The Annals of Statistics
2018, Vol. 46, No. 6A, 2939–2959
https://doi.org/10.1214/17-AOS1644

ON MSE-OPTIMAL CROSSOVER DESIGNS[1]

BY CHRISTOPH NEUMANN AND JOACHIM KUNERT

Technische Universität Dortmund

In crossover designs, each subject receives a series of treatments one after the other. Most papers on optimal crossover designs consider an estimate which is corrected for carryover effects. We look at the estimate for direct effects of treatment, which is not corrected for carryover effects. If there are carryover effects, this estimate will be biased. We try to find a design that minimizes the mean square error, that is, the sum of the squared bias and the variance. It turns out that the designs which are optimal for the corrected estimate are highly efficient for the uncorrected estimate.

1. Introduction. In crossover designs, each experimental unit receives a series of treatments in consecutive periods. There is concern that a treatment applied in a given period may, in addition to its direct effect, also have a carryover effect, that is, it may affect the measurement in the next period. In most cases, the experimenter is interested in the direct effects of the treatments. So the experimenter will try to ensure that there are no carryover effects or at least try to minimize them. Attempts to remove carryover effects include washout periods or consumption of a neutral taste to neutralize lingering flavors.

If the carryover effects cannot be eliminated completely, the experimenter may want to apply a model that allows for carryover. Kunert and Sailer (2006) warn against the illusion that the model with carryover effects solves the problem of carryover completely. They state as one of the main disadvantages of the model with carryover effects that experimenters might put less effort in avoiding carryover when they use it. Senn (2002) gives "5 reasons for believing that the simple carry-over model is not useful," Senn (2002), Chapter 10.3. He also argues that experimenters should be more interested in avoiding carryover than in adjusting for it.

On the other hand, Ozan and Stufken (2010) recommend adjusting for carryover effects in each experiment. They showed, however, that the variance of the corrected estimators can get large, especially in more complicated models like the model with self- and mixed-carryover effects or the model with proportional carryover effects, and recommend using designs which minimize the increase of the variance.

Received December 2016; revised October 2017.

[1]Supported by the Deutsche Forschungsgemeinschaft (SFB 823, Teilprojekt C2). Part of the research was carried out while the second author was visiting the Free University of Bozen-Bolzano.

MSC2010 subject classifications. Primary 62K05; secondary 62K10.

Key words and phrases. Optimal design, crossover design, MSE-optimality.

A possible compromise might be analyzing in a model without carryover effects but choosing the design in such a way that the carryover effects have as little impact on the estimates as possible. David et al. (2001) showed that this approach can be quite useful, at least in agricultural studies.

Compared to the vast literature on the optimality of designs in the model with carryover effects, there is only a very small number of papers on the choice of designs if the carryover effect is neglected.

The most relevant paper for our work is Azaïs and Druilhet (1997) who present a bias-criterion, which is similar to the optimality criterion by Kiefer (1975). We note that, apart from the disadvantage of having biased estimates, there is the advantage of a smaller variance of the estimators neglecting the carryover effects. The present paper considers an optimality criterion that gives a compromise between these two opposing attributes. This criterion is the well-known mean square error (MSE).

2. Calculating the MSE. We consider the set of crossover designs $\Omega_{t,n,p}$ with t treatments, n units and p periods. If $d \in \Omega_{t,n,p}$ is applied, then y_{ij}, the jth observation on unit i, arises from a model with additive carryover effects, that is,

$$y_{ij} = \alpha_i + \tau_{d(i,j)} + \rho_{d(i,j-1)} + \varepsilon_{ij}.$$

Here, α_i, $1 \leq i \leq n$, is the effect of the ith unit, $\tau_{d(i,j)}$ is the effect of the treatment given to the ith unit in the jth period by the design d, $\rho_{d(i,j-1)}$ is the carryover effect of the treatment given to unit i in period $(j-1)$, where $\rho_{d(i,0)} = 0$, and ε_{ij} is the error. The errors are independent, identically distributed with expectation 0 and variance σ^2.

In vector notation, this model can be written as

$$y = U\alpha + T_d\tau + F_d\rho + \varepsilon.$$

Here, y is the vector of the y_{ij} and ε is the vector of the errors. The vectors α, τ and ρ are the vectors of the unit, direct and carryover effects, respectively. The matrices U, T_d and F_d are the corresponding design-matrices.

We assume that the analysis of the data is done with a model without carryover effects, that is,

$$y = U\alpha + T_d\tau + \varepsilon.$$

It is hoped that, due to the precautions taken by the experimenter, the carryover effects are vanishingly low or zero. In that case, the uncorrected estimate is unbiased and the estimate which is corrected for carryover effects will have a unnecessarily large variance.

If, however, there are carryover effects, then the uncorrected estimate of the treatment effects is biased. We try to determine a design that minimizes the mean square error (MSE) as a performance measure combining bias and variance. Because the MSE in general is not convex, it is neither a criterion in the sense of Kiefer (1975) nor in the sense of Azaïs and Druilhet (1997).

For an $(n \times a)$-matrix A, we define $\omega^{\perp}(A) = I_n - A(A^T A)^{-} A^T$. Here, A^T is the transpose of A and $(A^T A)^{-}$ is a generalized inverse of $A^T A$.

Using this notation, the joint information matrix of direct and carryover effects can be written as

$$M_d = \begin{bmatrix} M_{d11} & M_{d12} \\ M_{d12}^T & M_{d22} \end{bmatrix},$$

where

$$M_{d11} = T_d^T \omega^{\perp}(U) T_d,$$
$$M_{d12} = T_d^T \omega^{\perp}(U) F_d,$$
$$M_{d22} = F_d^T \omega^{\perp}(U) F_d;$$

see Bose and Dey (2009), page 15. Note that $T_d 1_t = 1_{np} = U 1_n$. Therefore, $1_t^T T_d^T \omega^{\perp}(U) = 0$, implying that M_{d11} and M_{d12} have column-sums 0.

In what follows, we restrict attention to designs which allow estimation of all contrasts of direct effects in the model without carryover effects. Because M_{d11} is the information matrix for direct effects in the model without carryover effects, this is the set of all designs for which $\operatorname{rank}(M_{d11}) = t - 1$. In the model with carryover-effects, we see that for any pair $(i, j), i \neq j$ the MSE of the uncorrected estimate $\widehat{\tau_i - \tau_j}$ then equals

$$E\big(\widehat{\tau_i - \tau_j} - (\tau_i - \tau_j)\big)^2 = \sigma^2 \ell_{ij}^T M_{d11}^{+} \ell_{ij} + \big(\ell_{ij}^T M_{d11}^{+} M_{d12} \rho\big)^2,$$

where M_{d11}^{+} is the Moore–Penrose generalized inverse of M_{d11} and ℓ_{ij} is a t-dimensional vector with +1 in position i, -1 in position j and all other entries 0. If $\operatorname{tr}(M)$ denotes the trace of a matrix M, this can be rewritten as

$$E\big(\widehat{\tau_i - \tau_j} - (\tau_i - \tau_j)\big)^2 = \sigma^2 \operatorname{tr}\big(M_{d11}^{+} \ell_{ij} \ell_{ij}^T\big) + \rho^T \big(M_{d12}^T M_{d11}^{+} \ell_{ij} \ell_{ij}^T M_{d11}^{+} M_{d12}\big) \rho.$$

Noting that

$$\sum_i \sum_{j>i} \ell_{ij} \ell_{ij}^T = t I_t - 1_t 1_t^T$$

and averaging over all pairs $(i, j), i < j$, we observe that the average MSE equals

$$\frac{2}{t-1} \big(\sigma^2 \operatorname{tr}(M_{d11}^{+} H_t) + \rho^T \big(M_{d12}^T M_{d11}^{+} H_t M_{d11}^{+} M_{d12}\big) \rho\big),$$

where $H_t = I_t - \frac{1}{t} 1_t 1_t^T$. Since M_{d11} has row- and column-sums zero, this simplifies to

$$\frac{2}{t-1} \big(\sigma^2 \operatorname{tr}(M_{d11}^{+}) + \rho^T \big(M_{d12}^T M_{d11}^{+} M_{d11}^{+} M_{d12}\big) \rho\big).$$

To reduce the dependence on the unknown parameter ρ, we consider the worst case for given $\rho^T\rho = \sum \rho_i^2 = \delta$, say, that is, we consider

$$\begin{aligned}\frac{2}{t-1}\max_{\rho^T\rho=\delta}&\left(\sigma^2 \operatorname{tr}(M_{d11}^+) + \rho^T(M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12})\rho\right)\\ &= \frac{2}{t-1}\left(\sigma^2 \operatorname{tr}(M_{d11}^+) + \delta\lambda_1(M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12})\right),\end{aligned}$$

where $\lambda_i(M)$ denotes the ith ordered eigenvalue of a symmetric matrix M.

