Journal of Statistical Theory and Practice (2020) 14:23
https://doi.org/10.1007/s42519-020-00089-5

ORIGINAL ARTICLE

Optimal Crossover Designs for Generalized Linear Models

Jeevan Jankar[1] · **Abhyuday Mandal**[1] · **Jie Yang**[2]

Published online: 31 March 2020

Abstract
We identify locally *D*-optimal crossover designs for generalized linear models. We use generalized estimating equations to estimate the model parameters along with their variances. To capture the dependency among the observations coming from the same subject, we propose six different correlation structures. We identify the optimal allocations of units for different sequences of treatments. For two-treatment crossover designs, we show via simulations that the optimal allocations are reasonably robust to different choices of the correlation structures. We discuss a real example of multiple-treatment crossover experiments using Latin square designs. Using a simulation study, we show that a two-stage design with our locally *D*-optimal design at the second stage is more efficient than the uniform design, especially when the responses from the same subject are correlated.

Keywords Approximate designs · *D*-optimality · Compound symmetric correlation · AR(1) correlation structure · Generalized estimating equations · Two-stage design

1 Introduction

Pharmaceutical companies frequently conduct clinical trials where the outcome is either success or failure of a particular therapy. Crossover designs, also known as repeated measurements designs or changeover designs, have been used extensively in pharmaceutical research. There is a rich literature on optimal crossover designs

Part of special issue guest edited by Pritam Ranjan and Min Yang—Algorithms, Analysis and Advanced Methodologies in the Design of Experiments.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s42519-020-00089-5) contains supplementary material, which is available to authorized users.

✉ Abhyuday Mandal
amandal@stat.uga.edu

[1] Department of Statistics, University of Georgia, Athens, GA 30602, USA

[2] Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA

when the response can be adequately modeled by normal distributions. However, for a binary outcome, where the response needs to be described using generalized linear models (GLMs), limited results are known. Consequently, these trials are usually designed using the guidelines of traditional crossover designs obtained using the theory of linear models. However, these designs can be quite inefficient for GLMs. Our goal is to bridge this gap in the literature and determine efficient designs specifically for crossover experiments with responses under univariate GLMs, including binary, binomial, Poisson, gamma, inverse Gaussian responses, etc.

Among different types of experiments that are available for treatment comparisons with multiple periods, the crossover designs are among the most important ones. In these experiments, every subject is exposed to a sequence of treatments over different time periods, i.e., subjects crossover from one treatment to another. One of the most important aspects of crossover designs is that we can get the same number of observations as other designs but with less number of subjects. This is an important consideration since human participants are often scarce in clinical trials. The order in which treatments are applied to subjects is known as a *sequence*, and the time at which these sequences are applied is known as a *period*. In most of the cases, the main aim of such experiments is to compare t treatments over p periods. In each period, each subject receives a treatment, and the corresponding response is recorded. In different periods, a subject may receive different treatments, but treatment may also be repeated on the same subject. Naturally, crossover designs also provide within-subject information about treatment differences.

Most of the research in the crossover design literature dealt with continuous response variables (see, for example, Kershner and Federer [10], Laska and Meisner [15], Matthews [18], Carriere and Huang [3] and the references therein). The problem of determining optimal crossover designs for continuous responses has been studied extensively (see, for example, Bose and Dey [1, 2], for a review of results). For examples of practical cases where the responses are discrete in nature, such as binary responses, one may refer to Jones and Kenward [9] and Senn [21].

Among many fixed effects models proposed in the literature, the following linear model is used extensively to formulate crossover designs.

$$Y_{ij} = \lambda + \beta_i + \alpha_j + \tau_{d(i,j)} + \rho_{d(i-1,j)} + \epsilon_{ij}, \tag{1}$$

where Y_{ij} is the observation from the jth subject in the ith time period, with $i = 1, \ldots, p$ and $j = 1, \ldots, n$. Here, $d(i, j)$ stands for the treatment assignment to the jth subject at time period i and $\lambda, \beta_i, \alpha_j, \tau_{d(i,j)}, \rho_{d(i-1,j)}$ are the corresponding overall mean, the ith period effect, the jth subject effect, the direct treatment effect and the carryover treatment effect, respectively. Here, ϵ_{ij}'s are the uncorrelated error terms which follow a normal distribution with zero mean and constant variance. Model (1) is sometimes referred to as the traditional model due to its extensive use in the literature.

As all the effects are fixed, for the linear model (1), the Fisher information matrix is independent of model parameters. Various optimality criteria such as A-, D-, E-optimality depend on this information matrix (see, for example, Pukelsheim [20]). Numerous results corresponding to the optimality of crossover designs for

linear models are available in the literature. Hedayat and Afsarinejad [7], Cheng and Wu [4] and Kunert [13] studied the optimality of balanced, uniform designs. Cheng and Wu [4] formulated theorems for the optimality of a strongly balanced design. Kunert [12] produced results for optimality of designs that are neither balanced nor strongly balanced. Similar results can also be found in Stufken [23]. Dey et al. [6] were among the first ones to provide results for optimality of designs when $p \leq t$. Considering arbitrary p and t with both $p \leq t$ and $p \geq t$, Kushner [14] obtained conditions for universal optimality through approximate theory. Such results cannot be readily extended for binary responses since the Fisher information matrix for GLMs depends on the model parameters [19, 24]. In this paper, we focus on local optimality to circumvent this problem [11].

This paper is organized as follows. We describe a preliminary setup of a model for crossover designs for GLMs in Sect. 2.1 and then discuss generalized estimating equations in Sect. 2.2. We propose different correlation structures in Sect. 2.3 and formulate locally optimal crossover designs along with an algorithm for obtaining such designs, in Sect. 2.4. In Sect. 3, we provide examples of optimal design for two-treatment crossover trials. We calculate optimal designs for examples with binary response in Sect. 3.1 and for example with Poisson response in Sect. 3.2. In Sect. 4.1, we provide examples of optimal designs for multi-treatment crossover trials, where we use Latin square design. Sensitivity study and relative D-efficiency are presented in Sect. 4.2. Simulation studies are presented in Sect. 4.3. The paper concludes with comments in Sect. 5. Some technical details and additional results are presented in "Appendix" and Supplementary Materials.

2 Crossover Designs for GLM

Although there is a rich literature on optimal crossover designs for linear models, the results on crossover designs under generalized linear models (GLMs) are meager. Before identifying optimal crossover design, we first formally introduce the GLM and the associated optimal crossover designs.

2.1 Preliminary Setup

We consider a crossover trial with t treatments, n subjects and p periods. The responses obtained from these n subjects are denoted as $Y_1, \ldots, Y_n$, where the response from the jth subject is $Y_j = (Y_{1j}, \ldots, Y_{pj})'$. As discussed above, we use a generalized linear model (GLM) to describe the marginal distribution of Y_{ij} as in Liang and Zeger [17]. Let μ_{ij} denote the mean of a binary response Y_{ij}. To fix ideas, first we consider the logistic regression, which models the marginal mean μ_{ij} for crossover trial as

$$\text{logit}(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \eta_{ij} = \lambda + \beta_i + \tau_{d(i,j)} + \rho_{d(i-1,j)}, \tag{2}$$

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 $s = 1, \ldots, t$.

Remark 1 Unlike model (1), model (2) does not contain a subject effect term α_j. Note that the response here is described by a GLM, where the Fisher information matrix depends on model parameters. In this paper, we consider the local optimality approach of Chernoff [5], in which the parameters are replaced by assumed values. In the linear model, the subject effect can be estimated from the data, but for our local optimality approach for the GLM, an educated guess for the subject effect is needed. It would be reasonable to guess the fixed treatment effects from prior knowledge, while from a design point of view the subject effect, if included, has to be treated as random. Instead of incorporating a random effects term, in this paper, the mean response is modeled through the logit link function in Eq. (2) with an extra assumption that the responses from a particular subject are mutually correlated, while the responses from different subjects are uncorrelated. In the case of generalized linear models, only the mean response is modeled through the link function, and hence, we are free to choose a variance–covariance matrix as long as that is positive definite. So, in this paper, we use this opportunity of choosing the covariance matrix and capture the subject effect by putting different meaningful structures on this matrix and studying the robustness of the design. In this way, we can exclude a random subject effect from the model and calculate optimal designs more easily.

As the main interest is in estimating the treatment effects and variance of its estimator, carryover effects are treated as nuisance parameters. To ensure estimability of the model parameters, we set the baseline constraints as $\beta_1 = \tau_1 = \rho_1 = 0$. Consider $\beta = (\beta_2, \ldots, \beta_p)'$, $\tau = (\tau_2, \ldots, \tau_t)'$ and $\rho = (\rho_2, \ldots, \rho_t)'$, which define the parameter vector $\theta = (\lambda, \beta, \tau, \rho)'$. Then, the linear predictor corresponding to the *jth* subject, $\eta_j = (\eta_{1j}, \ldots, \eta_{pj})'$, can be written as

$$\eta_j = X_j\theta.$$

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

If the number of subjects is fixed to n and the number of periods is p, then we determine the proportion of subjects assigned to a particular treatment sequence. As the number of periods is fixed to p, 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 Ω 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 $n = \Sigma_{\omega \in \Omega} n_\omega, n_\omega \geq 0$. A crossover design ζ in approximate theory is specified by

the set $\{p_\omega, \omega \in \Omega\}$, where $p_\omega = n_\omega/n$ is the proportion of subjects assigned to treatment sequence ω. Such a crossover design ζ can be denoted as follows:

$$\zeta = \begin{Bmatrix} \omega_1 & \omega_2 & \dots & \omega_k \\ p_{\omega_1} & p_{\omega_2} & \cdots & p_{\omega_k} \end{Bmatrix}$$

where k is the number of treatment sequences involved, such that $\sum_{i=1}^{k} p_{\omega_i} = 1$, for $i = 1, \dots, k$. From the definitions of matrices T_j and F_j, it can be noted that they depend only on the treatments sequence ω that subject j receives. So it can be inferred that $T_j = T_\omega$ and $F_j = F_\omega$. This implies, $X_j = X_\omega$ as $P_j = [0_{(p-1)1}, I_{p-1}]'$.

