. Then the total number of subjects n can be written as $n = \Sigma_{\omega \in \boldsymbol{\Omega}} n_\omega, n_\omega \geq 0$. A crossover design ξ in approximate theory is specified by the set $\{p_\omega, \omega \in \boldsymbol{\Omega}\}$, where $p_\omega = n_\omega/n$ is the proportion of subjects assigned to treatment sequence ω. As denoted by Silvey (1980), such a crossover design ξ can be written as follows:

$$\xi = \begin{Bmatrix} \omega_1 \ \omega_2 \ \ldots \ \omega_k \\ p_1 \ p_2 \ \ldots \ p_k \end{Bmatrix}, \tag{2}$$

where k is the number of treatment sequences involved, ω_i is the ith treatment sequence and p_i is the corresponding proportion of units allocated to that support point, such that $\sum_{i=1}^{k} p_i = 1$, for $i = 1, \ldots, k$. Note Jankar et al. (2020) observed that, in the case of non-uniform allocations, only a few sequences have non-zero proportions. Hence in our illustrations, we consider $\boldsymbol{\Omega}$ to be the collection of only those sequences that have non-zero allocations.

Generalized estimating equations are quasi-likelihood equations that allow us to estimate quasi-likelihood estimators (Prentice, 1988, Zeger et al., 1988). In crossover trials, it is typical to make an assumption that the observations from the same subject are correlated while the observations from different subjects are independent (Jones and Kenward 2014). This dependency among repeated observations from the same subject can be modeled by the "working correlation" matrix $\boldsymbol{C_\alpha}$, which is a function of correlation coefficient α. If $\boldsymbol{C_\alpha}$ is the true correlation matrix of $\boldsymbol{Y_j}$, then from the definition of covariance we can write

$$Cov(\boldsymbol{Y_j}) = \boldsymbol{D_j}^{1/2} \boldsymbol{C_\alpha} \boldsymbol{D_j}^{1/2},$$

where $\boldsymbol{D_j} = diag\Big(Var(Y_{1j}), \ldots, Var(Y_{pj})\Big)$. Let us denote $Cov(\boldsymbol{Y_j})$ by $\boldsymbol{W_j}$. In Zeger et al. (1988) (equation (3.1)) it was shown that for repeated

measurement models, the generalized estimating equations (GEE) are defined to be

$$\sum_{j=1}^{n} \frac{\partial \boldsymbol{\mu_j}'}{\partial \boldsymbol{\theta}} \boldsymbol{W_j}^{-1} \left(\boldsymbol{Y_j} - \boldsymbol{\mu_j} \right) = 0,$$

where $\boldsymbol{\mu_j} = (\mu_{1j}, \ldots, \mu_{pj})'$ and the asymptotic variance for the GEE estimator $\hat{\boldsymbol{\theta}}$ (see Zeger et al. (1988), equation (3.2)) is

$$\text{Var}(\hat{\boldsymbol{\theta}}) = \left[\sum_{j=1}^{k} n p_j \frac{\partial \boldsymbol{\mu_j}'}{\partial \boldsymbol{\theta}} \boldsymbol{W_j}^{-1} \frac{\partial \boldsymbol{\mu_j}}{\partial \boldsymbol{\theta}} \right]^{-1} = \boldsymbol{M}^{-1}, \tag{3}$$

where $\frac{\partial \boldsymbol{\mu_j}'}{\partial \boldsymbol{\theta}} = \boldsymbol{X_j}' \text{diag} \left\{ (g^{-1})'(\eta_{1j}), \ldots, (g^{-1})'(\eta_{pj}) \right\}$ and j stands for the jth treatment sequence. In Section 3, we will define $\boldsymbol{M}$ explicitly for crossover designs. Later, we consider the situation where direct treatment effects are studied specifically.

Remark 1 Note that the subject effect term is not included in the model (1). In this work, GLM is used to describe the response and hence the Fisher information matrix depends on the model parameters. Since we are considering the local optimality approach, an educated guess for the subject effect, if included, will be needed. But, from a design point of view, the subject effect has to be treated random. In model (1), the link function is used to model only the mean response and hence we are free to choose a variance-covariance matrix. So, instead of including a random subject effect in the model, we choose a working variance-covariance matrix through GEE to capture the effect of a subject (see Appendix A.3 in Stroup et al. (2018), Fitzmaurice et al. (2011), Jankar et al. (2020) and references therein).

Remark 2 The general equivalence theorem describes the optimality criteria in terms of the Fisher information matrix. The information matrix for optimal crossover designs under GLMs is defined as the inverse of the variance-covariance matrix of parameters of interest through GEE, which is easier to obtain and works similarly to the Fisher information matrix. Here we assume that the responses from a particular subject are mutually correlated, while the responses from different subjects are uncorrelated. According to Jankar et al. (2020), the obtained optimal designs are robust to the choices of such working correlation matrices.

As mentioned in Atkinson et al. (2007), the general equivalence theorem can be viewed as a consequence of the result that the derivative of a smooth function over an unconstrained region is zero at its minimum. In this manuscript, we derive the general equivalence theorem for crossover designs by calculating the directional derivative of an objective function $\Phi(\xi)$ expressed in terms of $\boldsymbol{M}(\xi)$. Consider $\bar{\xi}_i$ to be the design that puts unit mass at the point x_i, i.e., the design supported only at x_i, where $i = 1, 2, \ldots, k$. Let $\xi_i' = (1 - h)\xi + h\bar{\xi}_i$. Then the derivative of $\Phi(\xi)$ in the direction $\bar{\xi}_i$ or x_i in case of D-optimal criterion is

$$\phi(x_i, \xi) = \lim_{h \to 0^+} \frac{1}{h}[\Phi(\xi_i') - \Phi(\xi_i)] = -\lim_{h \to 0^+} \frac{1}{h}[\ln det(\boldsymbol{M}(\xi_i')) - \ln det(\boldsymbol{M}(\xi_i))],$$

and ξ is D-optimal if and only if $min_i\phi(x_i, \xi) = 0$ and $\phi(x_i, \xi) = 0$ if $p_{\omega_i} > 0$, where this minimum is occurring at the points of support of design.

In the case of crossover designs and estimates using generalized estimating equations, a different approach compared to the one mentioned above is needed as the design points are finite and pre-specified for crossover designs. We use the technique used in the supplement materials of Yang et al. (2016). Instead of using $\xi_i' = (1 - h)\xi + h\bar{\xi}_i = \xi + h(\bar{\xi}_i - \xi)$, they used $\boldsymbol{p_r} + u\boldsymbol{\delta}_i^{(r)}$, where $\boldsymbol{p_r}$ and $\boldsymbol{\delta}_i^{(r)}$ are defined below. Therefore, the directional derivative $\phi(u, \boldsymbol{p_r})$ of the objective function is equal to $\frac{\partial \Phi(\boldsymbol{p_r} + u\boldsymbol{\delta}_i^{(r)})}{\partial u}\Big|_{u=0}$.