DEFINITION 1. For any $d \in \Omega_{t,n,p}$, we define

$$\operatorname{MSE}(d) = \frac{2}{t-1}\left(\sigma^2 \operatorname{tr}(M_{d11}^+) + \delta\lambda_1(M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12})\right).$$

The advantage of this criterion is that the multivariate purpose of minimizing the bias and maximizing the precision of the estimators can be calculated as a number in $\mathbb{R}$. Our aim is to find a design that minimizes $\operatorname{MSE}(d)$.

Note that $\operatorname{MSE}(d)$ depends on the two unknown parameters σ^2 and δ. The comparison of two designs, however, only depends on the quotient $\frac{\delta}{\sigma^2}$. We therefore assume without loss of generality that $\sigma^2 = 1$.

Define $\mathcal{S}_t$ as the set of all $(t \times t)$-permutation matrices. For any design d, we define the symmetrized version $\bar{M}_{dij}$ of the matrix M_{dij} as

$$\bar{M}_{dij} = \frac{1}{t!}\sum_{\Pi\in\mathcal{S}_t} \Pi^T M_{dij}\Pi,$$

for $1 \le i \le j \le 2$. We call a square matrix M *completely symmetric*, if there are numbers a and b such that all diagonal elements of M are equal to a, while all off-diagonal elements are equal to b. Note that all $\bar{M}_{dij}$ are completely symmetric and that $\operatorname{tr} M_{dij} = \operatorname{tr} \bar{M}_{dij}$. Since M_{d11} and M_{d12} have column-sums zero, it hence is easy to see that

$$\bar{M}_{dij} = \operatorname{tr}(M_{dij})\frac{1}{t-1}H_t$$

for $(i, j) \in \{(1,1), (1,2)\}$.

PROPOSITION 1. *For any design $d \in \Omega_{t,n,p}$, there is a lower bound for* $\operatorname{MSE}(d)$, *namely*

$$\operatorname{MSE}(d) \ge \frac{2(t-1)}{\operatorname{tr}(M_{d11})} + \frac{2\delta}{t-1}\frac{(\operatorname{tr}(M_{d12}))^2}{(\operatorname{tr}(M_{d11}))^2}.$$

Equality holds if M_{d11} and M_{d12} are completely symmetric.

PROOF. The fact that

$$\operatorname{tr}\big(M_{d11}^{+}\big) \geq \frac{(t-1)^2}{\operatorname{tr}(M_{d11})}$$

is standard knowledge. It follows immediately from Kiefer's (1975) Proposition 1.

The lower bound of $\lambda_1(M_{d12}^T M_{d11}^{+} M_{d11}^{+} M_{d12})$ is derived as follows. Note that $\lambda_1(M_{d12}^T M_{d11}^{+} M_{d11}^{+} M_{d12}) = \lambda_1(M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+})$.

Observing that $M_{d11} M_{d11}^{+} M_{d12} = M_{d12}$, we get

$$M_{d12} M_{d12}^T = M_{d11} M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+} M_{d11}.$$

Because $M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+}$ has row- and column-sums 0, we have that

$$M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+} \leq \lambda_1\big(M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+}\big) H_t$$

in the Loewner-sense and, consequently,

$$M_{d12} M_{d12}^T \leq M_{d11} M_{d11} \lambda_1\big(M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+}\big).$$

It is well known [see, e.g., Horn and Johnson (2013), Corollary 7.7.4 (c)] that this implies the same ordering for all eigenvalues, that is, for all $1 \leq i \leq t$ we get

$$\lambda_i\big(M_{d12} M_{d12}^T\big) \leq \lambda_i(M_{d11} M_{d11}) \lambda_1\big(M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+}\big).$$

Since $\lambda_i(M_{d11}) > 0$ and, therefore, $\lambda_i(M_{d11} M_{d11}) > 0$ for $1 \leq i \leq t-1$, we conclude that

$$\lambda_1\big(M_{d11}^{+} M_{d12} M_{d12}^T M_{d11}^{+}\big) \geq \frac{\lambda_i(M_{d12} M_{d12}^T)}{\lambda_i(M_{d11} M_{d11})}. \tag{1}$$

Consider the singular values of M_{d12},

$$\sigma_1(M_{d12}) \geq \cdots \geq \sigma_{t-1}(M_{d12}) \geq \sigma_t(M_{d12}) = 0.$$

From the singular value decomposition, it follows that

$$\operatorname{tr}(M_{d12}) = \operatorname{tr}\left(\begin{bmatrix} \sigma_1(M_{d12}) & & & \\ & \ddots & & \\ & & \sigma_{t-1}(M_{d12}) & \\ & & & \sigma_t(M_{d12}) \end{bmatrix} G\right),$$

where G is an orthonormal matrix. Consequently,

$$\big|\operatorname{tr}(M_{d12})\big| = \left|\sum_{i=1}^{t} \sigma_i(M_{d12}) g_{ii}\right| \leq \sum_{i=1}^{t} \sigma_i(M_{d12}) |g_{ii}|,$$

where g_{ij} is the (i,j)th entry of G. Since G is an orthogonal matrix, all $|g_{ij}| \leq 1$ and we get the well-known inequality between the trace and the sum of the singular values

$$\big|\operatorname{tr}(M_{d12})\big| \leq \sum_{i=1}^{t} \sigma_i(M_{d12}). \tag{2}$$

Could it be that

$$\frac{\sigma_i(M_{d12})}{\lambda_i(M_{d11})} < \frac{|\operatorname{tr}(M_{d12})|}{\operatorname{tr}(M_{d11})}$$

for all $1 \le i \le t-1$?

It would follow that

$$\sum_{i=1}^{t-1} \sigma_i(M_{d12}) < \frac{|\operatorname{tr}(M_{d12})|}{\operatorname{tr}(M_{d11})} \sum_{i=1}^{t-1} \lambda_i(M_{d11}).$$

Since $\sigma_t(M_{d12}) = 0$ and $\lambda_t(M_{d11}) = 0$, this implies that

$$\sum_{i=1}^{t} \sigma_i(M_{d12}) < \left|\operatorname{tr}(M_{d12})\right|.$$

This, however, contradicts inequality (2). Hence, there is an i_0, such that

$$\frac{\sigma_{i_0}(M_{d12})}{\lambda_{i_0}(M_{d11})} \ge \frac{|\operatorname{tr}(M_{d12})|}{\operatorname{tr}(M_{d11})}.$$

Note that $\lambda_i(M_{d11}M_{d11}) = (\lambda_i(M_{d11}))^2$ and that $\lambda_i(M_{d12}M_{d12}^T) = (\sigma_i(M_{d12}))^2$. Inserting this in inequality (1), we have shown that

$$\lambda_1\big(M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12}\big) \ge \frac{(\sigma_{i_0}(M_{d12}))^2}{(\lambda_{i_0}(M_{d11}))^2} \ge \frac{|\operatorname{tr}(M_{d12})|^2}{(\operatorname{tr}(M_{d11}))^2}.$$