2.2 Generalized Estimating Equations

Generalized estimating equations are quasi-likelihood equations which allow us to estimate quasi-likelihood estimators. In this paper, instead of using maximum likelihood estimation (MLE) or ordinary least squares (OLS) to estimate the parameters, we use quasi-likelihood estimation. Earlier we made one important assumption in crossover trials that observations from each subject are mutually correlated, while the observations from different subjects are uncorrelated. This dependency between repeated observations from a subject is modeled using what is called "working correlation" matrix C. If C is the true correlation matrix of Y_j, then from the definition of covariance we can write

$$\text{Cov}(Y_j) = D_j^{1/2} C D_j^{1/2},$$

where $D_j = \text{diag}\Big(\mu_{1j}(1-\mu_{1j}), \dots, \mu_{pj}(1-\mu_{pj})\Big)$. Let us denote $\text{Cov}(Y_j)$ by W_j. In Zeger et al. [25, Eq. (3.1)], it has been shown that for repeated measurement model, the generalized estimating equations (GEE) are defined to be

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

where $\mu_j = \left(\mu_{1j}, \dots, \mu_{pj}\right)'$ and the asymptotic variance for the GEE estimator $\hat{\theta}$ (see Zeger et al. [25], Eq. (3.2)) is

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

where $W_j = \text{Cov}(Y_j)$. As mentioned by Singh and Mukhopadhyay [22] in the paper (Zeger et al. [25], Eq. (3.2)), it has also been shown that if the true correlation structure varies from "working correlation" structure, then $\text{Var}(\hat{\theta})$ is given by the sandwich formula

$$\text{Var}(\hat{\theta}) = U^{-1} V U^{-1},$$

where the U and V in above equation are as follows:

$$U = \sum_{\omega \epsilon \Omega} np_\omega \frac{\partial \mu'_\omega}{\partial \theta} W_\omega^{-1} \frac{\partial \mu_\omega}{\partial \theta}, \quad V = \sum_{\omega \epsilon \Omega} np_\omega \frac{\partial \mu'_\omega}{\partial \theta} W_\omega^{-1} \mathrm{Cov}(Y_\omega) W_\omega^{-1} \frac{\partial \mu_\omega}{\partial \theta}. \tag{4}$$

So it is expected that the effect of variance misspecification on the locally optimal designs will be minimal. Table 10 presented in "Appendix" confirms this.

From Eqs. (3) and (4), it can be seen that if the true correlation of Y_j is equal to C, then $\mathrm{Var}(\hat{\theta}) = U^{-1}$.We have considered carryover effects to be nuisance parameters as the main interest usually lies in estimating the direct treatment effect contrasts. So, instead of working with the full variance–covariance matrix of parameter estimator $\hat{\theta}$, we concentrate only on the variance of the estimator of treatment effect $\mathrm{Var}(\hat{\tau})$ where

$$\mathrm{Var}(\hat{\tau}) = H \mathrm{Var}(\hat{\theta}) H', \tag{5}$$

H is a $(t-1) \times m$ matrix given by $[0_{(t-1)1}, 0_{(t-1)(p-1)}, I_{t-1}, 0_{(t-1)(t-1)}]$ where $m = p + 2t - 2$ is the total number of parameters in θ and $0_{(t-1)(p-1)}$ is a $(t-1) \times (p-1)$ matrix of zeros.

We calculate optimal proportions such that the variances of estimators of treatment effect are minimized. In this paper, we focus on D-optimality and use the determinant of $\mathrm{Var}(\hat{\tau})$ as our objective function. Note that other optimality criteria such as A-, E-optimality can be applied similarly. Then, an optimal design ζ^* minimizes the determinant of $\mathrm{Var}(\hat{\tau})$ in Eq. (5) with respect to p_ω such that $\sum_{w \in \Omega} p_w = 1$. For illustration, we give an explicit expression of the information matrix and present the associated calculations for a crossover design in the Supplementary Materials.

2.3 Proposed Correlation Structures

As mentioned in the above section, to calculate the variance matrix of parameter estimates, a predefined working correlation structure for the responses is needed. Any correlation structure can be assumed for the responses, but if the design is not robust, then the optimal proportions will vary as the correlation structure varies. So, to check the robustness of design and to make the design more practically acceptable, optimal proportions using different correlation structures are calculated. For the design in Eq. (2) with two treatments A and B, six different types of correlation structures are proposed, and optimal proportions are calculated. Out of these six correlation structures, the correlation matrices defined by the first three correlation structures are fixed and do not depend on treatment sequence, whereas the correlation matrices of the fourth, fifth and sixth types depend on treatment sequences and vary along with treatment sequences.

The first correlation structure is a compound symmetric correlation structure, i.e.,

$$\mathrm{Corr}(1) = (1 - \rho) I_p + \rho J_p,$$

where I_p is the identity matrix of order p, and J_p is a $p \times p$ matrix with all elements unity.

The second correlation structure is the AR(1) correlation structure, i.e.,

$$\text{Corr}(2) = \left(\rho^{|i-i'|}\right),$$

so that the correlation between responses decreases as the time gap between responses increases.

The third correlation structure is as follows:

$$\text{Corr}(3) = \begin{pmatrix} 1 & \rho & 0 & \dots & 0 & 0 & 0 \\ \rho & 1 & \rho & \dots & 0 & 0 & 0 \\ \vdots & & & \vdots & & & \vdots \\ 0 & 0 & 0 & \dots & \rho & 1 & \rho \\ 0 & 0 & 0 & \dots & 0 & \rho & 1 \end{pmatrix}.$$

For each correlation structure different correlation matrices using different ρ values are considered.

To understand the other three correlation structures, we denote the correlation coefficient between the response when a subject receives treatment A first and the response when the same subject receives treatment B afterward as ρ_{AB} and, ρ_{BA} when the subject receives B first and A afterward. Note that in general ρ_{AB} is not necessarily the same as ρ_{BA}. In a similar manner, we define ρ_{AA} and ρ_{BB}. To define the fourth type of correlation structure, we will use the same structure as Corr(3) but with different values of correlation coefficient for different treatment sequences. For fourth type of correlation, we use $\rho_{AB} = 0.2, \rho_{BA} = 0.5$ and $\rho_{AA} = 0.1, \rho_{BB} = 0.3$.

To define fifth and sixth type of correlation structures, we use AR(1) correlation structure with correlation coefficient depending on treatment sequence. For the fifth type, we use the same values for ρ_{AB} and ρ_{BA}, and for the sixth type of correlation structure, we use different values for ρ_{AB} and ρ_{BA}. For both fifth and sixth type of correlation structure, we keep $\rho_{AA} = \rho_{BB}$. These values might vary from example to example and would depend on what treatments A and B are. As the entries of the correlation matrix depend on which treatment the subject receives in a particular period, these correlation matrices are different for different treatment sequences. Here, our aim is to see how optimal proportions vary as we vary values of ρ_{AB} and ρ_{BA}.

As an illustration, we consider $p = 2$ with treatment sequences AB, BA. Then, the third type correlation matrices for both treatment sequences AB and BA will have same structure as Corr(1). The fourth, fifth and sixth type correlation matrices will have same structure as follows with different ρ values,

$$\text{Corr}(4/5/6)_{AB} = \begin{pmatrix} 1 & \rho_{AB} \\ \rho_{AB} & 1 \end{pmatrix},$$

$$\text{Corr}(4/5/6)_{BA} = \begin{pmatrix} 1 & \rho_{BA} \\ \rho_{BA} & 1 \end{pmatrix}.$$

For $p = 3$ case, we consider an example with treatment sequences ABB, BAA. The fourth type of correlation matrix will have values as mentioned above. The fifth type

correlation matrices for both treatment sequences *ABB* and *BAA* will be the same if in treatment sequences, A and B are interchangeable and $\rho_{AB} = \rho_{BA}$ along with $\rho_{AA} = \rho_{BB}$. The sixth type correlation matrices for both treatment sequences *ABB* and *BAA* will be different as ρ_{AB} and ρ_{BA} are different. We get

$$\text{Corr}(4)_{ABB} = \begin{pmatrix} 1 & \rho_{AB} & 0 \\ \rho_{AB} & 1 & \rho_{BB} \\ 0 & \rho_{BB} & 1 \end{pmatrix},$$

$$\text{Corr}(4)_{BAA} = \begin{pmatrix} 1 & \rho_{BA} & 0 \\ \rho_{BA} & 1 & \rho_{AA} \\ 0 & \rho_{AA} & 1 \end{pmatrix},$$

and

$$\text{Corr}(5)_{ABB} = \text{Corr}(5)_{BAA} = \begin{pmatrix} 1 & \rho_{AB} & \rho_{AB}^2 \\ \rho_{AB} & 1 & \rho_{BB} \\ \rho_{AB}^2 & \rho_{BB} & 1 \end{pmatrix},$$

and

$$\text{Corr}(6)_{ABB} = \begin{pmatrix} 1 & \rho_{AB} & \rho_{AB}^2 \\ \rho_{AB} & 1 & \rho_{BB} \\ \rho_{AB}^2 & \rho_{BB} & 1 \end{pmatrix},$$

$$\text{Corr}(6)_{BAA} = \begin{pmatrix} 1 & \rho_{BA} & \rho_{BA}^2 \\ \rho_{BA} & 1 & \rho_{AA} \\ \rho_{BA}^2 & \rho_{AA} & 1 \end{pmatrix}.$$