Here is the outline of the general equivalence theorem in the case of crossover designs. Note that $0 \leq p_i < 1$ for $i = 1, \ldots, k$, and since $\sum_{i=1}^{k} p_i = 1$ we may assume without any loss of generality that $p_k > 0$. Define $\boldsymbol{p_r} = (p_1, \ldots, p_{k-1})'$, and $\Phi(\boldsymbol{p_r}) = -\ln det(\boldsymbol{M}(p_1, \ldots, p_{k-1}, 1 - \sum_{i=1}^{k-1} p_i))$. Let $\boldsymbol{\delta}_i^{(r)} = (-p_1, \ldots, -p_{i-1}, 1-p_i, -p_{i+1}, \ldots, -p_{k-1})'$ for $i = 1, \ldots, k-1$. $\boldsymbol{\delta}_i^{(r)}$ are defined in such a way that the determinant $|(\boldsymbol{\delta}_1^{(r)}, \ldots, \boldsymbol{\delta}_{k-1}^{(r)})| = p_k \neq 0$. Hence, $\boldsymbol{\delta}_1^{(r)}, \ldots, \boldsymbol{\delta}_{k-1}^{(r)}$ are linearly independent and thus can serve as the new basis of

$$\boldsymbol{S_r} = \{(p_1, \ldots, p_{k-1})' | \sum_{i=1}^{k-1} p_i < 1, \text{and } p_i \geq 0, i = 1, \ldots, k - 1\}.$$

Note that negative $\ln det$ is a convex function on a set of positive definite matrices. Hence, $\boldsymbol{p_r}$ minimizes $\Phi(\boldsymbol{p_r})$ if and only if along each direction $\boldsymbol{\delta_i^r}$,

$$\left.\frac{\partial\Phi(\boldsymbol{p_r}+u\boldsymbol{\delta_i^{(r)}})}{\partial u}\right|_{u=0} \left\{ \begin{array}{l} = 0 \text{ if } p_i > 0 \\ \geq 0 \text{ if } p_i = 0 \end{array} \right.$$

3   Equivalence Theorems for Crossover Designs

As defined earlier, $\boldsymbol{C_\alpha}$ is the "working correlation" matrix and hence is a positive definite and symmetric. So, there exists a square matrix $\boldsymbol{R}$ such that $\boldsymbol{C_\alpha}^{-1} = \boldsymbol{R^T R}$. Then the inverse of the variance of the parameter estimates through GEEs is as follows:

$$\boldsymbol{M} = \sum_{j=1}^{k} np_j \frac{\partial\boldsymbol{\mu_j}'}{\partial\boldsymbol{\theta}} \boldsymbol{W_j}^{-1} \frac{\partial\boldsymbol{\mu_j}}{\partial\boldsymbol{\theta}} = \sum_{j=1}^{k} np_j \boldsymbol{X_j}^T \boldsymbol{G_j} \boldsymbol{D_j}^{-\frac{1}{2}} \boldsymbol{C_\alpha}^{-1} \boldsymbol{D_j}^{-\frac{1}{2}} \boldsymbol{G_j} \boldsymbol{X_j} \text{ (4)}$$

,where $\boldsymbol{G_j} = \text{diag}\left\{(g^{-1})'(\eta_{1j}), \ldots, (g^{-1})'(\eta_{pj})\right\}$. Equation 4 can be further simplified as,

$$\boldsymbol{M} = \sum_{j=1}^{k} np_j (\boldsymbol{X_j}^*)^T (\boldsymbol{X_j}^*),$$

where $\boldsymbol{X_j}^* = \boldsymbol{R}\boldsymbol{D_j}^{-\frac{1}{2}}\boldsymbol{G_j}\boldsymbol{X_j}$.

3.1   Equivalence Theorem when Objective Function is Variance of Parameter Estimates   In this section, we present the equivalence theorem for crossover design when the objective function is a determinant of the variance of parameter estimates. We also present a special case of the theorem when there are only two treatment sequences involved in the design.

Theorem 1 (*General Equivalence Theorem for Crossover Design when the objective function is $|\boldsymbol{Var(\hat{\theta})}|$*): *Consider the design ξ with k treatment sequences as defined in (2). Then, (a) The set of optimal designs is convex.*
(b) The design ξ is D-optimal if and only if

$$trace\left(\boldsymbol{X_i}^* \boldsymbol{M}(\xi)^{-1} \boldsymbol{X_i}^{*T}\right) \left\{ \begin{array}{l} = m \text{ if } p_i > 0 \\ \leq m \text{ if } p_i = 0 \end{array} \right.,$$

for each $p_i \in [0,1]$, where p_i is the allocation corresponding to point ω_i of design ξ for all $i = 1, 2, \ldots, k$, and m is the number of parameters in $\boldsymbol{\theta}$.

Proof of Theorem 3.1:

Let k be the number of treatment sequences involved in the experiment and ξ be any design, then $\Phi(\boldsymbol{M}(\xi)) = -\ln det(\boldsymbol{M}(\xi))$.

Proof of (a):

Let ξ_1^* and ξ_2^* be optimal designs i.e.,

$$\Phi[\boldsymbol{M}(\xi_1^*)] = \Phi[\boldsymbol{M}(\xi_2^*)] = \min_\xi \Phi[\boldsymbol{M}(\xi)]$$

and let $\xi^* = (1-\gamma)\xi_1^* + \gamma\xi_2^*$, for $0 \le \gamma \le 1$. $\Phi[\boldsymbol{M}(\xi)] = -\ln det(\boldsymbol{M}(\xi))$ is convex on set of positive definite matrices (Boyd and Vandenberghe, 2004). Therefore,

$$\Phi[\boldsymbol{M}(\xi^*)] \le (1-\gamma)\Phi[\boldsymbol{M}(\xi_1^*)] + \gamma\Phi[\boldsymbol{M}(\xi_2^*)] = \min_\xi \Phi[\boldsymbol{M}(\xi)],$$

which proves the optimality of ξ^*.

Proof of (b):

We have $\boldsymbol{p_r} = (p_1, p_2, \ldots, p_{k-1})'$ and $\boldsymbol{\delta_1^{(r)}} = (1 - p_1, -p_2, \ldots, -p_{k-1})'$, $\boldsymbol{\delta_2^{(r)}} = (-p_1, 1 - p_2, \ldots, -p_{k-1})', \ldots, \boldsymbol{\delta_{k-1}^{(r)}} = (-p_1, -p_2, \ldots, 1 - p_{k-1})'$.