It is easy to verify that complete symmetry of M_{d11} and M_{d12} implies equality. Note that, since M_{d11} and M_{d12} have column sums zero, both matrices are completely symmetric if and only if they are multiples of H_t. This completes the proof. $\square$

For any design d, define $q_{dij} = \frac{1}{n}\operatorname{tr}(M_{dij})$ for $1 \le i \le j \le 2$. Then the bound in Proposition 1 can be written as

$$\operatorname{MSE}(d) \ge \frac{2(t-1)}{nq_{d11}} + \frac{2\delta}{t-1}\left(\frac{q_{d12}}{q_{d11}}\right)^2.$$

Each subject in the design d receives a sequence s of treatments. Denote by $T(s)$ and $F(s)$ the part of T_d and F_d that corresponds to s and define $q_{11}(s) = \operatorname{tr}(T(s)^T \omega^\perp(1_p) T(s))$ and $q_{12}(s) = \operatorname{tr}(T(s)^T \omega^\perp(1_p) F(s))$. Two sequences s_1 and s_2 are equivalent if s_1 can be transformed to s_2 by relabelling the treatments. Two equivalent sequences have the same $q_{ij}(s)$. If for given t and p there are K, say, equivalence classes, we choose a representative sequence s_k, $1 \le k \le K$ for each class. As pointed out by Kushner (1997), the q_{dij} then are weighted means of the $q_{ij}(s_k)$. More precisely, we get

$$q_{dij} = \sum_{k=1}^{K} q_{ij}(s_k)\pi_d(k),$$

where $\pi_d(k)$ is the proportion of units of d receiving a sequence from class k, $1 \le k \le K$.

At this point, it makes sense to consider approximate designs. For approximate designs, we remove the restriction that the number of experimental units to receive a given sequence s must be an integer. For an approximate design, the $\pi_d(k)$, $1 \le k \le K$ can be any set of nonnegative real numbers, subject to the condition that $\sum_{k=1}^{K} \pi_d(k) = 1$. An exact design $d \in \Omega_{t,n,p}$ then is a special instance of an approximate design, where each sequence is assigned to an integral number of units. We denote the set of all approximate designs for given t, n and p by $\Delta_{t,n,p}$. Note that the number n of units is not important for an approximate design $d \in \Delta_{t,n,p}$. It plays a role in the calculation of MSE(d), however. A design $d \in \Delta_{t,n,p}$ is called *symmetric*, if each sequence s from class k, $1 \le k \le K$ appears $\pi_d(k)/m_k$ times in the design, where m_k is the number of sequences in class k. Each M_{dij} of a symmetric design is equal to its symmetrized version $\bar{M}_{dij}$. Thus for any combination of q_{d11} and q_{d12} there is a symmetric design with these traces. If the design d is symmetric, then all M_{dij} are completely symmetric, $1 \le i \le j \le 2$.

3. Optimal designs. For any sequence s, we can calculate $q_{11}(s)$ and $q_{12}(s)$ as follows [cf. Kushner (1998) or Bose and Dey (2009), Lemma 4.4.1]:

$$q_{11}(s) = p - \frac{1}{p}\sum_{m=1}^{t} f_{s,m}^2 \quad \text{and} \quad q_{12}(s) = \frac{1}{p}\left(pB_s + f_{s,t_p} - \sum_{m=1}^{t} f_{s,m}^2\right).$$

Here, $f_{s,m}$ is the frequency of treatment m in the sequence, f_{s,t_p} is the frequency of the treatment given in the last period and B_s is the number of periods in which the treatment of the preceding period is repeated.

Four classes of sequences are of special interest. If $p \le t$, we consider classes A and B with representative sequences

$$s_1 = [1, \dots, p-1, p],$$
$$s_2 = [1, \dots, p-1, p-1],$$

respectively. For $p > t$, we can write $p = \ell t + r$, with integers ℓ and r, such that $0 < r \le t$. We then consider classes C and D with representative sequences

$$s_3 = [1, 2, \dots, t, 1, 2, \dots, t, \dots, 1, 2, \dots, t, 1, 2, \dots, r],$$
$$s_4 = [1, \dots, 1, 2, \dots, 2, \dots, t-1, \dots, t-1, t, t, \dots, t],$$

respectively. In s_3 and s_4 each treatment appears either ℓ or $\ell + 1$ times. We assume in s_4 that the first $t - r$ treatments appear ℓ times and the last r treatments are given $\ell + 1$ times. The numbers $q_{11}(s)$ and $q_{12}(s)$ for the four sequences can be seen in Table 1.

Finding an MSE-optimal design is easier for $p > t$. We get the following result.

TABLE 1
The numbers $q_{11}(s)$ and $q_{12}(s)$ for four classes of sequences

Class	**Sequence**	$q_{11}(s)$	$q_{12}(s)$
A	s_1	$p-1$	$-(p-1)/p$
B	s_2	$(p^2-p-2)/p$	0
C	s_3	$((p^2-r)t-p^2+r^2)/pt$	$-(p(p-1)+(r-1)(t-r))/(pt)$
D	s_4	$((p^2-r)t-p^2+r^2)/(pt)$	$(pt(p-t)-p(p-1)-(r-1)(t-r))/(pt)$

PROPOSITION 2. *Assume $p > t$ and consider a symmetric approximate design $d^* \in \Delta_{t,n,p}$ consisting of sequences from class C with proportion*

$$\pi_3 = 1 - \frac{(r-1)(t-r)+p(p-1)}{pt(p-t)}$$

and of sequences from class D with proportion $\pi_4 = 1 - \pi_3$. Then d^* is MSE-optimal in $\Delta_{t,n,p}$: for all $d \in \Delta_{t,n,p}$ we have $\mathrm{MSE}(d) \geq \mathrm{MSE}(d^*)$.

PROOF. In each sequence used by d^*, each treatment in each unit appears either ℓ or $\ell+1$ times. We therefore have that

$$q_{d^*11} = \frac{1}{n}\operatorname{tr}(M_{d^*11}) = \frac{1}{n}\max\{\operatorname{tr}(M_{d11})\} = \max\{q_{d11}\};$$

see, for example, Shah and Sinha (1989), page 17. Therefore, $\frac{2(t-1)}{nq_{d11}}$ is minimal for $d = d^*$ and $\mathrm{MSE}(d) \geq \frac{2(t-1)}{nq_{d^*11}} + \frac{2\delta}{t-1}(\frac{q_{d12}}{q_{d11}})^2 \geq \frac{2(t-1)}{nq_{d^*11}}$. Noting that d^* is symmetric and that $q_{d^*12} = 0$, we observe that $\mathrm{MSE}(d^*) = \frac{2(t-1)}{nq_{d^*11}}$. $\square$

Since $q_{d^*12} = 0$ for the design d^* in Proposition 2, the uncorrected estimate for the direct effects is unbiased. This implies that this design is also bias-optimal in the sense of Azaïs and Druilhet (1997), Proposition 2.3.1.

It is easy to see that the design d^* from Proposition 2 is also universally optimal for the corrected estimate in the model with carryover effects: the information matrix in the simpler model M_{d^*11} is completely symmetric with maximal trace. Since $q_{d^*12} = 0$, the information matrix in the finer model with carryover effects is the same as in the simpler model without carryover effects. This implies that d^* is also universally optimal in the finer model, see Kunert (1983).

For $p \leq t$, we define a set of designs of interest. For any $b \in [0, 1]$, define a symmetric design $g(b) \in \Delta_{t,n,p}$ consisting of a proportion $\pi_{g(b)}(1) = 1 - b$ of sequences from class A and of proposition $\pi_{g(b)}(2) = b$ from class B. We then define $\Gamma_{t,n,p} = \{g(b) \in \Delta_{t,n,p} : b \in [0, 1]\}$. Note that $g(1)$ consists of sequences of class B only. This is the bias optimal design.