Same as the above two cases, for $p = 4$ case, we consider an example with treatment sequences *AABB*, *BBAA*. The fourth type of correlation matrix will be as given below. The fifth type of correlation matrices for both treatment sequences *AABB* and *BBAA* will be same because in treatment sequences A, B are interchangeable and $\rho_{AA} = \rho_{BB}$ and $\rho_{AB} = \rho_{BA}$. Sixth type of correlation matrices for both treatment sequences *ABB* and *BAA* will be different as ρ_{AB} and ρ_{BA} are different. We get

$$\text{Corr}(4)_{AABB} = \begin{pmatrix} 1 & \rho_{AA} & 0 & 0 \\ \rho_{AA} & 1 & \rho_{AB} & 0 \\ 0 & \rho_{AB} & 1 & \rho_{BB} \\ 0 & 0 & \rho_{BB} & 1 \end{pmatrix},$$

$$\text{Corr}(4)_{BBAA} = \begin{pmatrix} 1 & \rho_{BB} & 0 & 0 \\ \rho_{BB} & 1 & \rho_{BA} & 0 \\ 0 & \rho_{BA} & 1 & \rho_{AA} \\ 0 & 0 & \rho_{AA} & 1 \end{pmatrix},$$

and

$$\text{Corr}(5)_{AABB} = \text{Corr}(5)_{BBAA} = \begin{pmatrix} 1 & \rho_{BB} & \rho_{BA}^2 & \rho_{BA}^3 \\ \rho_{BB} & 1 & \rho_{BA} & \rho_{BB}^2 \\ \rho_{BA}^2 & \rho_{BA} & 1 & \rho_{BB} \\ \rho_{BA}^3 & \rho_{BA}^2 & \rho_{BB} & 1 \end{pmatrix},$$

and

$$\text{Corr}(6)_{AABB} = \begin{pmatrix} 1 & \rho_{AA} & \rho_{AB}^2 & \rho_{AB}^3 \\ \rho_{AA} & 1 & \rho_{AB} & \rho_{AB}^2 \\ \rho_{AB}^2 & \rho_{AB} & 1 & \rho_{BB} \\ \rho_{AB}^3 & \rho_{AB}^2 & \rho_{BB} & 1 \end{pmatrix},$$

$$\text{Corr}(6)_{BBAA} = \begin{pmatrix} 1 & \rho_{BB} & \rho_{BA}^2 & \rho_{BA}^3 \\ \rho_{BB} & 1 & \rho_{BA} & \rho_{BA}^2 \\ \rho_{BA}^2 & \rho_{BA} & 1 & \rho_{AA} \\ \rho_{BA}^3 & \rho_{BA}^2 & \rho_{AA} & 1 \end{pmatrix}.$$

For $p = 4$ case, we discuss another interesting example with four treatments A, B, C and D. The set of treatment sequences for this example is $\Omega = \{ABCD, BDAC, CADB, DCBA\}$. This experiment will be discussed in detail later in Sect. 4. Note that the treatment sequences are given by a Latin square design and the treatments are interchangeable.

$$\begin{matrix} A & B & C & D \\ B & D & A & C \\ C & A & D & B \\ D & C & B & A \end{matrix}.$$

For this example, above six different types of correlation matrices are considered. The first three correlation matrices will be the same as above with $\rho = 0.3$, $\rho = 0.2$ and $\rho = 0.1$, respectively. The fourth type correlation structure will be defined in similar manner as discussed above. The fifth type correlation matrix is defined using AR(1) correlation structure with $\rho_{AB} = \rho_{AC} = \rho_{AD} = \rho_{BA} = \rho_{CA} = \rho_{DA} = 0.4$, $\rho_{BC} = \rho_{BD} = \rho_{CB} = \rho_{DB} = 0.3$ and $\rho_{CD} = \rho_{DC} = 0.2$. For fourth type and sixth type of correlation matrix, $\rho_{AB} = \rho_{AC} = \rho_{AD}$ is taken to be 0.4. In a similar manner, $\rho_{BA} = \rho_{BC} = \rho_{BD}$ is taken to be 0.3 and $\rho_{CA} = \rho_{CB} = \rho_{CD}$ is taken to be 0.2 and $\rho_{DA} = \rho_{DB} = \rho_{DC}$ is taken to be 0.1. As the entries of the correlation matrix depend on which treatment the subject receives in a particular period, these correlation matrices are different for different treatment sequences and are listed as follows:

$$\text{Corr}(4)_{ABCD} = \begin{pmatrix} 1 & \rho_{AB} & 0 & 0 \\ \rho_{AB} & 1 & \rho_{BC} & 0 \\ 0 & \rho_{BC} & 1 & \rho_{CD} \\ 0 & 0 & \rho_{CD} & 1 \end{pmatrix},$$

$$\text{Corr}_{BDAC} = \begin{pmatrix} 1 & \rho_{BD} & 0 & 0 \\ \rho_{BD} & 1 & \rho_{DA} & 0 \\ 0 & \rho_{DA} & 1 & \rho_{AC} \\ 0 & 0 & \rho_{AC} & 1 \end{pmatrix},$$

$$\text{Corr}(4)_{CADB} = \begin{pmatrix} 1 & \rho_{CA} & 0 & 0 \\ \rho_{CA} & 1 & \rho_{AD} & 0 \\ 0 & \rho_{AD} & 1 & \rho_{DB} \\ 0 & 0 & \rho_{DB} & 1 \end{pmatrix},$$

$$\text{Corr}(4)_{DCBA} = \begin{pmatrix} 1 & \rho_{DC} & 0 & 0 \\ \rho_{DC} & 1 & \rho_{CB} & 0 \\ 0 & \rho_{CB} & 1 & \rho_{BA} \\ 0 & 0 & \rho_{BA} & 1 \end{pmatrix},$$

and

$$\text{Corr}(5/6)_{ABCD} = \begin{pmatrix} 1 & \rho_{AB} & \rho_{AC}^2 & \rho_{AD}^3 \\ \rho_{AB} & 1 & \rho_{BC} & \rho_{BD}^2 \\ \rho_{AC}^2 & \rho_{BC} & 1 & \rho_{CD} \\ \rho_{AD}^3 & \rho_{BD}^2 & \rho_{CD} & 1 \end{pmatrix},$$

$$\text{Corr}(5/6)_{BDAC} = \begin{pmatrix} 1 & \rho_{BD} & \rho_{BA}^2 & \rho_{BC}^3 \\ \rho_{BD} & 1 & \rho_{DA} & \rho_{DC}^2 \\ \rho_{BA}^2 & \rho_{DA} & 1 & \rho_{AC} \\ \rho_{BC}^3 & \rho_{DC}^2 & \rho_{AC} & 1 \end{pmatrix},$$

$$\text{Corr}(5/6)_{CADB} = \begin{pmatrix} 1 & \rho_{CA} & \rho_{CD}^2 & \rho_{CB}^3 \\ \rho_{CA} & 1 & \rho_{AD} & \rho_{AB}^2 \\ \rho_{CD}^2 & \rho_{AD} & 1 & \rho_{DB} \\ \rho_{CB}^3 & \rho_{AB}^2 & \rho_{DB} & 1 \end{pmatrix},$$

$$\text{Corr}(5/6)_{DCBA} = \begin{pmatrix} 1 & \rho_{DC} & \rho_{DB}^2 & \rho_{DA}^3 \\ \rho_{DC} & 1 & \rho_{CB} & \rho_{CA}^2 \\ \rho_{DB}^2 & \rho_{CB} & 1 & \rho_{BA} \\ \rho_{DA}^3 & \rho_{CA}^2 & \rho_{BA} & 1 \end{pmatrix}.$$

In the above, we only specified the forms of correlation structures. Note that for this particular example, the form of Corr(5) is the same as that of Corr(6) since the treatment sequences are obtained using a Latin square design. In Sect. 4, we will consider the above six types of correlation structures and calculate the corresponding optimal proportions. We will also perform a simulation analysis using this example. For simulation analysis, AR(1) correlation structure will be considered with different ρ values. We have performed robustness in "Appendix" and provided explicit expressions on how to obtain objective function in Supplementary Section S1.2.

2.4 Algorithm for Locally Optimal Crossover Trials

In this section, we propose an algorithm to find locally optimal designs for crossover trials. Assumed values of the model parameters are obtained from some prior knowledge or pilot studies. To identify the locally optimal crossover design, the major challenge is in minimizing the objective function. The complexity of the objective function increases with the increase in t, p and k. We use the `solnp` function in R for numerical optimization.

Algorithm : Pseudo-code for finding locally optimal crossover designs.