Hence, $\boldsymbol{p_r} + u\boldsymbol{\delta_1^{(r)}} = (p_1 + u(1 - p_1), (1 - u)p_2, \ldots, (1 - u)p_{k-1})'$, $\boldsymbol{p_r} + u\boldsymbol{\delta_2^{(r)}} = ((1 - u)p_1, p_2 + u(1 - p_2), \ldots, (1 - u)p_{k-1})', \ldots,$ $\boldsymbol{p_r} + u\boldsymbol{\delta_{k-1}^{(r)}} = ((1 - u)p_1, (1 - u)p_2, \ldots, p_{k-1} + u(1 - p_{k-1}))'$.

The determinant of $(\boldsymbol{\delta_1^{(r)}}, \cdots, \boldsymbol{\delta_{k-1}^{(r)}})$ is equal to $1 - (p_1 + p_2 + \cdots + p_{k-1}) = p_k$. Then for design with k treatment sequences we can write $\boldsymbol{M}$ as,

$$\boldsymbol{M}(\boldsymbol{p_r}) = \sum_{j=1}^{k} np_j(\boldsymbol{X_j}^*)^T(\boldsymbol{X_j}^*) = np_1(\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*) + np_2(\boldsymbol{X_2}^*)^T(\boldsymbol{X_2}^*) + \cdots$$

$$+ np_{k-1}(\boldsymbol{X_{k-1}}^*)^T(\boldsymbol{X_{k-1}}^*) + n\left(1 - (p_1 + p_2 + \cdots + p_{k-1})\right)(\boldsymbol{X_k}^*)^T(\boldsymbol{X_k}^*)$$

For illustration purpose consider the direction $\boldsymbol{\delta_1^{(r)}}$, and calculations for other directions can be done similarly,

$$\Phi(\boldsymbol{p_r} + u\boldsymbol{\delta_1^{(r)}}) = -\ln det\left[\boldsymbol{M}\left(\{p_1 + u(1 - p_1), (1 - u)p_2, \ldots, (1 - u)p_{k-1}\}'\right)\right]$$

$$= -\ln det\left[n\{p_1 + u(1 - p_1)\}(\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*)\right.$$

$$\left. + n\{(1-u)p_2\}(\boldsymbol{X_2}^*)^T(\boldsymbol{X_2}^*) + \cdots + n\{(1-u)p_{k-1}\}(\boldsymbol{X_{k-1}}^*)^T(\boldsymbol{X_{k-1}}^*)\right.$$

$$+ n(1 - u)\left\{1 - (p_1 + p_2 + \cdots + p_{k-1})\right\}(\boldsymbol{X_k}^*)^T(\boldsymbol{X_k}^*)]$$
$$= -m\ln n - \ln det[\boldsymbol{M}(u, \boldsymbol{p_r})] = -m\ln n + \Phi^{(r)}(u),$$

where $\boldsymbol{M}(u, \boldsymbol{p_r}) = \frac{\boldsymbol{M}(\boldsymbol{p_r} + u\boldsymbol{\delta_1}^{(r)})}{n}$, and $\Phi^{(r)}(u) = -\ln det[\boldsymbol{M}(u, \boldsymbol{p_r})]$.
The directional derivative of the above objective function along one specific direction for a design with k treatment sequences can be calculated as follows:

$$\phi(u, \boldsymbol{p_r}) = \frac{\partial\Phi(\boldsymbol{p_r} + u\boldsymbol{\delta_1}^{(r)})}{\partial u} = \lim_{h \to 0}\frac{1}{h}\left[\Phi^{(r)}(u + h) - \Phi^{(r)}(u)\right]$$

$$= -\lim_{h \to 0}\frac{1}{h}\left\{\ln det\left[\boldsymbol{M}(u + h, \boldsymbol{p_r})\right] - \ln det\left[\boldsymbol{M}(u, \boldsymbol{p_r})\right]\right\}$$

$$= -\lim_{h \to 0}\frac{1}{h}\left\{\ln det\left[\boldsymbol{M}(u, \boldsymbol{p_r}) + h(1 - p_1)\boldsymbol{X_1}^{*T}\boldsymbol{X_1}^* - hp_2\boldsymbol{X_2}^{*T}\boldsymbol{X_2}^* - \cdots\right.\right.$$
$$\left.\left. - hp_{k-1}\boldsymbol{X_{k-1}}^{*T}\boldsymbol{X_{k-1}}^* - h\left(1 - (p_1 \cdots + p_{k-1})\right)\boldsymbol{X_k}^{*T}\boldsymbol{X_k}^*\right]det\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\right\}$$

$$= -\lim_{h \to 0}\frac{1}{h}\left\{\ln det\left[\boldsymbol{M}(u, \boldsymbol{p_r})\boldsymbol{M}(u, \boldsymbol{p_r})^{-1} + h\left\{\boldsymbol{X_1}^{*T}\boldsymbol{X_1}^* - \boldsymbol{M}(\boldsymbol{p_r})\right\}\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\right]\right\}$$

$$= -\lim_{h \to 0}\frac{1}{h}\left\{\ln det\left[\boldsymbol{I_p} + h\left\{\boldsymbol{X_1}^{*T}\boldsymbol{X_1}^* - \boldsymbol{M}(\boldsymbol{p_r})\right\}\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\right]\right\}$$

Using the approximation of determinant $\det(\boldsymbol{I} + h\boldsymbol{A}) = 1 + h\,\text{trace}(\boldsymbol{A}) + \mathcal{O}(h^2)$ (Bornemann 2010) we get,

$$= -\lim_{h \to 0}\frac{1}{h}\left\{\ln\left(1 + h\,\text{trace}\left[\left\{\boldsymbol{X_1}^{*T}\boldsymbol{X_1}^* - \boldsymbol{M}(\boldsymbol{p_r})\right\}\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\right] + \mathcal{O}(h^2)\right)\right\}$$

And using $\ln(1 + t) = t + \mathcal{O}(t^2)$ we get,

$$= -\lim_{h \to 0}\frac{1}{h}\left\{h\,\text{trace}\left[(\boldsymbol{X_1}^{*T}\boldsymbol{X_1}^* - \boldsymbol{M}(\boldsymbol{p_r}))\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\right] + \mathcal{O}(h^2)\right\}$$

$$= -\text{trace}\left[(\boldsymbol{X_1}^{*T}\boldsymbol{X_1}^* - \boldsymbol{M}(\boldsymbol{p_r}))\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\right]$$

$$= \text{trace}\left(\boldsymbol{M}(\boldsymbol{p_r})\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\right) - \text{trace}\left(\boldsymbol{X_1}^*\boldsymbol{M}(u, \boldsymbol{p_r})^{-1}\boldsymbol{X_1}^{*T}\right)$$

$$\left.\frac{\partial\Phi(\boldsymbol{p_r} + u\boldsymbol{\delta_1}^{(r)})}{\partial u}\right|_{u=0} = m - \text{trace}\left(\boldsymbol{X_1}^*\boldsymbol{M}(\boldsymbol{p_r})^{-1}\boldsymbol{X_1}^{*T}\right) \tag{5}$$

The proof follows by equating the above expression in (5) to zero.