PROPOSITION 3. *In the case $p \le t$, consider a symmetric design $g(1) \in \Gamma_{t,n,p}$. Then $q_{g(1),12} = 0$ and the uncorrected estimate for direct effects is unbiased.*

PROOF. Since $g(1)$ uses only sequences which are equivalent to s_2, it follows that $q_{g(1),12} = q_{12}(s_2) = 0$; see Table 1. $\square$

The design $g(1)$ from Proposition 3 is universally bias-optimal in the sense of Azaïs and Druilhet (1997). It is, however, not MSE-optimal: the design performs relatively poorly in the model without carryover effects.

The MSE-optimal design for the case $p \le t$ is more complicated. The following boundaries help to restrict the class of competing designs.

PROPOSITION 4. *Let $p \le t$ and let B_s denote the number of periods in which the treatment of the preceding period is repeated. We then have that*

$$\sum_{m=1}^{t} f_{s,m}^2 \ge (p + 2B_s).$$

PROOF. Without loss of generality, we assume that only treatments $1, \dots, l$ occur in the sequence. Then $l \le p$. For $1 \le m \le p$, define $a_m = f_{s,m} - 1$. Then $\sum_{m=1}^{p} f_{s,m} = p$, and thus $\sum_{m=1}^{p} a_m = 0$. Define $M^* = \{m : f_{s,m} \ge 2\}$, the set of all treatments that occur more than once. We get that

$$\sum_{m=1}^{p} f_{s,m}^2 = \sum_{m=1}^{p} (a_m + 1)^2 = \sum_{m=1}^{p} a_m^2 + p = \sum_{m \in M^*} a_m^2 + \sum_{m \notin M^*} a_m^2 + p.$$

It is easy to see that $\sum_{m \in M^*} a_m^2 \ge \sum_{m \in M^*} a_m \ge B_s$.

Since $\sum_{m \notin M^*} a_m^2 \ge \sum_{m \notin M^*} (-a_m)$ and $\sum_{m \notin M^*} (-a_m) = \sum_{m \in M^*} a_m$, it follows that $\sum f_{s,m}^2 \ge 2B_s + p$. $\square$

Proposition 4 has a corollary, which can be seen easily.

COROLLARY 1. *Under the conditions of Proposition 4, it holds for every sequence s that*

$$q_{11}(s) \le p - \frac{1}{p}(p + 2B_s) = p - 1 - \frac{2}{p} B_s,$$
$$q_{12}(s) \le (B_s - 1)\frac{p-1}{p}.$$

For a design $d \in \Delta_{t,n,p}$, we define $B(d) = \sum_k B_{s_k} \pi_d(k)$, the average number of consecutive identical treatments per unit in the design. Then the corresponding *inequalities hold for the* q_{dij}, *namely*,

$$q_{d11} \leq p - 1 - \frac{2}{p} B(d),$$

$$q_{d12} \leq \big(B(d) - 1\big) \frac{p-1}{p}.$$

With these results, we can determine MSE-optimal designs for the case $p \leq t$. At first, we take a look at the case $p = 2$. Since there are only two possible sequence classes, we get the following proposition.

PROPOSITION 5. *Let* $2 = p \leq t$ *and consider* $g(0) \in \Gamma_{t,n,2}$, *which only consists of sequences being equivalent to* $s_1 = [1, 2]$.
We then have that $\min_{d \in \Delta_{t,n,2}} \mathrm{MSE}(d) = \mathrm{MSE}(g(0))$.

PROOF. Any sequence that is not equivalent to s_1, is equivalent to $s_2 = [1, 1]$, with $B_{s_2} = 1$. Defining π as the proportion of sequences being equivalent to s_1, we get from Corollary 1 that

$$\mathrm{MSE}(d) \geq \frac{2(t-1)}{n\pi} + \left(\frac{-\pi}{2\pi}\right)^2 \frac{2\delta}{t-1} = \frac{2(t-1)}{n\pi} + \frac{\delta}{2(t-1)} \geq \mathrm{MSE}\big(g(0)\big). \qquad \square$$

Now consider the case $3 \leq p \leq t$. When searching for an optimal MSE-optimal design, we then observe that the set $\Gamma_{t,n,p}$ is a complete class. This is shown in the next proposition.

PROPOSITION 6. *Assume* $3 \leq p \leq t$ *and consider an arbitrary design* $d \in \Delta_{t,n,p}$. *Define* $B(d)$ *as in Corollary 1 and* $b_d = \min\{B(d), 1\}$. *For the design* $g(b_d) \in \Gamma_{t,n,p}$, *we then have that*

$$\mathrm{MSE}(d) \geq \mathrm{MSE}\big(g(b_d)\big).$$

The inequality holds true for any $\delta \in [0, \infty)$.

PROOF. From Corollary 1, we know that $q_{d11} \leq p - 1 - \frac{2}{p} B(d)$ and $q_{d12} \leq (B(d) - 1)\frac{p-1}{p}$. Equality holds for $g(b_d)$, since the design consists of sequences from classes A and B only.

As long as $B(d) \leq 1$, we have $q_{d12} \leq 0$ and, therefore, $|q_{d12}| \geq |B(d) - 1| \frac{p-1}{p}$. This implies that

$$\begin{aligned}
\mathrm{MSE}(d) &\geq \frac{2(t-1)}{n q_{d11}} + \left(\frac{q_{d12}}{q_{d11}}\right)^2 \frac{2\delta}{t-1} \\
&\geq \frac{2(t-1)}{n(p - 1 - \frac{2}{p} B(d))} + \left(\frac{\frac{p-1}{p}(B(d) - 1)}{p - 1 - \frac{2}{p} B(d)}\right)^2 \frac{2\delta}{t-1} \\
&= \mathrm{MSE}\big(g\big(B(d)\big)\big).
\end{aligned}$$

If $B(d) > 1$, we see that $q_{d11} < p - 1 - \frac{2}{p} = q_{g(1),11}$ and $|q_{d12}| \geq 0 = q_{g(1),12}$. This implies that

$$\mathrm{MSE}(d) \geq \frac{2(t-1)}{n(p-1-\frac{2}{p})} = \mathrm{MSE}\big(g(1)\big). \qquad \square$$

We know that the design $g(0)$ is universally optimal in the model without carryover effects. Hence, it is MSE-optimal if $\delta = 0$ and it has to be at least highly efficient for small δ. We find that there even is a $\delta^* > 0$ such that $g(0)$ is MSE-optimal for all $\delta \leq \delta^*$. For $\delta > \delta^*$, there is a $b^* > 0$ (depending on δ) such that $g(b^*)$ is MSE-optimal.