Given assumed values of the parameters, construct the design matrix, correlation matrix, and the parameter vector.

for

Each subject in each period
Calculate the mean of the response

end
for

Each treatment sequence
Calculate the covariance matrix using the correlation matrix
Diagonal entries of covariance matrix are variances of observations
Variance depends on the distribution of the response
Calculate the inverse of covariance matrix

end
for

Each treatment sequence
Calculate the corresponding derivative matrix
Using calculated matrices and variables corresponding to each treatment sequence, compute the variance matrix of parameter estimates
Calculate variance matrix of treatment effects. Its determinant is the required objective function

end
function

Define the objective function along with the constraints, i.e., sum of proportions is equal to one

end
solnp Using this constraint optimization function calculate optimal proportions

3 Optimal Designs for Two-Treatment Crossover Trials

The crossover designs for which we will calculate the optimal proportions are similar to those discussed by Laska and Meisner [15] and Carriere and Huang [3]. Optimal proportions are listed below for $p = 2, 3, 4$ for binary response and for $p = 2$ for poisson response under two sets of parameter estimates. In this section, we consider only two treatments A and B. Considering our baseline constraint to be $\tau_A = \rho_A = 0$ and $\beta_1 = 0$, we only have $p + 2$ parameters in vector θ. So, when

Table 1 Color scheme for different correlation structures

Correlation Structure	Color
Corr(1) $(1-\rho)I_p + \rho J_p$ with $\rho = 0.1$	
Corr(2) $\rho^{\lvert i-i' \rvert}, i \neq i'$ with $\rho = 0.1$	
Corr(3) with $\rho = 0.1$	
Corr(4) with $\rho_{AB} = 0.2, \rho_{BA} = 0.5$	
Corr(5) with $\rho_{AB} = \rho_{BA} = 0.4$	
Corr(6) with $\rho_{AB} = 0.4, \rho_{BA} = 0.3$	

Table 2 Optimal proportions for $p = 2$ case

Design points	Corr	Optimal proportions under θ_1	Optimal proportions under θ_2
$\{AB, BA\}$	Corr(1)	$\{0.1770, 0.8230\}$	$\{0.5070, 0.4930\}$
	Corr(2)	$\{0.1770, 0.8230\}$	$\{0.5070, 0.4930\}$
	Corr(3)	$\{0.1770, 0.8230\}$	$\{0.5070, 0.4930\}$
	Corr(4)	$\{0.1770, 0.8230\}$	$\{0.5070, 0.4930\}$
	Corr(5)	$\{0.1770, 0.8230\}$	$\{0.5070, 0.4930\}$
	Corr(6)	$\{0.1770, 0.8230\}$	$\{0.5070, 0.4930\}$
$\{AB, BA, AA, BB\}$	Corr(1)	$\{0.0908, 0.5207, 0.0315, 0.3570\}$	$\{0.2633, 0.2425, 0.2722, 0.2220\}$
	Corr(2)	$\{0.0908, 0.5207, 0.0315, 0.3570\}$	$\{0.2633, 0.2425, 0.2722, 0.2220\}$
	Corr(3)	$\{0.0908, 0.5207, 0.0315, 0.3570\}$	$\{0.2633, 0.2425, 0.2722, 0.2220\}$
	Corr(4)	$\{0.0957, 0.4960, 0.0338, 0.3745\}$	$\{0.2534, 0.2393, 0.2661, 0.2412\}$
	Corr(5)	$\{0.1002, 0.4941, 0.0379, 0.3678\}$	$\{0.2496, 0.2359, 0.2801, 0.2344\}$
	Corr(6)	$\{0.0972, 0.5050, 0.0367, 0.3611\}$	$\{0.2502, 0.2400, 0.2808, 0.2290\}$

there are only two treatments involved in the crossover trial, the parameter vector θ is $[\lambda, \beta_2, \ldots, \beta_p, \tau_2, \rho_2]$.

Optimal proportions for different crossover designs are calculated with each of the six different correlation structures mention above. For each correlation matrix that we consider, an optimal design ζ^* is the one minimizing the determinant of $\text{Var}(\hat{\tau})$ in Eq. (5) with respect to p_ω such that $\sum_{w \in \Omega} p_w = 1$.

We use different colors to represent different correlation structures. The color scheme that we use is given in Table 1.

3.1 Optimal Designs for Binary Response

In case of binary response, we calculate locally optimal designs under model (2) for different crossover designs.

We first consider the local optimality approach, for $p = 2$ case. For illustration purpose, we assume that the parameter values are $\theta_1 = [\lambda, \beta_2, \tau_B, \rho_B] = [0.5, -1.0, 4.0, -2.0]$ which gives us non-uniform optimal

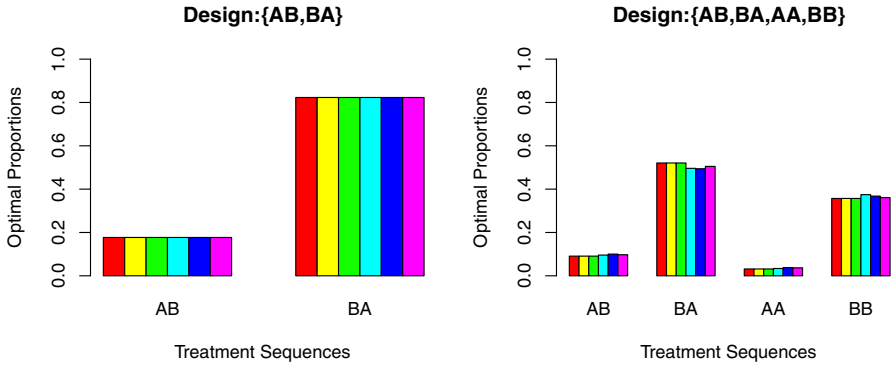

Fig. 1 Optimal proportions for $p = 2$ case under θ_1

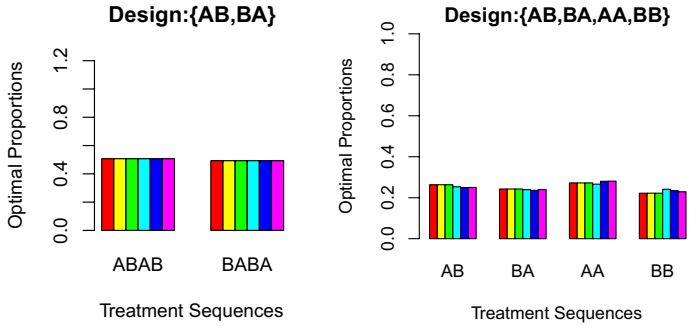

Fig. 2 Optimal proportions for $p = 2$ case under θ_2

allocations and $\theta_2 = [\lambda, \beta_2, \tau_B, \rho_B] = [0.5, 0.06, -0.35, 0.73]$ which gives us approximately uniform allocations. Note that we need to know the parameter values before calculating the optimal proportions. If the initial guess for the model parameters changes, the obtained optimal proportions will change as well. For different correlation structures, the optimal designs (proportions) are stated in Table 2. The same information is presented in Figs. 1 and 2 as well.

It can be seen from the graphs in Figs. 1 and 2 that in case of $p = 2$ the optimal proportions do not vary when correlation structure changes under both θ_2 and θ_1. Uniform designs (same proportions for each sequence) are often used in practice. It is clear that those uniform designs are sub-optimal under θ_1.

For $p = 3$ case, as before suppose our guess for the parameter values are $\theta_1 = [\lambda, \beta_2, \beta_3, \tau_B, \rho_B] = [0.5, -1.0, 2.0, 4.0, -2.0]$ which gives us non-uniform optimal allocations and $\theta_2 = [\lambda, \beta_2, \beta_3, \tau_B, \rho_B] = [0.5, 0.06, -0.53, -0.35, 0.73]$ which gives us approximately uniform optimal allocations. The designs are presented in Table 3 and Fig. 3 for the first example, and in Table 4 and Fig. 4 for the second example. It can be seen that in case of $p = 3$ also the optimal proportions do not vary much when correlation structure changes under both θ_1 and θ_2. Similar to $p = 2$

Table 3 Optimal proportions for $p = 3$ case for designs with two treatment sequences

Design points	Corr	Optimal proportions under θ_1	Optimal proportions under θ_2
$\{ABB, BAA\}$	Corr(1)	$\{0.5756, 0.4244\}$	$\{0.4880, 0.5120\}$
	Corr(2)	$\{0.5761, 0.4239\}$	$\{0.4887, 0.5113\}$
	Corr(3)	$\{0.5762, 0.4238\}$	$\{0.4888, 0.5112\}$
	Corr(4)	$\{0.6120, 0.3880\}$	$\{0.5416, 0.4584\}$
	Corr(5)	$\{0.5921, 0.4079\}$	$\{0.4917, 0.5083\}$
	Corr(6)	$\{0.5721, 0.4279\}$	$\{0.4700, 0.5300\}$
$\{ABA, BAB\}$	Corr(1)	$\{0.1768, 0.8232\}$	$\{0.5070, 0.4930\}$
	Corr(2)	$\{0.1766, 0.8234\}$	$\{0.5072, 0.4928\}$
	Corr(3)	$\{0.1766, 0.8234\}$	$\{0.5072, 0.4928\}$
	Corr(4)	$\{0.1756, 0.8244\}$	$\{0.5217, 0.4783\}$
	Corr(5)	$\{0.1714, 0.8286\}$	$\{0.5088, 0.4912\}$
	Corr(6)	$\{0.1715, 0.8285\}$	$\{0.5043, 0.4957\}$
$\{AAB, BBA\}$	Corr(1)	$\{0.2713, 0.7287\}$	$\{0.4927, 0.5073\}$
	Corr(2)	$\{0.2738, 0.7262\}$	$\{0.4926, 0.5074\}$
	Corr(3)	$\{0.2740, 0.7260\}$	$\{0.4926, 0.5074\}$
	Corr(4)	$\{0.2685, 0.7315\}$	$\{0.5181, 0.4819\}$
	Corr(5)	$\{0.2771, 0.7229\}$	$\{0.4911, 0.5089\}$
	Corr(6)	$\{0.2740, 0.7260\}$	$\{0.4702, 0.5298\}$