3.2   Equivalence Theorem when Objective Function is Variance of Treatment Effect Estimates   As the main interest usually lies in estimating the

direct treatment effect contrasts, instead of working with the full variance-covariance matrix of parameters estimate, in this section, we concentrate only on the variance of the estimator of treatment effects $\text{Var}(\hat{\boldsymbol{\tau}})$ given as

$$\text{Var}(\hat{\boldsymbol{\tau}}) = \boldsymbol{H}\text{Var}(\hat{\boldsymbol{\theta}})\boldsymbol{H}', \tag{6}$$

where $\boldsymbol{H}$ is a $(t-1) \times m$ matrix given by $[\boldsymbol{0}_{(t-1)1}, \boldsymbol{0}_{(t-1)(p-1)}, \boldsymbol{I}_{t-1}, \boldsymbol{0}_{(t-1)(t-1)}]$ and $m = p + 2t - 2$ is the total number of parameters in $\boldsymbol{\theta}$. Below, we present the equivalence theorem for crossover design when the objective function is a determinant of the variance of treatment effects estimate i.e., the determinant of dispersion matrix.

Lemma 1 *Consider function* $f : \mathbb{R}^n_{>0} \to \mathbb{R}_{>0}$, *such that* $f(\boldsymbol{x}) = \frac{1}{\prod_{i=1}^{n} x_i}$ *where* $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}^n_{>0}$. *Then* $f(\boldsymbol{x})$ *is a strictly convex function.*
Proof of Lemma 1:

Let H be the Hessian matrix, i.e., the matrix of second-order partial derivatives.

Then $H = f(x)(D + qq')$, where D is the diagonal matrix with elements $1/(x_1)^2, \ldots, 1/(x_n)^2$ and q is the column vector with elements $1/(x_1), \ldots, 1/(x_n)$.
The lemma follows as H is positive definite. An alternative proof is provided in the supplementary material.

Theorem 2 *General Equivalence Theorem for Crossover Design when objective function is* $|\boldsymbol{Var(\hat{\tau})}|$: *Consider the design* ξ *with* k *treatment sequences as defined in* (2). *Then, (a) The set of optimal designs is convex. (b) The design* ξ *is D-optimal if and only if*

$$trace\left\{\boldsymbol{A}(\boldsymbol{X_i^*})^T(\boldsymbol{X_i^*})\right\} \begin{cases} = t - 1 \ \ if \ p_i > 0 \\ \leq t - 1 \ \ if \ p_i = 0 \end{cases}$$

for each $p_i \in [0,1]$, *where* $\boldsymbol{A} = \boldsymbol{M}^{-1}\boldsymbol{H}'\left(\boldsymbol{H}\boldsymbol{M}^{-1}\boldsymbol{H}'\right)^{-1}\boldsymbol{H}\boldsymbol{M}^{-1}$, p_i *is the allocation corresponding to point* ω_i *of design* ξ *for all* $i = 1, 2, \ldots, k$, *and* t *is number of treatments.*

Proof of Theorem 3.2: Let k be the number of treatment sequences involved in the experiment and ξ be any design, then $\Phi(\boldsymbol{M}(\xi)) = \ln det(\boldsymbol{H}\boldsymbol{M}(\xi)^{-1}\boldsymbol{H}')$. Proof of (a): Let ξ_1^* and ξ_2^* be optimal designs i.e.,

$$\Phi[\boldsymbol{M}(\xi_1^*)] = \Phi[\boldsymbol{M}(\xi_2^*)] = \min_{\xi} \Phi[\boldsymbol{M}(\xi)]$$

and let $\xi^* = (1 - \gamma)\xi_1^* + \gamma\xi_2^*$, for $0 \leq \gamma \leq 1$.

Since we are using the D-optimality criterion, we prove the following Eq. 7 to prove the optimality of ξ^*.

$$|\boldsymbol{H}\boldsymbol{M}(\xi^*)^{-1}\boldsymbol{H}'| \leq (1 - \gamma)|\boldsymbol{H}\boldsymbol{M}(\xi_1^*)^{-1}\boldsymbol{H}'| + \gamma|\boldsymbol{H}\boldsymbol{M}(\xi_2^*)^{-1}\boldsymbol{H}'|. \qquad (7)$$

Since both $\boldsymbol{M}(\xi_1^*)$ and $\boldsymbol{M}(\xi_2^*)$ are positive definite, we can find a non-singular matrix $\boldsymbol{O}^{-1}$ such that $\boldsymbol{M}(\xi_1^*) = \boldsymbol{O}\boldsymbol{O}^T$ and $\boldsymbol{M}(\xi_2^*) = \boldsymbol{O}\boldsymbol{\Lambda}\boldsymbol{O}^T$, where $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_m\}$ is a $m \times m$ diagonal matrix (see page 41 Rao 1973). In this situation, $\boldsymbol{M}(\xi^*) = \boldsymbol{O}((1 - \gamma)\boldsymbol{I} + \gamma\boldsymbol{\Lambda})\boldsymbol{O}^T$. Then (7) is equivalent to

$$|\boldsymbol{G}((1 - \gamma)\boldsymbol{I} + \gamma\boldsymbol{\Lambda})^{-1}\boldsymbol{G}^T| \leq (1 - \gamma)|\boldsymbol{G}\boldsymbol{G}^T| + \gamma|\boldsymbol{G}\boldsymbol{\Lambda}^{-1}\boldsymbol{G}^T|, \qquad (8)$$

where $\boldsymbol{G} = \boldsymbol{H}(\boldsymbol{O}^T)^{-1}$. According to Theorem 1.1.2 in Fedorov (1972),

$$|\boldsymbol{G}((1 - \gamma)\boldsymbol{I} + \gamma\boldsymbol{\Lambda})^{-1}\boldsymbol{G}^T| = \sum_{1 \leq i_1 < \cdots < i_q \leq m} |\boldsymbol{G}^T[i_1, \ldots, i_q]|^2 \prod_{l=1}^{q} \frac{1}{(1 - \gamma) + \gamma\lambda_{i_l}},$$

where $\boldsymbol{G}^T[i_1, \ldots, i_q]$ is the $q \times q$ sub-matrix of $\boldsymbol{G}^T$ consisting of the $i_1, \ldots, i_q$ rows of $\boldsymbol{G}^T$. Similarly,

$$(1 - \gamma)|\boldsymbol{G}\boldsymbol{G}^T| + \gamma|\boldsymbol{G}\boldsymbol{\Lambda}^{-1}\boldsymbol{G}^T| = \sum_{1 \leq i_1 < \cdots < i_q \leq m} |\boldsymbol{G}^T[i_1, \ldots, i_q]|^2 \left(1 - \gamma + \gamma\prod_{l=1}^{q} \frac{1}{\lambda_{i_l}}\right).$$