PROPOSITION 7. *Assume $3 \leq p \leq t$. Depending on the true δ, we define $b^* \in [0, 1]$ by*

$$b^* = \begin{cases} 0, & \text{if } \delta \leq \dfrac{(t-1)^2 p^2}{n(p-1)(p+1)(p-2)} = \delta^*, \qquad \text{say,} \\ 1 - \dfrac{(t-1)^2 p(p^2-p-2)}{n\delta(p-1)^2(p+1)(p-2) - 2(t-1)^2 p}, & \text{if } \delta > \delta^* \end{cases}$$

Then the design $g(b^*) \in \Gamma_{t,n,p}$ is MSE-optimal over $\Delta_{t,n,p}$, that is, for any $d \in \Delta_{t,n,p}$ we have

$$\mathrm{MSE}(d) \geq \mathrm{MSE}\big(g(b^*)\big).$$

PROOF. Consider an arbitrary design $d \in \Delta_{t,n,p}$. Because of Proposition 6, there is a $b \in [0, 1]$ such that

$$\mathrm{MSE}(d) \geq \frac{2(t-1)}{n(p-1-\frac{2}{p}b)} + \left(\frac{\frac{p-1}{p}(b-1)}{p-1-\frac{2}{p}b}\right)^2 \frac{2\delta}{t-1} = G(b),$$

say. The derivative of $G(b)$ (with respect to b) equals

$$G^{|}(b) = \frac{2}{(p-1-\frac{2}{p}b)^3}\left[\left\{\frac{2(t-1)}{np}(p-1) - \frac{2\delta(p-1)^2}{p^2(t-1)}\frac{p(p-1)-2}{p}\right\}\right.$$
$$\left. + \left\{-\frac{4(t-1)}{np^2} + \frac{2\delta(p-1)^2}{p^2(t-1)}\frac{p(p-1)-2}{p}\right\}b\right].$$

Since $p - 1 - \frac{2}{p}b$ is positive for all $b \in [0, 1]$, the sign of $G^{|}$ is equal to the sign of

$$\frac{2(t-1)}{np}(p-1) - \frac{2\delta(p-1)^2}{p^2(t-1)}\frac{p(p-1)-2}{p}$$
$$+ \left\{-\frac{4(t-1)}{np^2} + \frac{2\delta(p-1)^2}{p^2(t-1)}\frac{p(p-1)-2}{p}\right\}b = N(b),$$

say. Note that $N(b)$ is linear in b and that

$$N(1) = \frac{2(t-1)}{np}\left\{p - 1 - \frac{2}{p}\right\} > 0.$$

On the other hand, $N(0)$ is nonnegative if and only if

$$\delta \le \left(\frac{2(t-1)}{np}(p-1)\right) \Big/ \left(\frac{2(p-1)^2}{p^2(t-1)}\frac{p(p-1)-2}{p}\right) = \delta^*.$$

Hence, for $\delta \le \delta^*$ the bound $G(b)$ is minimal for $b = 0 = b^*$.

For all $\delta > \delta^*$, however, we have $G^{|}(0) < 0$. This implies that $G(b)$ must have a local minimum. Since $N(b)$ is linear, there is only one b for which $G^{|}(b) = 0$, namely

$$b = \frac{\frac{2\delta(p-1)^2}{p^2(t-1)}\frac{p(p-1)-2}{p} - \frac{2(t-1)}{np}(p-1)}{\frac{2\delta(p-1)^2}{p^2(t-1)}\frac{p(p-1)-2}{p} - \frac{4(t-1)}{np^2}} = b^*. \qquad \square$$

Hence, for $p \le t$ the various criteria lead to different designs, all from the family of designs $\Gamma_{t,n,p} = \{g(b) : b \in [0, 1]\}$. While the design $g(1)$ minimizes the bias, the design $g(0)$ is optimal in the reduced model without carryover effects. It is also MSE-optimal, as long as $\delta \le \delta^*$. For $\delta > \delta^*$, the MSE-optimal design is a $g(b)$, where $0 < b < 1$ depends on δ.

Finally, the universally optimal design for the corrected estimate in the model with carryover-effects was determined by Kushner (1998). It is the design $d_{\text{opt}} = g(b)$ with $b = \frac{1}{(p-1)t}$.

4. Optimal designs for the model with period effects.

We extend the model and include period effects, that is, the model becomes

$$y = U\alpha + P\beta + T_d\tau + F_d\rho + \varepsilon,$$

with β the vector of the period effects and P the corresponding design matrix. The information matrices for this model are [see Bose and Dey (2009), page 15]

$$\tilde{M}_{d11} = T_d^T \omega^{\perp}([U, P])T_d,$$
$$\tilde{M}_{d12} = T_d^T \omega^{\perp}([U, P])F_d,$$
$$\tilde{M}_{d22} = F_d^T \omega^{\perp}([U, P])F_d.$$

For a design $d \in \Delta_{t,n,p}$, we define the average mean square error of the uncorrected estimate in the model with period effects as

$$\widetilde{\text{MSE}}(d) = \frac{2}{t-1}(\sigma^2 \operatorname{tr}(\tilde{M}_{d11}^+) + \delta\lambda_1(\tilde{M}_{d12}^T\tilde{M}_{d11}^+\tilde{M}_{d11}^+\tilde{M}_{d12})).$$

Define $\tilde{q}_{d11} = \frac{1}{n}\operatorname{tr}(\tilde{M}_{d11})$ and $\tilde{q}_{d12} = \frac{1}{n}\operatorname{tr}(\tilde{M}_{d12})$. With exactly the same arguments as in Section 2, we then can show that

$$\widetilde{\mathrm{MSE}}(d) \geq \frac{2(t-1)}{\tilde{q}_{d11}} + \frac{2\delta}{t-1}\left(\frac{\tilde{q}_{d12}}{\tilde{q}_{d11}}\right)^2.$$

Equality holds if $\tilde{M}_{d11}$ and $\tilde{M}_{d12}$ are completely symmetric.

Unfortunately, $\operatorname{tr}(\tilde{M}_{d11})$ and $\operatorname{tr}(\tilde{M}_{d12})$ are no longer weighted means of terms that depend on the single sequences only. As shown by Chêng and Wu (1980), $\tilde{q}_{d11}$ and $\tilde{q}_{d12}$ can be written as

$$\operatorname{tr}(\tilde{M}_{d11}) = n\tilde{q}_{d11} = nq_{d11} - \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p} l_{dij}^2 + \frac{1}{np}\sum_{i=1}^{t} r_{di}^2,$$

$$\operatorname{tr}(\tilde{M}_{d12}) = n\tilde{q}_{d12} = nq_{d12} - \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p} l_{dij}\tilde{l}_{dij} + \frac{1}{np}\sum_{i=1}^{t} r_{di}\tilde{r}_{di},$$

where l_{dij} is the number on appearances of treatment i in period j, $\tilde{l}_{dij}$ the number on appearances of treatment i in period $j-1$ with $\tilde{l}_{di1} = 0$, r_{di} the total number of appearances of treatment i in the design and $\tilde{r}_{di}$ the total number of appearances of treatment i in the first $p-1$ periods. The numbers q_{d11} and q_{d12} are as in Section 2.

If $d \in \Delta_{t,n,p}$ is a symmetric design, then $l_{dij} = n/t$, for each i and j. Hence, $\tilde{q}_{d11} = q_{d11}$ and $\tilde{q}_{d12} = q_{d12}$. Note that for any design d, we have $\tilde{q}_{d11} \leq q_{d11}$; see Kunert (1983). Therefore, in the case $p > t$, the results of Proposition 2 extend to the model with period effects.

For $p \leq t$, however, the MSE-optimal design $g(b^*)$ determined in Proposition 7 does not have $q_{g(b^*)12} = 0$. Example 4.6 in Kunert (1983) shows that there are nonsymmetric designs d such that $\tilde{M}_{d12} = 0$, while $M_{d12} \neq 0$. Hence, there are designs that have a smaller bias in the model with period effects than in the model without period effects and the boundaries of Corollary 1 do not hold for $\tilde{q}_{12}(s)$. However, $\tilde{q}_{d12} \neq q_{d12}$ can only be achieved if $\tilde{q}_{d11} < q_{d11} \leq p - 1 - \frac{2}{p}B(d)$. We now show for a wide class of designs that the loss in $\tilde{q}_{d11}$ is higher than the gain in $|\tilde{q}_{d12}|$.