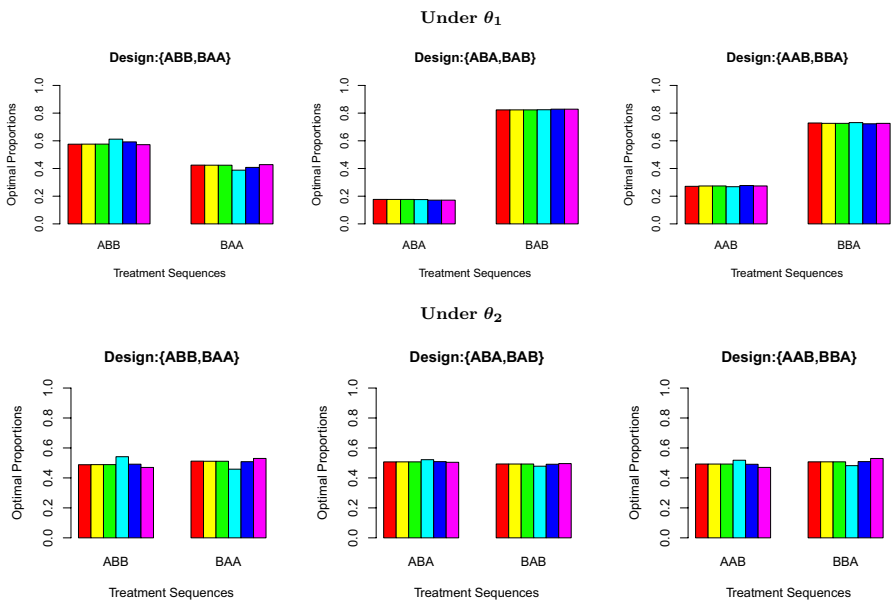

Fig. 3 Optimal proportions for $p = 3$ case with two treatment sequences under θ_1 and θ_2, respectively

Table 4 Optimal proportions for $p = 3$ case for designs with four treatment sequences

Design points	Corr	Optimal proportions under θ_1	Optimal proportions under θ_2
$\{ABB, BAA, AAA, BBB\}$	Corr(1)	$\{0.1222, 0.5344, 0.0000, 0.3434\}$	$\{0.4880, 0.5120, 0.0000, 0.0000\}$
	Corr(2)	$\{0.1199, 0.5316, 0.0022, 0.3463\}$	$\{0.4887, 0.5113, 0.0000, 0.0000\}$
	Corr(3)	$\{0.1197, 0.5312, 0.0025, 0.3466\}$	$\{0.4888, 0.5112, 0.0000, 0.0000\}$
	Corr(4)	$\{0.1115, 0.4975, 0.0100, 0.3720\}$	$\{0.5398, 0.4556, 0.0046, 0.0000\}$
	Corr(5)	$\{0.1313, 0.5113, 0.0000, 0.3574\}$	$\{0.4917, 0.5083, 0.0000, 0.0000\}$
	Corr(6)	$\{0.1233, 0.5236, 0.0018, 0.3513\}$	$\{0.4700, 0.5300, 0.0000, 0.0000\}$
$\{ABB, AAB, BAA, BBA\}$	Corr(1)	$\{0.0413, 0.1130, 0.4384, 0.4073\}$	$\{0.3544, 0.1646, 0.3908, 0.0902\}$
	Corr(2)	$\{0.0316, 0.1196, 0.4373, 0.4115\}$	$\{0.4266, 0.0957, 0.4777, 0.0000\}$
	Corr(3)	$\{0.0304, 0.1204, 0.4371, 0.4121\}$	$\{0.4271, 0.0953, 0.4776, 0.0000\}$
	Corr(4)	$\{0.0005, 0.1440, 0.4471, 0.4084\}$	$\{0.1512, 0.3503, 0.1854, 0.3131\}$
	Corr(5)	$\{0.0811, 0.1033, 0.4297, 0.3858\}$	$\{0.4420, 0.0747, 0.4833, 0.0000\}$
	Corr(6)	$\{0.0749, 0.1070, 0.4270, 0.3911\}$	$\{0.4094, 0.0955, 0.4951, 0.0000\}$
$\{ABB, ABA, BAA, BAB\}$	Corr(1)	$\{0.5755, 0.0000, 0.4244, 0.0000\}$	$\{0.4606, 0.0194, 0.4710, 0.0490\}$
	Corr(2)	$\{0.5761, 0.0000, 0.4239, 0.0000\}$	$\{0.4430, 0.0391, 0.4526, 0.0653\}$
	Corr(3)	$\{0.5762, 0.0000, 0.4238, 0.0000\}$	$\{0.4408, 0.0415, 0.4504, 0.0673\}$
	Corr(4)	$\{0.6120, 0.0000, 0.3880, 0.0000\}$	$\{0.4634, 0.1036, 0.4152, 0.0178\}$
	Corr(5)	$\{0.5921, 0.0000, 0.4079, 0.0000\}$	$\{0.4582, 0.0280, 0.4642, 0.0496\}$
	Corr(6)	$\{0.5721, 0.0000, 0.4279, 0.0000\}$	$\{0.4420, 0.0142, 0.4787, 0.0651\}$

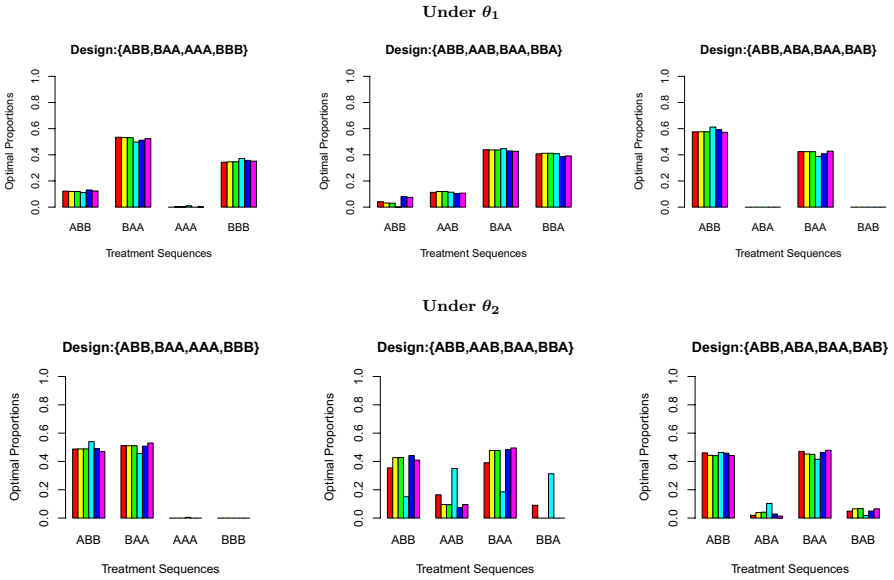

Fig. 4 Optimal proportions for $p = 3$ case with four treatment sequences under θ_1 and θ_2, respectively

case, it is clear from Tables 3 and 4 that uniform designs are sub-optimal for $p = 3$ case with two and four treatment sequences under θ_1.

An interesting thing to observe from Fig. 4 is that, unlike the previous examples, here under θ_1 the optimal proportions vary a little for different correlation structures. Also, as before, not only the uniform design is sub-optimal here, the first and third designs have optimal allocations very low for some sequences. Also it can be observed from Fig. 4 that under θ_2 for different correlation structures some of the optimal proportions are zero for all the three designs. Hence, under θ_2, these designs fail to have uniform allocations.

For $p = 4$ case, in a similar way, we calculate locally optimal designs with nominal parameter values as $\theta_1 = [\lambda, \beta_2, \beta_3, \beta_4, \tau_B, \rho_B] = [0.5, -1.0, 2.0, -1.5, 4.0, -2.0]$ which gives us non-uniform allocations and $\theta_2 = [\lambda, \beta_2, \beta_3, \beta_4, \tau_B, \rho_B] = [0.5, 0.06, -0.53, -0.6, -0.35, 0.73]$ which gives us approximately uniform allocations. From Table 5 and Fig. 5, it is clear that similar to $p = 2$ and $p = 3$ cases the uniform designs are sub-optimal for $p = 4$ case under θ_1.

In most cases, we may not have a clear idea about true correlation structure for responses, and hence, we choose an working correlation structure. The results in this section show that no matter what correlation structure we choose or what correlation coefficient (we obtained optimal proportions for different values of correlation coefficient but details are omitted here) we choose, the proposed design gives almost similar optimal proportions in each case, which suggests that optimal designs are robust.

Table 5 Optimal proportions for $p = 4$ case

Design points	Corr	Optimal proportions under θ_1	Optimal proportions under θ_2
$\{AABB, BBAA\}$	Corr(1)	$\{0.2723, 0.7277\}$	$\{0.4953, 0.5047\}$
	Corr(2)	$\{0.2743, 0.7257\}$	$\{0.4949, 0.5051\}$
	Corr(3)	$\{0.2744, 0.7256\}$	$\{0.4949, 0.5051\}$
	Corr(4)	$\{0.2690, 0.7310\}$	$\{0.5244, 0.4756\}$
	Corr(5)	$\{0.2772, 0.7228\}$	$\{0.4937, 0.5063\}$
	Corr(6)	$\{0.2745, 0.7255\}$	$\{0.4700, 0.5300\}$
$\{ABBA, BAAB\}$	Corr(1)	$\{0.6075, 0.3925\}$	$\{0.4992, 0.5008\}$
	Corr(2)	$\{0.6045, 0.3955\}$	$\{0.4998, 0.5002\}$
	Corr(3)	$\{0.6042, 0.3958\}$	$\{0.4998, 0.5002\}$
	Corr(4)	$\{0.5815, 0.4185\}$	$\{0.4927, 0.5073\}$
	Corr(5)	$\{0.6444, 0.3556\}$	$\{0.5021, 0.4979\}$
	Corr(6)	$\{0.6419, 0.3581\}$	$\{0.5007, 0.4993\}$
$\{ABAB, BABA\}$	Corr(1)	$\{0.1763, 0.8237\}$	$\{0.5071, 0.4929\}$
	Corr(2)	$\{0.1767, 0.8233\}$	$\{0.5071, 0.4929\}$
	Corr(3)	$\{0.1767, 0.8233\}$	$\{0.5071, 0.4929\}$
	Corr(4)	$\{0.1722, 0.8278\}$	$\{0.5086, 0.4914\}$
	Corr(6)	$\{0.1714, 0.8286\}$	$\{0.5031, 0.4969\}$