Then (8) is true if

$$\prod_{l=1}^{q} \frac{1}{(1 - \gamma) + \gamma\lambda_{i_l}} \leq 1 - \gamma + \gamma\prod_{l=1}^{q} \frac{1}{\lambda_{i_l}}. \qquad (9)$$

Since $f(\boldsymbol{x}) = \frac{1}{\prod_{i=1}^{q} x_i}$ is convex function (from Lemma 1), we have $f((1 - \gamma)\boldsymbol{1} + \gamma\boldsymbol{\lambda}) \leq (1 - \gamma)f(\boldsymbol{1}) + \gamma f(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_{i_1}, \cdots, \lambda_{i_q})$ and hence the result follows. Proof of (b):

$$\boldsymbol{M}(\boldsymbol{p_r}) = np_1(\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*) + np_2(\boldsymbol{X_2}^*)^T(\boldsymbol{X_2}^*) + \cdots + np_{k-1}(\boldsymbol{X_{k-1}}^*)^T(\boldsymbol{X_{k-1}}^*)$$
$$+ n\left(1 - (p_1 + p_2 \cdots + p_{k-1})\right)(\boldsymbol{X_k}^*)^T(\boldsymbol{X_{k-1}}^*).$$
$$\Phi(\boldsymbol{p_r} + u\boldsymbol{\delta_1}^{(r)}) = \Phi\left(\{p_1 + u(1 - p_1), (1 - u)p_2, \ldots, (1 - u)p_{k-1}\}'\right)$$
$$= \ln \det\left[\boldsymbol{H}\left\{\boldsymbol{M}\left(\{p_1 + u(1 - p_1), (1 - u)p_2, \ldots, (1 - u)p_{k-1}\}'\right)\right\}^{-1}\boldsymbol{H}'\right]$$
$$= -(t - 1)\ln n + \ln \det\left[\boldsymbol{H}\left\{\{p_1 + u(1 - p_1)\}(\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*)\right.\right.$$

$$+ \{(1-u)p_2\} \left(X_2^*\right)^T (X_2^*) + \cdots + \{(1-u)p_{k-1}\} \left(X_{k-1}^*\right)^T (X_{k-1}^*)$$

$$+ (1-u) \{1 - (p_1 + p_2 + \cdots + p_{k-1})\} \left(X_k^*\right)^T (X_k^*)\Big\}^{-1} H'\Big]$$

$$= -(t-1)\ln n + \ln det \left[HM(u,p_r)^{-1}H'\right] = -(t-1)\ln n + \Phi^{(r)}(u),$$

where now $\Phi^{(r)}(u) = \ln det \left[HM(u,p_r)^{-1}H'\right]$.

Consider direction $\boldsymbol{\delta}_1^{(r)}$, then the directional derivative of the above objective function for a design with k treatment sequences can be calculated as follows:

$$\phi(u,\boldsymbol{p_r}) = \frac{\partial \Phi(\boldsymbol{p_r} + u\boldsymbol{\delta}_1^{(r)})}{\partial u} = \lim_{h\to 0}\frac{1}{h}\left[\Phi^{(r)}(u+h) - \Phi^{(r)}(u)\right]$$

$$= \lim_{h\to 0}\frac{1}{h}\left\{\ln det\left[HM(u+h,\boldsymbol{p_r})^{-1}H'\right] - \ln det\left[HM(u,\boldsymbol{p_r})^{-1}H'\right]\right\}$$

$$= \lim_{h\to 0}\frac{1}{h}\Big\{\ln det\left[H\left\{(1-\mu-h)M(\boldsymbol{p_r}) + (\mu+h)(X_1^*)^T(X_1^*)\right\}^{-1}H'\right]$$

$$- \ln det\left[HM(\mu,\boldsymbol{p_r})^{-1}H'\right]\Big\}$$

$$= \lim_{h\to 0}\frac{1}{h}\Big\{\ln det\left[H\left\{M(u,\boldsymbol{p_r}) - h\left(M(\boldsymbol{p_r}) - (X_1^*)^T(X_1^*)\right)\right\}^{-1}H'\right]$$

$$- \ln det\left[HM(u,\boldsymbol{p_r})^{-1}H'\right]\Big\}$$

$$= \lim_{h\to 0}\frac{1}{h}\Big\{\ln det\left[H\left\{[M(u,\boldsymbol{p_r})]\left[I - hM(u,\boldsymbol{p_r})^{-1}\left(M(\boldsymbol{p_r}) - (X_1^*)^T(X_1^*)\right)\right]\right\}^{-1}H'\right]$$

$$\times det\left[HM(u,\boldsymbol{p_r})^{-1}H'\right]^{-1}\Big\}$$

$$= \lim_{h\to 0}\frac{1}{h}\Big\{\ln det\left[H\left\{\left[I - hM(u,\boldsymbol{p_r})^{-1}\left(M(\boldsymbol{p_r}) - (X_1^*)^T(X_1^*)\right)\right]^{-1}[M(u,\boldsymbol{p_r})]^{-1}\right\}H'\right]$$

$$\times det\left[HM(u,\boldsymbol{p_r})^{-1}H'\right]^{-1}\Big\}$$

Assuming h is sufficiently small we use the binomial series expansion $(I + hX)^{-1} = \sum_{i=0}^{\infty}(-tX)^i$ to obtain,

$$\phi(u,\boldsymbol{p_r}) = \lim_{h\to 0}\frac{1}{h}\left\{\ln det\left[I + hB + \mathcal{O}(h^2)\right]\right\},$$

$$B = HM(u,\boldsymbol{p_r})^{-1}\left[M(\boldsymbol{p_r}) - (X_1^*)^T(X_1^*)\right]M(u,\boldsymbol{p_r})^{-1}H'\left[HM(u,\boldsymbol{p_r})^{-1}H'\right]^{-1}.$$