PROPOSITION 8. *Define $\tilde{\Delta}_{t,n,p}$ as the set of all those approximate designs with t treatments, n units and p periods, where all treatments appear equally often, that is, $r_{di} = \frac{np}{t}$, $1 \leq i \leq t$. For any $d \in \tilde{\Delta}_{t,n,p}$, it holds that*

$$q_{d11} - \tilde{q}_{d11} \geq |q_{d12} - \tilde{q}_{d12}|.$$

PROOF. Defining $Q = \frac{1}{n}PP^T - \frac{1}{np}1_{np}1_{np}^T$, we find that

$$nq_{d11} - n\tilde{q}_{d11} = \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p} l_{dij}^2 - \frac{1}{np}\sum_{i=1}^{t} r_{di}^2 = \operatorname{tr}(T_d^T Q T_d)$$

and that

$$nq_{d12} - n\tilde{q}_{d12} = \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p} l_{dij}\tilde{l}_{dij} - \frac{1}{np}\sum_{i=1}^{t} r_{di}\tilde{r}_{di} = \operatorname{tr}\big(T_d^T QF_d\big).$$

Observing that $1_t^T T_d^T Q = 1_{np}^T Q = 0$, we get that $T_d^T QF_d = H_t T_d^T QF_d$ and, therefore, that

$$nq_{d12} - n\tilde{q}_{d12} = \operatorname{tr}\big(H_t T_d^T QF_d\big) = \operatorname{tr}\big(T_d^T QF_d H_t\big).$$

The last equality holds because $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

By the Cauchy–Schwarz inequality we have

$$\big(\operatorname{tr}\big(T_d^T QF_d H_t\big)\big)^2 \le \big(\operatorname{tr}\big(T_d^T QT_d\big)\big)\big(\operatorname{tr}\big(H_t F_d^T QF_d H_t\big)\big).$$

Since $d \in \tilde{\Delta}_{t,n,p}$, it holds that

$$\operatorname{tr}\big(T_d^T QT_d\big) = \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p} l_{dij}^2 - \frac{1}{np}t\left(\frac{np}{t}\right)^2 = \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p} l_{dij}^2 - \frac{np}{t}.$$

Defining $\bar{l}_{d.j} = \frac{1}{t}\sum_{i=1}^{t} l_{dij}$, we get for $d \in \tilde{\Delta}_{t,n,p}$ that $\bar{l}_{d.j} = \frac{n}{t}$ for all j. Hence, $\sum_{j=1}^{p}\bar{l}_{d.j}^2 = p(\frac{n}{t})^2$. It follows that

$$\operatorname{tr}\big(T_d^T QT_d\big) = \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p} l_{dij}^2 - \frac{t}{n}\sum_{j=1}^{p}\bar{l}_{d.j}^2 = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{t}(l_{dij} - \bar{l}_{d.j})^2.$$

Since $\frac{1}{n}PP^T - Q$ is nonnegative definite, we have that $\operatorname{tr}(H_t F_d^T QF_d H_t) \le \frac{1}{n}\operatorname{tr}(H_t F_d^T PP^T F_d H_t)$. For all $1 \le i \le t$, observe that

$$\tilde{l}_{dij} = \begin{cases} 0, & \text{if } j = 1, \\ l_{di,j-1}, & \text{if } j \ge 2. \end{cases}$$

Hence, the entries in the first column of $F_d^T P$ are 0, while the ith entry of column $j+1$ is l_{dij}. Consequently, the entries in the first column of $H_t F_d^T P$ are 0, while the ith entry of column $j+1$ of $H_t F_d^T P$ is $l_{dij} - \bar{l}_{d.j}$. Therefore,

$$\operatorname{tr}\big(H_t F_d^T QF_d H_t\big) \le \frac{1}{n}\operatorname{tr}\big(H_t F_t^T PP^T F_d H_t\big) = \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p-1}(l_{dij} - \bar{l}_{d.j})^2.$$

It follows that

$$\operatorname{tr}\big(H_t F_d^T QF_d H_t\big) \le \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p}(l_{dij} - \bar{l}_{d.j})^2 = \operatorname{tr}\big(T_d^T QT_d\big).$$

Inserting this in the Cauchy–Schwarz inequality, we get that

$$\big(\operatorname{tr}\big(T_d^T QF_d H_t\big)\big)^2 \le \operatorname{tr}\big(T_d^T QT_d\big)^2.$$

This implies the desired inequality. $\square$

We are now in a position to prove the main result of this section.

PROPOSITION 9. *Assume the model with period effects holds and $d \in \tilde{\Delta}_{t,n,p}$. Define the set $\Gamma_{t,n,p} = \{g(b) : b \in [0, 1]\} \subset \tilde{\Delta}_{t,n,p}$ as in Section 3. Then there is a $b_d \in [0, 1]$, such that $|\tilde{q}_{d12}| \geq |q_{g(b_d)12}|$ and $q_{g(b_d)11} \geq \tilde{q}_{d11}$.*

This implies that for all $\delta \in [0, \infty)$, we have

$$\widetilde{\mathrm{MSE}}(d) \geq \widetilde{\mathrm{MSE}}\big(g(b_d)\big).$$

PROOF. We consider two cases.

Case 1: $|\tilde{q}_{d12}| \geq \frac{p-1}{p}$.

The design $g(0)$ consists of sequences from class A only. Therefore, $q_{g(0)12} = -\frac{p-1}{p} \geq -|\tilde{q}_{d12}|$.

Further, $q_{g(0)11} = p - 1$. We know from Corollary 1 that $\tilde{q}_{d11} \leq q_{d11} \leq p - 1 - \frac{2B(d)}{p} \leq p - 1$. Hence, we have shown that

$$\tilde{q}_{d11} \leq q_{g(0)11} \quad \text{and} \quad |\tilde{q}_{d12}| \geq |q_{g(0)12}|.$$

This implies that

$$\begin{aligned}\widetilde{\mathrm{MSE}}(d) &\geq \frac{2(t-1)}{n}\frac{1}{\tilde{q}_{d11}} + \frac{2\delta}{t-1}\left(\frac{\tilde{q}_{d12}}{\tilde{q}_{d11}}\right)^2 \\ &\geq \frac{2(t-1)}{n}\frac{1}{q_{g(0)11}} + \frac{2\delta}{t-1}\left(\frac{q_{g(0)12}}{q_{g(0)11}}\right)^2 = \widetilde{\mathrm{MSE}}\big(g(0)\big),\end{aligned}$$

where the last equality holds because $g(0)$ is symmetric. Note that $\widetilde{\mathrm{MSE}}(d) \geq \widetilde{\mathrm{MSE}}(g(0))$ is true for all $\delta \in [0, \infty)$.

Case 2: $|\tilde{q}_{d12}| < \frac{p-1}{p}$.

Choose $b_d = 1 - |\tilde{q}_{d12}|\frac{p}{p-1} \in (0, 1]$. For the corresponding design $g(b_d)$, we get that

$$q_{g(b_d)12} = (1 - b_d)q_{12}(s_1) + b_d q_{12}(s_2) = -|\tilde{q}_{d12}|$$

and

$$q_{g(b_d)11} = (1 - b_d)q_{11}(s_1) + b_d q_{12}(s_2) = p - 1 - \frac{2b_d}{p}.$$

It remains to show that

$$q_{g(b_d)11} \geq \tilde{q}_{d11}.$$

Define $B(d)$ as in Corollary 1 and $\kappa = \tilde{q}_{d12} - q_{d12}$. We consider two subcases.

Subcase 2a: $B(d) \geq b_d$.