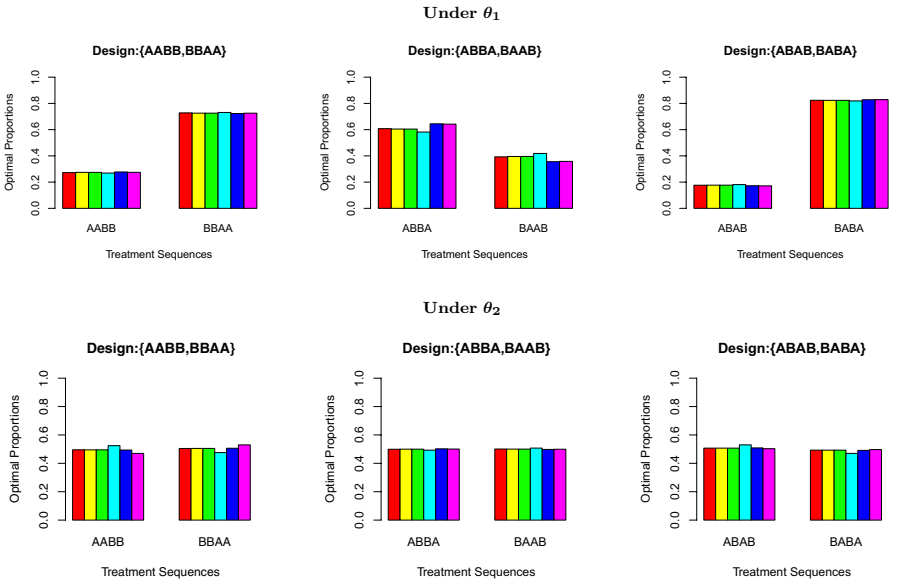

Fig. 5 Optimal proportions for $p = 4$ case under θ_1 and θ_2, respectively

3.2 Optimal Designs for Poisson Response

In the case of Poisson response, we calculate locally optimal design for the following example under the model,

$$\log(\mu_{ij}) = \eta_{ij} = \lambda + \beta_i + \tau_{d(i,j)} + \rho_{d(i-1,j)}, \tag{6}$$

where notations have the same meaning as in Eq. (2).

We consider an example described in Layard and Arvesen [16]. In a crossover clinical trial to test a standard anti-nausea treatment (drug A) against a proposed treatment (drug B), twenty subjects were tested, ten for each order of administration. The response variable is the number of episodes of nausea suffered by a patient during the first two hours after cancer chemotherapy, and for a given patient is approximately Poisson distributed.

We calculate optimal designs using two values of parameter estimates. $\theta_1 = [0.2, 0.34, -1.60, -1.65]$ represents those parameter estimates that give us non-uniform designs, and $\theta_2 = [-0.223, -0.875, 0.405, -0.105]$ corresponds to parameter estimates guessed from the data presented in Table 6.

It can be noted from Table 6 that when responses are Poisson in nature the optimal proportions do not vary much when correlation structure changes under both θ_1 and θ_2. This suggests us that even when responses are Poisson in nature the proposed design gives almost similar optimal proportions for different choices of correlation matrices. Hence, obtained optimal designs are robust.

Table 6 Optimal proportions for anti-nausea experiment

Design points	Correlation structure	Optimal design: θ_1
$\{AB, BA\}$	Corr(1) $(1-\rho)I_p + \rho J_p$ with $\rho = 0.1$	$\{0.3632, 0.6368\}$
	Corr(2) $\rho^{\lvert i-i' \rvert}, i \neq i'$ with $\rho = 0.1$	$\{0.3632, 0.6368\}$
	Corr(3) with $\rho = 0.1$	$\{0.3632, 0.6368\}$
	Corr(4) with $\rho_{AB} = 0.2, \rho_{BA} = 0.5$	$\{0.3632, 0.6368\}$
	Corr(5) with $\rho_{AB} = \rho_{BA} = 0.4$	$\{0.3632, 0.6368\}$
	Corr(6) with $\rho_{AB} = 0.4, \rho_{BA} = 0.3$	$\{0.3632, 0.6368\}$
Design points	Correlation structure	Optimal design: θ_2
$\{AB, BA\}$	Corr(1) $(1-\rho)I_p + \rho J_p$ with $\rho = 0.1$	$\{0.5505, 0.4495\}$
	Corr(2) $\rho^{\lvert i-i' \rvert}, i \neq i'$ with $\rho = 0.1$	$\{0.5505, 0.4495\}$
	Corr(3) with $\rho = 0.1$	$\{0.5505, 0.4495\}$
	Corr(4) with $\rho_{AB} = 0.2, \rho_{BA} = 0.5$	$\{0.5505, 0.4495\}$
	Corr(5) with $\rho_{AB} = \rho_{BA} = 0.4$	$\{0.5505, 0.4495\}$
	Corr(6) with $\rho_{AB} = 0.4, \rho_{BA} = 0.3$	$\{0.5505, 0.4495\}$

4 Optimal Design for Multiple-Treatment Crossover Trials

So far, we have considered crossover designs with two treatments only. In this section, we extend our study for multiple treatments. This is motivated by a four-period four-treatment trial which was first given in Kenward and Jones [8] and later discussed as Example 6.1 in their book [9], *Design and Analysis for Crossover Trials*.

4.1 Latin Square Design and Optimal Proportions

In this example, binary responses for four-period crossover trial were obtained. There were four treatments, and treatment sequences were allocated at random to eighty different subjects at four different periods. At the end of each period, efficacy measurement of each subject was recorded as success or failure, which resulted in joint outcome at the end of the trial. The dataset contains four different treatment sequences which were decided before the trial $\Omega = \{ABCD, BDAC, CADB, DCBA\}$ along with the joint outcome of four different periods from the same subject according to a particular treatment sequence. The numbers below each sequence denote how many subjects received that particular treatment sequence, and the particular response was recorded.

We use the correlation matrices defined in Sect. 2.3 and calculate the optimal proportions. As mentioned earlier for estimating parameters, we have considered the baseline constraints as $\beta_1 = \tau_A = \rho_A = 0$, so that the design matrix has full column rank and all other parameters are estimable.

Table 7 Optimal proportions for different correlation matrices

Correlation structure	θ_1				θ_2			
	ABCD	*BDAC*	*CADB*	*DCBA*	*ABCD*	*BDAC*	*CADB*	*DCBA*
Corr(1)	0.1725	0.2483	0.2223	0.3569	0.2463	0.2493	0.2504	0.2540
Corr(2)	0.1747	0.2490	0.2184	0.3579	0.2461	0.2493	0.2501	0.2546
Corr(3)	0.1714	0.2480	0.2236	0.3570	0.2461	0.2492	0.2507	0.2540
Corr(4)	0.1788	0.2556	0.2163	0.3493	0.2478	0.2634	0.2334	0.2554
Corr(5)	0.1784	0.2465	0.2101	0.3650	0.2480	0.2517	0.2442	0.2561
Corr(6)	0.1752	0.2531	0.2170	0.3547	0.2470	0.2656	0.2320	0.2554

Using these baseline constraints and `glm` function in R, we fit the model, which gives us parameter estimates for the given data. Then, we use these parameter estimates to make a guess for values of unknown parameters. Our nominal guess for the parameter values is $\theta_2 = [0.5, 0.06, -0.53, -0.6, -0.35, 0.025, -0.23, 0.73, 0.23, 0.30]$. Now, we follow the same procedure as mentioned in above pseudocode and calculate the optimal designs for different correlation structures. We also calculate optimal proportions by considering parameter estimates that gives non-uniform designs, i.e., $\theta_1 = [-2, 0.25, 0, 0.75, 1, 5, -1.5, -3.5, 2.75, 0.75]$. As seen from Table 7, for the Latin square design the optimal proportions that we obtain using θ_1 are non-uniform and that using θ_2 are nearly uniform.

We also calculate optimal design considering all 24 sequences. We consider Corr(2) and calculate optimal proportions for different values of ρ. Please refer the Supplementary Materials for details. From the tables in the Supplementary Materials, it can be noted that corresponding to θ_1 we have non-uniform allocations for the Latin Square design, and almost uniform allocation corresponding to θ_2. In case of non-uniform allocations, although nothing is uniform, the optimal design corresponding to θ_1 has more zeros. Also note that the allocations do not vary a lot as ρ changes, particularly for the sequences where we have zero allocations.