Using $\ln det \left[\boldsymbol{I} + h\boldsymbol{B} + \mathcal{O}(h^2) \right] = h\text{trace}(\boldsymbol{B}) + \mathcal{O}(h^2)$ (Withers and Nadarajah 2010),

$$
\begin{aligned}
\phi(u, \boldsymbol{p_r}) \quad &= \text{trace}\{\boldsymbol{HM}(u, \boldsymbol{p_r})^{-1} \left[\boldsymbol{M}(\boldsymbol{p_r}) - (\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*) \right] \boldsymbol{M}(u, \boldsymbol{p_r})^{-1} \boldsymbol{H'} \\
&\qquad \times \left[\boldsymbol{HM}(u, \boldsymbol{p_r})^{-1} \boldsymbol{H'} \right]^{-1} \} \\
\phi(u, \boldsymbol{p_r})|_{u=0} \quad &= \text{trace}\{\boldsymbol{HM}(\boldsymbol{p_r})^{-1} \left[\boldsymbol{M}(\boldsymbol{p_r}) - (\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*) \right] \boldsymbol{M}(\boldsymbol{p_r})^{-1} \boldsymbol{H'} \\
&\qquad \times \left[\boldsymbol{HM}(\boldsymbol{p_r})^{-1} \boldsymbol{H'} \right]^{-1} \} \\
&= \text{trace}\Big\{ \boldsymbol{I}_{(t-1)} - \boldsymbol{HM}(\boldsymbol{p_r})^{-1}(\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*)\boldsymbol{M}(\boldsymbol{p_r})^{-1} \boldsymbol{H'} \\
&\qquad \times \left[\boldsymbol{HM}(\boldsymbol{p_r})^{-1}\boldsymbol{H'} \right]^{-1} \Big\} \\
&= (t-1) - \text{trace}\Big\{ \boldsymbol{HM}(\boldsymbol{p_r})^{-1}(\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*)\boldsymbol{M}(\boldsymbol{p_r})^{-1}\boldsymbol{H'} \left(\boldsymbol{HM}(\boldsymbol{p_r})^{-1}\boldsymbol{H'} \right)^{-1} \Big\} \\
&= (t-1) \qquad - \text{trace}\Big\{ \left[\boldsymbol{M}^{-1}\boldsymbol{H'} \left(\boldsymbol{HM}^{-1}\boldsymbol{H'} \right)^{-1} \boldsymbol{HM}^{-1} \right] (\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*) \Big\} \qquad (10)
\end{aligned}
$$

The proof follows by equating the above expression in (10) to zero.

    3.3 Illustration   To illustrate the results of the above general equivalence theorems, we consider a design space $\{AB, BA\}$ has $k = 2, p = 2$. Since we are considering a local optimality approach, for illustration purposes we assume that the parameter values are $\boldsymbol{\theta} = (\lambda, \beta_2, \tau_B, \rho_B)' = (0.5, -1.0, 4.0, -2.0)'$. Note that we need to assume parameter values before calculating the optimal proportions. Considering the AR(1) correlation structure with $\alpha = 0.1$, i.e.,

$$
\boldsymbol{C_\alpha} = \left(\alpha^{|i-i'|} \right) = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix},
$$

for the assumed parameter values the optimal proportions are $p_1 = p_2 = 0.5$. The graph of the objective function, $\Phi(p_1) = -\ln det(\boldsymbol{M}(p_1))$ and its directional derivative $trace\left(\boldsymbol{X_1}^*\boldsymbol{M}(p_1)^{-1}\boldsymbol{X_1}^{*T} \right) - m$ w.r.t $p_1 \in [0,1]$ are shown in Fig. 1.

    Graphs in Fig. 1 verify that the minimum of the objective function is located at $p_1 = 0.5$ and directional derivative is zero at $p_1 = 0.5$. Using Theorem 1, we conclude that for assumed values of parameters, design

$$
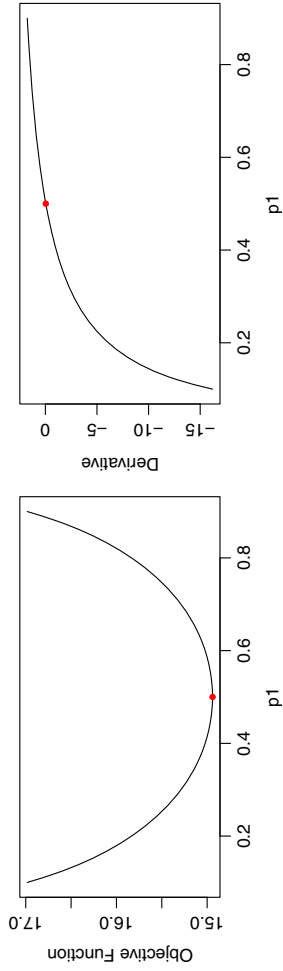\xi = \left\{ \begin{matrix} AB & BA \\ 0.5 & 0.5 \end{matrix} \right\}
$$

Figure 1: Objective function and its directional derivative for designs with two treatment sequences

is the D-optimal design when the objective function is $Var(\hat{\boldsymbol{\theta}})$.

Considering $Var(\hat{\boldsymbol{\tau}})$ as the objective function, the graph of the objective function, $\Phi(p_1) = \ln det[\boldsymbol{H}\boldsymbol{M}(p_1)^{-1}\boldsymbol{H}']$ and it's directional derivative $trace\left\{\boldsymbol{A}(\boldsymbol{X_1}^*)^T(\boldsymbol{X_1}^*)\right\} - (t-1)$ w.r.t $p_1 \in [0,1]$ are shown in Fig. 2.

Graphs in Fig. 2 verify that the minimum of the objective function is located at $p_1 = 0.177$ and directional derivative is zero at $p_1 = 0.177$. Using Theorem 2, we conclude that for assumed values of parameters, design

$$\xi = \left\{ \begin{matrix} AB & BA \\ 0.177 & 0.823 \end{matrix} \right\}$$

is the D-optimal design when the objective function is $Var(\hat{\boldsymbol{\tau}})$.

4 Real Example

In this section, we look at the application of the above equivalence theorems to a real-life example discussed earlier. We obtain the D-optimal designs by solving a system of equations given by the general equivalence theorems when the objective functions are $Var(\hat{\boldsymbol{\theta}})$ and $Var(\hat{\boldsymbol{\tau}})$, respectively.

4.1 Work Environment Experiment Consider the data obtained from the work environment experiment conducted at Booking.com (Pitchforth et al. 2020). In recent years, many corporate offices and organizations have adopted open office spaces over traditional cubical office spaces. Since there were no previous studies to examine the effects of office designs in workspaces, Booking.com conducted an experiment to assess different office spacing efficiency.