As in case 1, we have $\tilde{q}_{d11} \le p - 1 - \frac{2B(d)}{p}$. Since $B(d) \ge b_d$, we get

$$\tilde{q}_{d11} \le p - 1 - \frac{2b_d}{p} = q_{g(b_d)11}.$$

Subcase 2b: $B(d) < b_d$.
From the definition of b_d, we get

$$(b_d - 1)\frac{p-1}{p} = -|\tilde{q}_{d12}| \le \tilde{q}_{d12}.$$

From Corollary 1, we know that

$$q_{d12} \le \bigl(B(d) - 1\bigr)\frac{p-1}{p}.$$

Therefore,

$$\kappa = \tilde{q}_{d12} - q_{d12} \ge (b_d - 1)\frac{p-1}{p} - \bigl(B(d) - 1\bigr)\frac{p-1}{p} = \bigl(b_d - B(d)\bigr)\frac{p-1}{p} > 0.$$

Since $p \ge 3$, it follows that

$$|\kappa| \ge \bigl(b_d - B(d)\bigr)\frac{2}{p}.$$

We hence get from Proposition 8 that

$$q_{d11} - \tilde{q}_{d11} \ge |\kappa| \ge \frac{2(b_d - B(d))}{p},$$

and, consequently, that

$$\tilde{q}_{d11} \le q_{d11} - (q_{d11} - \tilde{q}_{d11}) \le q_{d11} - \frac{2(b_d - B(d))}{p}.$$

Once again making use of the fact that $q_{d11} \le p - 1 - 2B(d)/p$, we conclude that

$$\tilde{q}_{d11} \le p - 1 - \frac{2B(d)}{p} - \frac{2(b_d - B(d))}{p} = q_{g(b_d)11}.$$

The rest is shown as in Case 1. $\square$

As a direct consequence of Proposition 9, the MSE-optimal design from Proposition 7 remains MSE-optimal in the model with period effects, provided the competing designs are restricted to the equireplicated designs in $\tilde{\Delta}_{t,n,p}$.

5. Efficiency in terms of MSE and examples. MSE-optimality, at least for $p \leq t$, is a local optimality: which design is MSE-optimal depends on the unknown δ. It hence is useful to determine efficient designs, that is, designs that can compete with the locally best designs for a range of δ. That is, we look at various δ and compare the MSE of a given design d to the MSE of the respective MSE-optimal design.

As before, the cases $p > t$ and $p \leq t$ are different. For $p > t$, there is the design d^* from Proposition 2, which is optimum for all $\delta \in [0, \infty)$. So, under the MSE-criterion this design is the obvious choice.

From Proposition 2, we know that the MSE-optimal design is also universally optimal for the corrected estimator. As in example 4.6.4 of Bose and Dey (2009), we consider the case that $p = 6$ and $t = 3$. We search for an n such that an exact MSE-optimal design $d^* \in \Omega_{3,n,6}$ exists. With Proposition 2, we get $\pi_3 = 4/9$. Both for s_3 and for s_4 the number of equivalent sequences in the case $t = 3$ is $3! = 6$. To construct a symmetric MSE-optimal exact design in $\Omega_{3,n,6}$, we hence need that both $n \times \frac{4}{9} \times \frac{1}{6}$ and $n \times \frac{5}{9} \times \frac{1}{6}$ are integers. This is achieved by $n = 54$. We hence can calculate an exact optimal design d^* for $n = 54$, where there are 24 sequences equivalent to $s_3 = [1, 2, 3, 1, 2, 3]$ and 30 sequences to $s_4 = [1, 1, 2, 2, 3, 3]$.

If we do not insist on a symmetric design, we can construct an exact design $\tilde{d} \in \Omega_{3,9,6}$ with the same MSE as the MSE-optimal approximate design in $\Delta_{3,9,6}$: This is achieved by a strongly balanced design $\tilde{d}$ constructed with the construction method of Cheng and Wu (1980).

For $p \leq t$, however, the MSE-optimal design $g(b^*)$ determined in Proposition 7 depends on δ.

To determine the efficiency of a design $d \in \Delta_{t,n,p}$, we calculate

$$\operatorname{Eff}(d) = \operatorname{MSE}\bigl(g(b^*)\bigr)/\operatorname{MSE}(d).$$

For a good design, MSE(d) should not be much larger than the best possible MSE, hence Eff(d) should be as near to 1 as possible. Candidates for good overall designs seem to be the designs $g(0)$, $g(1)$ and the design d_{opt}, which is universally optimal for the corrected estimate.

First, consider $\delta \in [0, \delta^*]$. Then, obviously,

$$\operatorname{Eff}\bigl(g(0)\bigr) = 1$$

and we have to compare the other two designs to $g(0)$. For the design $g(1)$, we then get

$$\operatorname{Eff}\bigl(g(1)\bigr) = \frac{(n\delta(p-1) + p^2(t-1)^2)(p^2 - p - 2)}{(t-1)^2 p^3 (p-1)}.$$

The efficiency of $g(1)$ increases in δ, for $\delta = 0$ it is $1 - 2/(p(p-1))$. Hence, for larger p, $g(1)$ has a relatively high efficiency, even in its worst case $\delta = 0$.

For d_{opt}, the optimal design in the model with carryover effects, we get

$$\operatorname{Eff}(d_{\text{opt}}) = \frac{1}{p^2(p-1)^2} \times \frac{((t-1)^2p^2 + n\delta(p-1))(p(p-1)^2t-2)^2}{((p-1)((p-1)t-1)^2n\delta + (t-1)^2pt(p(p-1)^2t-2))}.$$

At point $\delta = 0$, we get an efficiency of $1 - 2/(pt(p-1)^2)$ which depends on t and p but is relatively near to 1.

Now consider $\delta > \delta^*$. For design $g(0)$, we get

$$\operatorname{Eff}\big(g(0)\big) = \frac{(t-1)^2p^3((p-1)^2(p^2-p-2)n\delta - (t-1)^2p)}{n\delta(p-1)(p+1)^2(p-2)^2((t-1)^2p^2 + n\delta(p-1))}$$

which is decreasing toward 0 when δ increases to ∞.

The design $g(1)$ gets more efficient with increasing δ (because in the optimal design the proportion of sequences s_2 increases). The efficiency calculates to

$$\operatorname{Eff}\big(g(1)\big) = 1 - \frac{(t-1)^2p}{n\delta(p-1)^2(p+1)(p-2)}.$$

The efficiency tends to 1 for $\delta \to \infty$.

Finally, we consider the efficiency of the universally optimal design d_{opt}. Here, we get

$$\operatorname{Eff}(d_{\text{opt}}) = \frac{((p^3-2p^2+p)t-2)^2}{n\delta(p-1)^3(p-2)^2(p+1)^2} \times \frac{(t-1)^2p((p-1)^2(p-2)(p+1)n\delta - (t-1)^2p)}{(p-1)(t(p-1)-1)^2n\delta - (t-1)^2pt(p(p-1)^2t-2)}.$$

The efficiency of d_{opt} first increases until that δ where $b^* = \frac{1}{(p-1)t}$, that is, where d_{opt} is MSE-optimal. For larger δ it decreases, but the efficiency of d_{opt} for those δ is always higher than $\operatorname{Eff}(g(0))$.

To exemplify these general findings, we consider the case $p = 3, t = 4$. The design d_{opt} for the corrected estimate should have proportion $1 - \frac{1}{(p-1)t} = \frac{7}{8}$ of sequences from class A, represented by $s_1 = [1,2,3]$ and $\frac{1}{8}$ of sequences from class B represented by $s_2 = [1,2,2]$. There are $4! = 24$ equivalent sequences in class A and $\frac{4!}{2!} = 12$ equivalent sequences in class B. Hence, the smallest n such that a symmetric exact design d_{opt} could exist is $n = 192$. However, in example 4.6.2 of Bose and Dey (2009) there is an exact design for $n = 48$, which consists of 32 sequences equivalent to s_1 and 6 sequences equivalent to s_2. This design is not symmetric, but it has M_{d11}, M_{d12} and M_{d22} completely symmetric and each treatment appears in each period equally often. Hence, it behaves exactly like a symmetric design $d_{\text{opt}} \in \Delta_{4,48,3}$. Note that there are exact symmetric designs $g(0)$ and $g(1)$ in $\Omega_{4,48,3}$.