4.2 Sensitivity Study and Relative *D*-efficiency

In this section, we study the performance of the proposed locally optimal designs via sensitivity study in terms of relative *D*-efficiencies. Let θ_t be true parameter values and θ_c be assumed parameter values. Then, we have corresponding objective function for these two choices of parameter values i.e $\det(\text{var}(\hat{\tau}_t))$ and $\det(\text{var}(\hat{\tau}_c))$, respectively. Hence, the relative loss of efficiency of choosing θ_c instead of θ_t can be formulated as

$$S(\tau_t, \tau_c) = \frac{\det(\text{var}(\hat{\tau}_t))^{\left(-\frac{1}{k}\right)} - \det(\text{var}(\hat{\tau}_c))^{\left(-\frac{1}{k}\right)}}{\det(\text{var}(\hat{\tau}_t))^{\left(-\frac{1}{k}\right)}},$$

Table 8 Assumed values for model parameters

Parameters θ_c	Case 1	Case 2
λ	$U(-0.5, 1.5)$	$U(-1.5, 2.5)$
β_2	$U(-0.04, 0.16)$	$U(-0.14, 0.26)$
β_3	$U(-1.53, 0.47)$	$U(-2.53, 1.47)$
β_4	$U(-1.6, 0.4)$	$U(-2.6, 1.4)$
τ_2	$U(-1.35, 0.65)$	$U(-2.35, 1.65)$
τ_3	$U(-0.075, 0.125)$	$U(-0.175, 0.225)$
τ_4	$U(-1.23, 0.77)$	$U(-2.23, 1.77)$
ρ_2	$U(-0.27, 1.73)$	$U(-1.27, 2.73)$
ρ_3	$U(-0.77, 1.23)$	$U(-1.77, 2.23)$
ρ_4	$U(-0.70, 1.30)$	$U(-1.70, 2.30)$

where k is the dimension of τ. Then, the relative D-efficiency of the original design ξ compared to the optimal design ξ^* can be computed using the formula:

$$E_\xi = \left[\frac{\det(\mathrm{var}(\hat{\tau}_c))_{\xi^*}}{\det(\mathrm{var}(\hat{\tau}_t))_\xi}\right]^{-\frac{1}{k}}.$$

For the Latin square design example, we consider two cases of assumed values of θ_c for model parameters as mentioned in Table 8. For each case, the values of parameters are simulated from a uniform distribution. The range of uniform distribution is obtained by ± 1 and ± 2 from true parameter values θ_t for each case, respectively. Here, we consider $\theta_t = [0.5, 0.06, -0.53, -0.6, -0.35, 0.025, -0.23, 0.73, 0.23, 0.30]$ (Fig. 6).

4.3 Simulation Studies with Two-Stage Designs

As stated earlier, the main aim of this paper is to determine optimal and efficient crossover designs for experiments where the generalized linear model adequately describes the process under study. Crossover trials are repeated measurement designs, where these repeated measurements on the same subject have great advantages, but there are also many potential disadvantages associated with it. Nevertheless, the impact of these disadvantages can be minimized or reduced if we choose a proper design and analysis method.

One of the major disadvantages of repeated measurement designs is that the effect of the treatment depends on the subject itself. Stronger subject effects cause more variation on estimated treatment effects.

The simulation studies are motivated by the real-life example of Latin square design mentioned above. Since all the correlation structures mentioned in Sect. 2.3 perform similarly in Table 7, we choose Corr(2) for illustration purpose. Note that in Corr(2), we have AR(1) structure, where the correlation between two responses decreases as the number of periods between responses increases, which makes

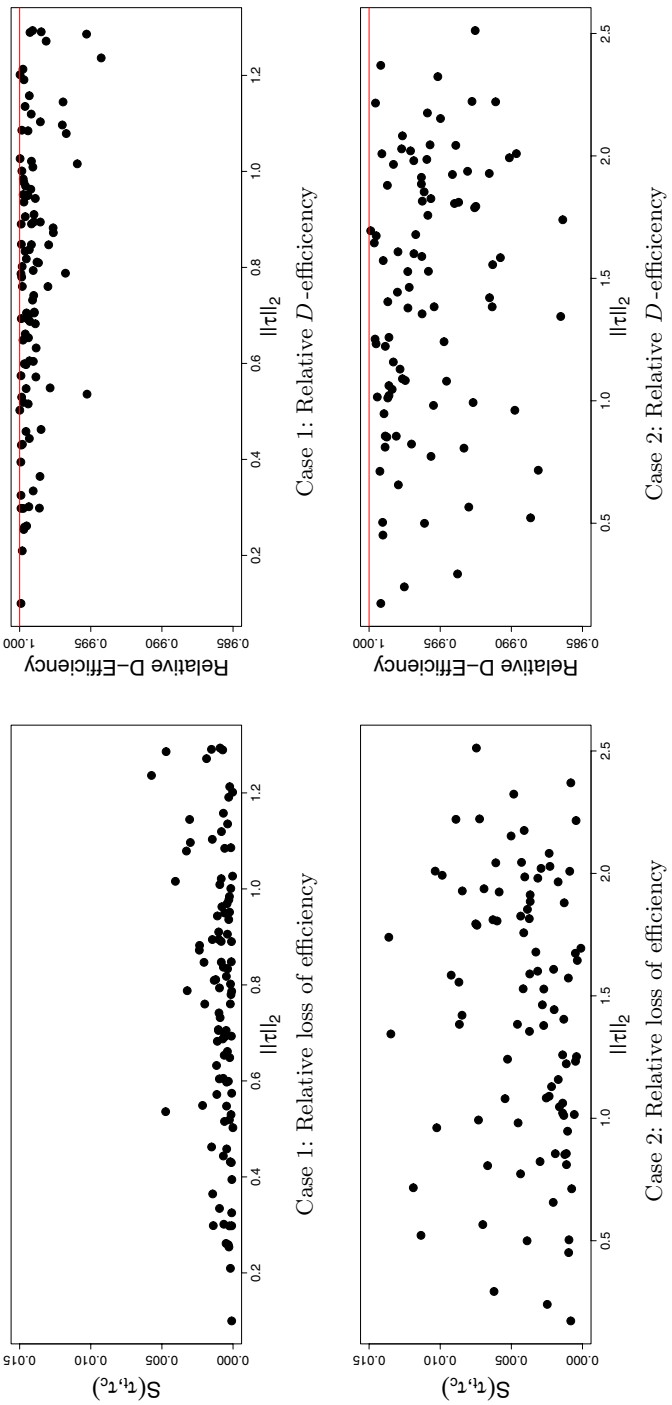

Fig. 6 Performance of the locally optimal designs

good practical sense. For these simulation studies, we are considering 400 observations and two different types of initial guess for θ values. In Case 1, we will use $\theta_2 = [0.5, 0.06, -0.53, -0.6, -0.35, 0.025, -0.23, 0.73, 0.23, 0.30]$ which is obtained from real data. This choice of θ_2 gives optimal allocations as (0.2460, 0. 2495, 0.2500, 0.2545) , which is approximately uniform. For Case 2, we will use $\theta_1 = [-2, 0.25, 0, 0.75, 1, 5, -1.5, -3.5, 2.75, 0.75]$ and this guess of θ_1 is such that optimal allocations are non-uniform. For example, for $\rho = 0.1$ the optimal allocations are (0.172, 0.248, 0.222, 0.358) . Optimal allocations are similar for other values of ρ.

The simulation process used here has two stages. First for a given parameter θ, we define a design matrix corresponding to each treatment sequence along with correlation matrix.

- First stage:

    1. In this stage, we use `rbin` function in R to simulate 30% of observations uniformly over all four treatment sequence. These observations serve as our pilot study. Note that we use uniform design for pilot study.
    2. From these observations obtained in above step, we estimate the correlation coefficient and regression parameters, which are used as the assumed parameter values for the second stage.

- Second stage:

    1. Based on the assumed parameter values obtained in the first stage and the algorithm described in Sect. 2.4, we calculate the optimal allocation for the remaining 70% of the subjects.
    2. Using these optimal allocations, we simulate observations for remaining 70% of subjects according to the assumed parameter values.
    3. In case of uniform design, we simulate total number of observations uniformly over all treatment sequence, i.e., one-fourth of the total observations correspond to each of the four treatment sequence.

During this process, we calculate the parameter estimates based on the simulated observations and calculate the corresponding mean square error (MSE) from the true parameter values for each simulation. Above simulation procedure is repeated 100 times. Finally, we take the average of those individual MSEs to calculate the overall MSE reported in Table 9. We repeat the above simulation process for different correlation coefficients and for two different sets of initial θ's, θ_1 and θ_2. It is clear from Table 9 and Fig. 7 that if the optimal allocations are non-uniform, then the proposed optimal design has a significant advantage over the traditional uniform

Table 9 Simulation results

Corr Corr(2) ρ	Mean Squared Errors Case 1		Case 2	
	Uniform Design	Optimal Design	Uniform Design	Optimal Design
0.1	0.109	0.108	2.834	0.393
0.2	0.103	0.100	2.718	0.659
0.3	0.101	0.140	4.925	0.490
0.4	0.094	0.127	4.896	0.484
0.5	0.100	0.123	2.596	0.428
0.6	0.088	0.109	2.632	0.469
0.7	0.086	0.095	5.110	0.458
0.8	0.066	0.077	2.705	0.586
0.9	0.050	0.051	2.761	0.559

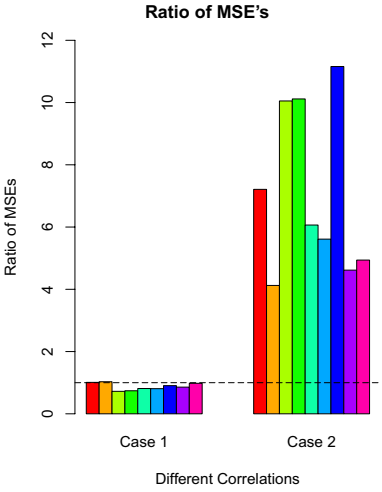

Fig. 7 Simulation results: ratios of MSEs of the uniform versus optimal designs, for different values of ρ, for each of the two cases

designs, for all values of the correlation coefficient. It should be noted that those high values of MSEs for uniform designs are mostly due to a handful of "bad" datasets. In our experience, the proposed optimal designs never give rise to such data.