In this experiment, there were a total of $n = 288$ participants. Participants were divided into four groups G_1, G_2, G_3, G_4 with each group having an equal number of 72 individual participants. It is essentially a uniform crossover design with $p = 4$ periods and $t = 4$ treatments. Periods were named Wave 1, Wave 2, Wave 3, and Wave 4, where each Wave had a duration of 2 weeks. The four treatments involved in this experiment are office designs named A (Activity-Based), B (Open Plan), C (Team Offices), and D (Zoned Open Plan), as shown in the Fig. 3. During the experiment, each group is exposed to different treatments over different periods depending on its treatment sequence. At a particular given period, there was no interaction between subjects from different groups. A Latin square design (see, for example, Wu and Hamada (2009)) of order four has been used to determine the sequence of exposure so that no group was exposed to the conditions in
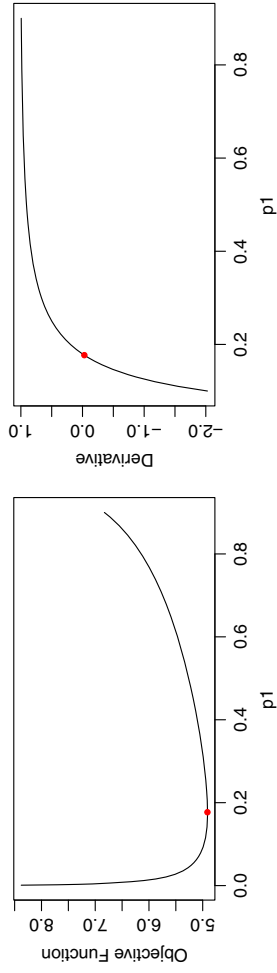
Figure 2: Objective function and its directional derivative for designs with two treatment sequences
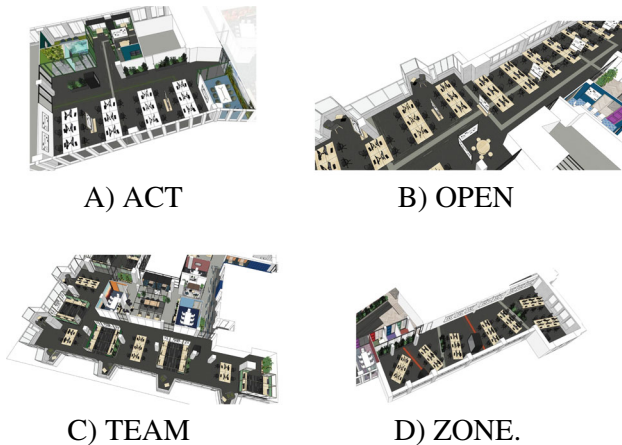
A) ACT

B) OPEN

C) TEAM

D) ZONE.

Figure 3: The four office designs involved in Work Environment Experiment

the same order as any other group. The design is shown below in Table 1. A total of $m = 23$ covariates was involved in the experiment, but we consider only the most important ones in our fitted model.

The images are reproduced from the manuscript (Pitchforth et al., 2020), under Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/).

For illustration purposes, consider the response *commit count* to illustrate the optimal crossover design for the Poisson response. The commit count is the number of commits submitted to the main git repository by each subject. In the fitted model, we examine three primary predictors: *area*, *wave*, and *carryover*. Here, *area* represents the direct treatment effect, *wave* denotes the period effect, and *carryover* represents the effect of the treatment from the previous period. To illustrate the local optimality approach, we assume specific parameter values $\boldsymbol{\theta} = (2.0,\ 0.3,\ 0.8,\ -0.1,\ -2.0,\ 0.40,$

Table 1: Latin square design

Groups ⇒ Period ⇓	G_1 (*BADC*)	G_2 (*CDAB*)	G_3 (*DBCA*)	G_4 (*ACBD*)
Wave 1	OPEN	TEAM	ZONE	ACT
Wave 2	ACT	ZONE	OPEN	TEAM
Wave 3	ZONE	ACT	TEAM	OPEN
Wave 4	TEAM	OPEN	ACT	ZONE

$-2.0, -1.0, 0.3, -1.0)'$, which lead to non-uniform allocations using the log link function and AR(1) correlation structure with $\alpha = 0.1$.

According to Theorem 3.1, the D-optimal design, i.e., the optimal proportions, can be obtained by solving the following system of equations instead of performing constrained optimization:

$$\text{trace}\left(\boldsymbol{X_i}^* \boldsymbol{M}(\boldsymbol{p_r})^{-1} \boldsymbol{X_i}^{*T}\right) = 10,$$

for $i = 1, 2, 3, 4$. The resulting D-optimal design is the same as the one obtained through constrained optimization, indicating that the design is given by:

$$\xi = \left\{ \begin{array}{cccc} BADC\ CDAB\ DBCA\ ACBD \\ 0.2375\ \ 0.2894\ \ 0.2246\ \ 0.2485 \end{array} \right\}$$

is the D-optimal design when the objective function is $Var(\hat{\boldsymbol{\theta}})$.

Similarly, according to Theorem 3.2, for the objective function $Var(\hat{\boldsymbol{\tau}})$, the D-optimal design can be obtained by solving the following system of equations:

$$\text{trace}\left\{\left[\boldsymbol{M}(\boldsymbol{p_r})^{-1}\boldsymbol{H}'\left(\boldsymbol{H}\boldsymbol{M}(\boldsymbol{p_r})^{-1}\boldsymbol{H}'\right)^{-1}\boldsymbol{H}\boldsymbol{M}(\boldsymbol{p_r})^{-1}\right](\boldsymbol{X_i}^*)^T(\boldsymbol{X_i}^*)\right\} = 3,$$

for $i = 1, 2, 3, 4$. Again, the resulting D-optimal design is the same as the one obtained through constrained optimization, indicating that the design is given by:

$$\xi = \left\{ \begin{array}{cccc} BADC\ CDAB\ DBCA\ ACBD \\ 0.2900\ \ 0.2963\ \ 0.1734\ \ 0.2403 \end{array} \right\}$$

is the D-optimal design.

Remark 3 In Jankar et al. (2020), we study the effect of misspecification of working correlation structures on optimal design. We calculate optimal designs under two choices of unknown parameters for a misspecified working correlation structure. Then we calculate relative D-efficiency under two parameter choices. The relative D-efficiency under two parameter choices suggests that the effect of variance misspecification on the local optimal designs is minimal. We also study the performance of proposed locally optimal designs via sensitivity study in terms of relative loss of efficiency for choosing assumed parameter values instead of true parameter values. The relative loss of efficiency increases as we move away from true parameter values. However, **Fig 6.** in Jankar et al. (2020) suggest that this loss of efficiency does not go beyond 2%. We also calculate the optimal designs with

all 24 sequences, by considering AR (1) correlation structure and different values of α. We observe that in the case of non-uniform allocations, the optimal design has more zeros than non-zero proportions; and these allocations do not vary a lot as α changes, particularly for the sequences where we have zero allocations.