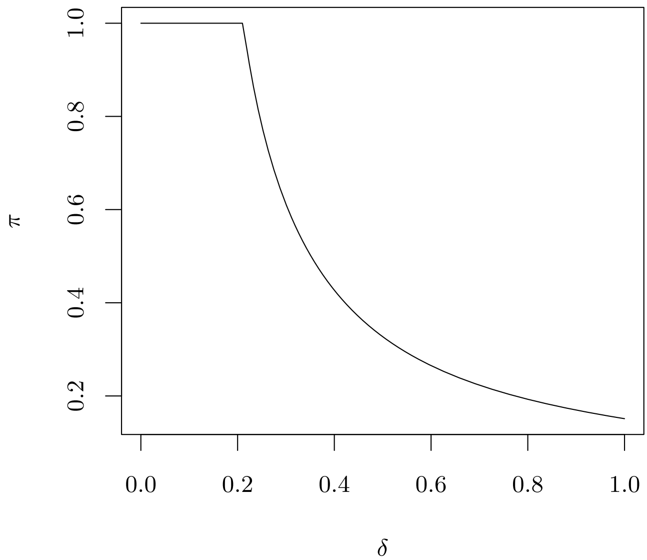

FIG. 1. *Proportion π of s_1 as a function of δ.*

We know from Proposition 7 that the optimal proportion of sequences equivalent to s_1 depends on δ and t. We calculate the boundary δ^* as $\frac{1\cdot 9\cdot 9}{48\cdot 1\cdot 4\cdot 2} = \frac{27}{128} \approx 0.21$. The optimal proportion $1 - b^*$ of s_1 as a function of δ can be seen in Figure 1. When $\delta \geq \delta^*$, we see that the proportion decreases with δ. For finite δ, it will always be positive. At the point $\delta = 1$, the proportion is still 0.15. However, for $\delta \to \infty$ the proportion of sequences from class A will go to 0.

The efficiencies of the three designs of interest are plotted in Figure 2. As predicted from the general case, the efficiency of design $g(1)$ increases with δ and the efficiency of $g(0)$ decreases. The efficiency of the design for the corrected estimator increases for small δ and then also decreases toward zero. As in the general case, the efficiency of d_{opt} for large δ always stays larger than that of $g(0)$.

Since the optimal b^* is in $[0, 1)$, there is a δ for which the design d_{opt} has efficiency 1 but there is no δ such that the efficiency of $g(1)$ equals 1.

However, if we allow for all $\delta \in [0, \infty)$, the minimal efficiency for any design d with $q_{d12} \neq 0$ is 0. The maximum minimal efficiency is achieved by $g(1)$ for which the minimal efficiency is attained for $\delta = 0$ and remains positive.

6. Discussion. We examine the performance of crossover designs in a model where there are carryover effects which are neglected in the analysis. Our criterion is the MSE which combines the variance and the bias of the uncorrected estimator.

We found for $p > t$ that there are designs which are universally optimal in the model without carryover effects, but for which the uncorrected estimator of treatment effects remains unbiased if carryover effects occur. Hence, these designs are clearly MSE-optimal, irrespective of the true size of the carryover effects.

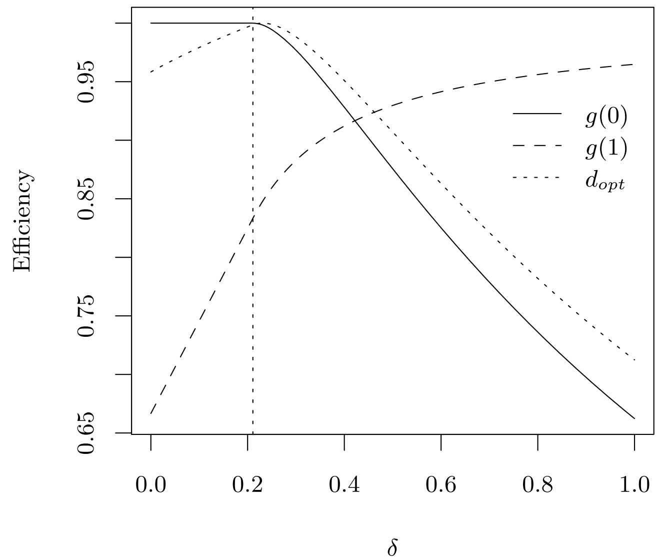

FIG. 2. *MSE-efficiencies for designs* $g(0)$ (*solid line*), $g(1)$ (*dashed line*) *and* d_{opt} (*dotted line*) *for parameters* $p = 3, t = 4, n = 48$ *and* $\delta \in [0, 1]$. *The vertical dashed line marks* δ^*.

If $p \leq t$, the situation is more complicated. Here, the number $B(d)$ of pairs of consecutive identical treatments is important. Among all designs with a given $B(d)$, it is best to have the design as balanced as possible. That includes that each treatment should be preceded by every other treatment equally often.

Note that the construction of an exact symmetric design with a desired $B(d)$ may require a very large number of units. It then makes sense to use a design that tries to come as near to symmetry as possible; see, for example, Kunert and Sailer (2006).

The optimal $B(d)$ depends on the true average size δ of the carryover effects. If there are no carryover effects, then we should choose $B(d) = 0$. When δ increases, the optimal $B(d)$ also increases. It is slightly surprising, however, that even for approximate designs, there is a $\delta^* > 0$, such that $B(d) = 0$ remains MSE-optimal for all $\delta \leq \delta^*$. This agrees with a recommendation that is often given for sensory experiments: use the uncorrected estimate for the treatment effects but a design [with $B(d) = 0$] which is balanced for pairs of consecutive treatments; see, for example, MacFie et al. (1989). This also is supported by the findings of David et al. (2001).

For large δ, however, our results show that designs with $B(d) > 0$ perform better under the MSE-criterion. Which $B(d)$ is the best, then depends on the unknown δ. If an experimenter carries out a series of similar experiments, the size of δ could be estimated from the data of past experiments. However, if there is indication of a large δ in these past experiments, it will be advisable to spend effort on trying to reduce the carryover effects in future experiments. Simply choosing a $B(d) > 0$ is unlikely to be very helpful.

Acknowledgments. We would like to thank the Editor and the two anonymous referees for their careful reading and their constructive remarks that largely improved the readability of this paper.

REFERENCES

AZAÏS, J.-M. and DRUILHET, P. (1997). Optimality of neighbour balanced designs when neighbour effects are neglected. *J. Statist. Plann. Inference* **64** 353–367. MR1621623

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

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

DAVID, O., MONOD, H., LORGEOU, J. and PHILIPPEAU, G. (2001). Control of interplot interference in Grain Maize. *Crop Science* **41** 406.

HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2978290

KIEFER, J. (1975). Construction and optimality of generalized youden designs. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) 333–366. North-Holland, Amsterdam.

KUNERT, J. (1983). Optimal design and refinement of the linear model with applications to repeated measurements designs. *Ann. Statist.* **11** 247–257. MR0684882

KUNERT, J. and SAILER, O. (2006). On nearly balanced designs for sensory trials. *Food Qual. Prefer.* **17** 219–227.

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

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

MACFIE, H. J. H., BRATCHELL, N., GRENHOFF, K. and VALLIS, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *J. Sens. Stud.* **4** 129–148.

OZAN, M. O. and STUFKEN, J. (2010). Assessing the impact of carryover effects on the variances of estimators of treatment differences in crossover designs. *Stat. Med.* **29** 2480–2485. MR2897362

SENN, S. (2002). *Cross-over Trials in Clinical Research*, 2nd ed. *Statistics in Practice*. Wiley, Chichester.

SHAH, K. R. and SINHA, B. K. (1989). *Theory of Optimal Designs*. *Lecture Notes in Statistics* **54**. Springer, New York.

FAKULTÄT STATISTIK
TECHNISCHE UNIVERSITÄT DORTMUND
44221 DORTMUND
GERMANY
E-MAIL: cneumann@statistik.tu-dortmund.de
joachim.kunert@tu-dortmund.de