5 Discussion

In practice, it is customary to use uniform designs where the same number of subjects is assigned to each treatment sequence. In the case of linear models, such uniform designs are optimal. However, optimal proportions obtained under generalized linear models are not uniform. We identified locally optimal designs under different correlation structures. Tables 2, 3, 4 and 5 and graphs in Figs. 1, 2, 3, 4 and 5 suggest that the optimal proportions do not vary much from one correlation structure to another. These results suggest that the identified designs are robust. The relative loss of efficiency increases as we move away from true parameter values. However, Fig. 6 suggest that this loss of efficiency does not go beyond 2% even for Case 2. Simulation studies and results in Table 9 and Fig. 7 suggest that these designs are more efficient than uniform designs as well.

The general equivalence theorem is frequently used to verify whether a design obtained numerically is indeed optimal or not. It considers the optimality criterion in terms of the Fisher information matrix. In the case of generalized estimating equations (GEE), instead of using the information matrix, we have an objective function in terms of the variance of the parameters of interest, as shown in Eq. (5). Naturally, the equivalence theorem cannot be directly used in this case. However, it is possible to use the expression in Eq. (5) to derive a necessary and sufficient condition that can be used to check whether the derived design is indeed optimal or not. Even a speedy algorithm for identifying optimal designs might be developed based on these results. This is a topic of the future research.

Acknowledgements The authors would like to thank Dr. Pritam Ranjan for many helpful suggestions.

Appendix

Effect of misspecification of working correlation structures

Table 10 lists the locally optimal design when true correlation structure varies from working correlation structure. In the table, first column represents true correlation structure and the corresponding optimal designs are calculated under θ_1 and θ_2 for each misspecified working correlation structure in second column. Also, relative D-efficiency is calculated under θ_1 and θ_2 for each design. It is clear from the relative D-efficiency values in the table that the effect of variance misspecification on the locally optimal design is minimal (see Table 10).

Table 10 Optimal design after variance misspecification

True correlation	Working correlation	Optimal proportions for θ_1				Optimal proportions for θ_2				Relative *D*-efficiency	
Structure	Structure	*ABCD*	*BDAC*	*CADB*	*DCBA*	*ABCD*	*BDAC*	*CADB*	*DCBA*	Under θ_1	Under θ_2
Corr(1)	Corr(2)	0.1723	0.2483	0.2222	0.3572	0.2463	0.2493	0.2504	0.2540	0.9999	0.9999
	Corr(3)	0.1726	0.2483	0.2223	0.3568	0.2463	0.2493	0.2504	0.2540	0.9999	0.9999
	Corr(4)	0.1723	0.2513	0.2202	0.3562	0.2500	0.2500	0.2500	0.2500	0.9997	0.9988
	Corr(5)	0.2447	0.1713	0.2495	0.2223	0.3569	0.2475	0.2557	0.2521	0.9994	0.9995
	Corr(6)	0.2500	0.1724	0.2508	0.2197	0.3571	0.2500	0.2500	0.2500	0.9999	0.9984
Corr(2)	Corr(1)	0.1745	0.2489	0.2183	0.3583	0.2462	0.2493	0.2500	0.2545	0.9999	0.9999
	Corr(3)	0.1744	0.2489	0.2182	0.3585	0.2462	0.2493	0.2500	0.2545	0.9999	0.9999
	Corr(4)	0.1745	0.2514	0.2177	0.3564	0.2500	0.2500	0.2500	0.2500	0.9998	0.9987
	Corr(5)	0.1740	0.2503	0.2180	0.3577	0.2450	0.2480	0.2530	0.2540	0.9997	0.9997
	Corr(6)	0.1744	0.2512	0.2174	0.3570	0.2463	0.2497	0.2505	0.2535	0.9999	0.9985
Corr(3)	Corr(1)	0.1714	0.2480	0.2236	0.3570	0.2461	0.2492	0.2507	0.2540	0.9999	0.9999
	Corr(2)	0.1711	0.2480	0.2235	0.3574	0.2462	0.2492	0.2506	0.2540	0.9999	0.9999
	Corr(4)	0.1713	0.2516	0.2209	0.3562	0.2500	0.2500	0.2500	0.2500	0.9996	0.9987
	Corr(5)	0.1700	0.2463	0.2235	0.3572	0.2441	0.2476	0.2561	0.2522	0.9992	0.9995
	Corr(6)	0.1713	0.2510	0.2204	0.3573	0.2500	0.2500	0.2500	0.2500	0.9999	0.9984
Corr(4)	Corr(1)	0.1783	0.2585	0.2140	0.3492	0.2500	0.2637	0.2347	0.2516	0.9994	0.9987
	Corr(2)	0.1784	0.2580	0.2156	0.3480	0.2486	0.2640	0.2344	0.2530	0.9996	0.9987
	Corr(3)	0.1782	0.2592	0.2131	0.3495	0.2498	0.2643	0.2342	0.2517	0.9992	0.9986
	Corr(5)	0.1778	0.2579	0.2167	0.3476	0.2470	0.2650	0.2343	0.2537	0.9992	0.9993
	Corr(6)	0.1790	0.2555	0.2165	0.3490	0.2485	0.2631	0.2337	0.2547	0.9999	0.9999

Table 10 (continued)

True correlation	Working correlation	Optimal proportions for θ_1				Optimal proportions for θ_2				Relative *D*-efficiency	
Structure	Structure	*ABCD*	*BDAC*	*CADB*	*DCBA*	*ABCD*	*BDAC*	*CADB*	*DCBA*	Under θ_1	Under θ_2
Corr(5)	Corr(1)	0.1774	0.2477	0.2092	0.3657	0.2466	0.2501	0.2486	0.2547	0.9994	0.9999
	Corr(2)	0.1776	0.2476	0.2099	0.3649	0.2470	0.2506	0.2470	0.2554	0.9997	0.9999
	Corr(3)	0.1770	0.2477	0.2087	0.3666	0.2462	0.2503	0.2485	0.2550	0.9992	0.9999
	Corr(4)	0.1776	0.2492	0.2108	0.3624	0.2472	0.2538	0.2450	0.2540	0.9996	0.9994
	Corr(6)	0.1774	0.2496	0.2110	0.3620	0.2465	0.2535	0.2456	0.2544	0.9998	0.9991
Corr(6)	Corr(1)	0.1748	0.2553	0.2142	0.3557	0.2482	0.2652	0.2332	0.2534	0.9997	0.9985
	Corr(2)	0.1748	0.2551	0.2160	0.3541	0.2470	0.2657	0.2329	0.2544	0.9999	0.9985
	Corr(3)	0.1748	0.2558	0.2133	0.3561	0.2482	0.2660	0.2325	0.2533	0.9996	0.9984
	Corr(4)	0.1754	0.2530	0.2172	0.3544	0.2476	0.2652	0.2324	0.2548	0.9999	0.9999
	Corr(5)	0.1741	0.2556	0.2180	0.3523	0.2452	0.2669	0.2339	0.2540	0.9994	0.9991

References

1. Bose M, Dey A (2009) Optimal crossover designs. World Scientific, Singapore
2. Bose M, Dey A (2015) Crossover designs. In: Dean AM, Morris M, Stufken J, Bingham D (eds) Handbook of design and analysis of experiments. CRC Press, Boca Raton
3. Carriere KC, Huang R (2000) Crossover designs for two-treatment clinical trials. J Stat Plan Inference 87:125–134
4. Cheng CS, Wu CF (1980) Balanced repeated measurements designs. Ann Stat 8:1272–1283
5. Chernoff H (1953) Locally optimal designs for estimating parameters. Ann Math Stat 24:586–602
6. Dey A, Gupta VK, Singh M (1983) Optimal change-over designs. Sankhya B45:233–239
7. Hedayat A, Afsarinejad K (1975) Repeated measurements designs, I. In: Srivastava JN (ed) A survey of statistical designs and linear models. North-Holland, Amsterdam, pp 229–242
8. Kenward MG, Jones B (1992) Alternative approaches to the analysis of binary and categorical repeated measurements. J Biopharm Stat 2:137–170
9. Kenward MG, Jones B (2014) Design and analysis of cross-over trials, 3rd edn. Chapman and Hall, London
10. Kershner RP, Federer WT (1981) Two-treatment crossover designs for estimating a variety of effects. J Am Stat Assoc D 76:612–619
11. Khuri AI, Mukherjee B, Sinha BK, Ghosh M (2006) Design issues for generalized linear models: a review. Stat Sci 21:376–399
12. Kunert J (1983) Optimal design and refinement of the linear model with applications to repeated measurements designs. Ann Stat 11:247–257
13. Kunert J (1984b) Optimality of balanced uniform repeated measurements designs. Ann Stat 12:1006–1017
14. Kushner HB (1997b) Optimal repeated measurements designs: the linear optimality equations. Ann Stat 25:2328–2344
15. Laska E, Meisner M (1985) A variational approach to optimal two treatment crossover designs: application to carryover effect models. J Am Stat Assoc 80:704–710
16. Layard MW, Arvesen JN (1978) Analysis of Poisson data in crossover experimental designs. Biometrics 34:421–428
17. Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22
18. Matthews JNS (1987) Recent developments in crossover designs. Int Stat Rev 56:117–127
19. McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London
20. Pukelsheim F (1993) Optimal design of experiments. Wiley, New York
21. Senn S (2003) Cross-over trials in clinical research, 2nd edn. Wiley, Chichester, England
22. Singh SP, Mukhopadhyay S (2016) Bayesian crossover design for generalized linear models. Comput Stat Data Anal 104:35–50
23. Stufken J (1996) Optimal crossover designs. In: Ghosh S, Rao CR (eds) Handbook of statistics, vol 13. North-Holland, Amsterdam, pp 63–90
24. Stufken J, Yang M (2012) Optimal designs for generalized linear models. In: Hinkelmann K (ed) Design and analysis of experiments, volume 3: special designs and applications. Wiley, New York
25. Zeger SL, Liang KY, Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. Biometrics 44:1049–1060