5 Summary and Conclusion

In many experiments in real life, uniform designs are typically used. Uniform designs are optimal in the case of a linear model i.e., when the response is normally distributed. However, in situations where responses are non-normal, the obtained optimal proportions are not necessarily uniform. In this manuscript, we derive an expression for the general equivalence theorem to check for the optimality of identified locally D-optimal crossover designs for generalized linear models. The equivalence theorem provides us with a system of equations that can calculate optimal proportions without performing constrained optimization of the objective function. We derive two different versions of the general equivalence theorem, one with the objective function $Var(\hat{\boldsymbol{\theta}})$ and the other with the objective function $Var(\hat{\boldsymbol{\tau}})$. We illustrate the application of these equivalence theorems on two real-life examples and obtain the same set of optimal proportions by solving the system of equations as obtained by performing constrained optimization. In our future work, we plan to use the Bayesian approach to avoid guessing the values of unknown parameters.

Declarations

Conflicts of interest There is no conflict of interest.

References

K. Afsarinejad. Repeated measurement designs - a review. Communications in Statistics - Theory and Methods., 19(11):3985–4028, 1990.

Anthony Atkinson, Alexander Donev, and Randall Tobias. Optimum Experimental Designs, with SAS. 01 2007.

R. Bailey and J. Kunert. On optimal crossover designs when carryover effects are proportional to direct effects. Biometrika, 93:613–625, 2006

Folkmar Bornemann. On the numerical evaluation of Fredholm determinants. Mathematics of Computation, 79(270):871–915, 2010.

Mausumi Bose and Aloke Dey. Optimal Crossover Designs. World Scientific, 2009.

Stephen P Boyd and Lieven Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

K. C. Carriere and Rong Huang. Crossover designs for two-treatment clinical trials. Journal of Statistical Planning and Inference, (87):125–134, 2002.

V.V. Fedorov. The design of experiments in the multiresponse case. Theory Probab Appl, (16):323–332, 1971.

V.V. Fedorov. Theory of Optimal Experiment. Academic Press, New York, 1972.

V.V. Fedorov and S.L. Leonov. Optimal Design for Nonlinear Response Models. Chapman & Hall/CRC, 2014.

V.V. Fedorov and M.B. Malyutov. Optimal designs in regression problems. Math. Operat. Statist., (3):281–308, 1972.

A. S. Hedayat and Min Yang. Optimal and efficient crossover designs for comparing test treatments with a control treatment. The Annals of Statistics, 33(2):915–943, 2005

Jeevan Jankar, Abhyuday Mandal, and Jie Yang. Optimal crossover designs for generalized linear models. Journal of Statistics Theory and Practice, 14(23), 2020.

B. Jones and M.G. Kenward. Design and Analysis of Cross-Over Trials. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2014.

M. G. Kenward and B. Jones. Alternative approaches to the analysis of binary and categorical repeated measurements. Journal of Biopharmaceutical Statistics, 2(2):137–170, 1992. PMID: 1300210.

Ronald P. Kershner and Walter T. Federer. Two-treatment crossover designs for estimating a variety of effects. Journal of the American Statistical Association, 76(375):612–619, 1981.

J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. Canadian Journal of Mathematics, 12:363–366, 1960.

Eugene M. Laska and Morris Meisner. A variational approach to optimal two-treatment crossover designs: Application to carryover-effect models. Journal of the American Statistical Association, 80(391):704–710, 1985.

J. N. S. Matthews. Recent developments in crossover designs. International Statistical Review / Revue Internationale de Statistique, 56(2):117–127, 1988.

J Pitchforth, E Nelson White, M van den Helder, and Oosting W. The work environment pilot: An experiment to determine the optimal office design for a technology company. PLoS ONE, 15(5), 2020.

Ross L. Prentice. Correlated binary regression with covariates specific to each binary observation. Biometrics, 44(4):1033–s1048, 1988.

C. Radhakrishna Rao. Linear Statistical Inference and its Applications. John Wiley Sons, Ltd, 1973.

Karlin S and Studden W.J. Tchebycheff systems: With applications in analysis and statistics. New York: Interscience Publishers, 1966.

S.S. Senn. Cross-over Trials in Clinical Research. Statistics in Practice. c, 2002.

S D Silvey. Optimal Design. 1980.

Christopher S. Withers and Saralees Nadarajah. Expansion for functions of determinants of power series. Canadian Applied Mathematics Quarterly, 18(1):107–114, 2010.

Cini Varghese, AR Rao, and V. Sharma. Robustness of Williams Square changeover designs. Metrika, 55:199–208, 2002.

P. Whittle and M.B. Malyutov. Some general points in the theory of optimal experimental design. J. Roy. Statist. Soc. Ser., (35):123–s130, 1973.

C.F.J. Wu and M. Hamada. Experiments: Planning, Analysis, and Parameter Design Optimization. Wiley India Pvt. Limited, 2009.

Jie Yang, Abhyuday Mandal, and Dibyen Majumdar. Optimal design for 2^k factorial experiments with binary response. Statistica Sinica, 26(1):385–411, 2016.

S. Zeger, K. Liang, and P. Albert. Models for longitudinal data: a generalized estimating equation approach. Biometrics, 44 4:1049–60, 1988.

Christopher Eager and Joseph Roy. Mixed Effects Models are Sometimes Terrible, 2017.

Fitzmaurice G.M., Laird N.M., and Ware J.H. Applied Longitudinal Analysis. Wiley Series in Probability and Statistics. Wiley, Hoboken, N.J., 2011.

Stroup, W.W., Milliken, G.A., Claasen, E.A., and Wolfinger, R.D. SAS for Mixed Models: Introduction and Basic Applications. SAS Institute, Cary, NC, 2018.

JEEVAN JANKAR
DEPARTMENT OF STATISTICS,
UNIVERSITY OF GEORGIA, ATHENS 30602
GA, GEORGIA
E-mail: jeevan.jankar@uga.edu

JIE YANG
DEPARTMENT OF MATHEMATICS,
STATISTICS, AND COMPUTER SCIENCE,

UNIVERSITY OF ILLINOIS AT CHICAGO,
CHICAGO IL 60607, USA
E-mail: jyang06@uic.edu

ABHYUDAY MANDAL
DEPARTMENT OF STATISTICS,
UNIVERSITY OF GEORGIA, ATHENS 30602
GA, GEORGIA
E-mail: amandal@stat.uga.edu