SFB 823

On MSE-optimal crossover designs

Christoph Neumann, Joachim Kunert

Discussion Paper

Nr. 1/2017

On MSE-optimal crossover designs

Christoph Neumann and Joachim Kunert

Technische Universität Dortmund
Fakultät Statistik
44221 Dortmund
Germany

Abstract: 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.

MSC 2010 subject classifications: Primary 62K05; secondary 62K10.
Keywords and phrases: Optimal Design, crossover Design, MSE-Optimality.

1. Introduction

In crossover designs, each experimental unit gets more than one treatment 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, i.e. it may effect the measurement in the subsequent 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.

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 j-th observation on unit i, arises from a model with additive carryover effects, i.e.

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

Here $\alpha_i, 1 \leq i \leq n$, is the effect of the i-th unit, $\tau_{d(i,j)}$ is the effect of the treatment given to the i-th unit in the j-th period by the design d, $\rho_{d(i,j-1)}$ is the carryover effect of the treatment given to unit i in period $(j-1)$, 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, period, direct and carryover effects, respectively. The matrices U, T_d and F_d are the respective design-matrices.

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

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

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).
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

$$\begin{aligned} M_{d11} &= T_d^T \omega^\perp(U) T_d, \\ M_{d12} &= T_d^T \omega^\perp(U) F_d, \\ M_{d22} &= F_d^T \omega^\perp(U) F_d, \end{aligned}$$

see (Bose and Dey, 2009, p. 17).

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 $rank(M_{d11}) = t - 1$. In the model with carryover-effects, we see that the MSE of the uncorrected estimate $\widehat{\tau_i - \tau_j}$ for any pair $(i, j), i \neq j$ then equals

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

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 $tr(M)$ denotes the trace of a matrix M, this can be rewritten as

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

Noting that

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

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

$$\sigma^2 tr(M_{d11}^+ H_t) + \rho^T (M_{d12}^T M_{d11}^+ H_t M_{d11}^+ M_{d12}) \rho,$$

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

$$\sigma^2 tr(M_{d11}^+) + \rho^T (M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12}) \rho.$$

To reduce the dependence on the unknown parameter ρ, we consider the worst case for given $\rho^T H_t \rho = \sum (\rho_i - \bar{\rho})^2 = \delta$, say, i.e. we consider

$$\begin{aligned} &\max_{\rho^T H_t \rho = \delta} (\sigma^2 tr(M_{d11}^+) + \rho^T (M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12}) \rho) \\ &\qquad = \sigma^2 tr(M_{d11}^+) + \delta \lambda_1 (M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12}), \end{aligned}$$

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

Definition 1. *Let $d \in \Omega_{t,n,p}$. Then we define*

$$MSE(d) = \sigma^2 tr(M_{d11}^+) + \delta\lambda_1(M_{d12}^T M_{d11}^+ M_{d11}^+ M_{d12}).$$

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 $MSE(d)$.
Note that $MSE(d)$ 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 can without any loss of generality assume 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 $\bar{M}_{ij}$ of the matrix M_{ij} as

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

for $1 \leq i \leq j \leq 2$. Note that $tr M_{dij} = tr \bar{M}_{dij}$. Since all M_{dij} have row-sums zero, $1_t^T M_{dij} = 0$, it hence is easy to see that

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

for $1 \leq i \leq j \leq 2$.

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

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

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

Proof. The fact that

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

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

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

in the Loewner-sense and, consequently,

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

This implies the same ordering for all eigenvalues, i.e. for all $1 \leq i \leq t$ we get

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

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

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

Consider the singular values of M_{d12},

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

From the singular value decomposition, it follows that

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

where G is an orthonormal matrix. Consequently,

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

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

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

Could it be that

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

for all $1 \leq i \leq t-1$?
It would follow that

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

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

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

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

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

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 equation (1), we have shown that

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

It is easy to verify that complete symmetry of M_{d11} and M_{d12} implies equality. This completes the proof. □

Thus we can restrict to symmetric designs and calculate the MSE as function of traces instead of matrices. For any design d define $q_{dij} = \frac{1}{n} tr\,(M_{dij})$ for $1 \leq i \leq j \leq 2$. It was shown by Kushner (1997) that the q_{dij} are weighted means of the sequences of treatments in the design. So, if there is a design d with l different sequences $s_1, \ldots, s_l$ with proportions $\pi_{s_1}, \ldots, \pi_{s_l}$, such that $\sum_{j=1}^{l} \pi_{s_j} = 1$ then it follows

$$q_{dij} = \sum_{k=1}^{l} q_{ij}(s_k)\pi_{s_k}.$$

That simplifies the calculation of the bound of MSE and we get

$$MSE(d) \geq \frac{(t-1)^2}{n\, q_{d11}} + \left(\frac{q_{d12}}{q_{d11}}\right)^2 \delta.$$

3. Optimal designs

The MSE of a design depends only on the sequences it uses and their proportions. It is known that each sequence within a symmetry group (i.e. all sequences which are equal when changing treatment labels only) has the same $q_{ij}(s)$, so that we only need to analyze a representative sequence. For any sequence we calculate $q_{11}(s)$ and $q_{12}(s)$ as follows (cf. Kushner (1998) or Bose and Dey (2009)):

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

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 the number of periods, in which the treatment of the period right before was repeated.

There are four special sequences to be investigated:

$$s_1 = [1, \ldots, p]$$

sequence	$q_{11}(s)$	$q_{12}(s)$
s_1	$p-1$	$(1-p)/p$
s_2	$(p^2-p-2)/p$	0
s_3	$((p^2-r)t-p^2+r^2)/(pt)$	$(p(1-p)+(r-1)(r-t))/(pt)$
s_4	$((p^2-r)t-p^2+r^2)/(pt)$	$(pt(p-t)+p(1-p)+(r-1)(r-t))/(pt)$

TABLE 1

Values for $q_{11}(s)$ and $q_{12}(s)$ for the investigated sequences

$$\begin{aligned} s_2 =& [1,\ldots,p-1,p-1] \\ s_3 =& [1,\ldots,t,1,\ldots,t,\ldots,1,\ldots,t,1,\ldots,r] \\ s_4 =& [1,1,\ldots,1,2,2,\ldots,2,\ldots,t,t,\ldots,t]. \end{aligned}$$

Here $r \leq t$. The sequences s_1 and s_2 are only for the case $p \leq t$ and s_3 and s_4 for the case $t > p$. In sequences s_3 and s_4 not all treatments are repeated equally often. While only the number of replications for the treatment given in the last period is of outstanding importance we assume that the last treatment in s_4 is given $(p-r)/t+1$ times. The values of $q_{11}(s)$ and $q_{12}(s)$ for the four mentioned classes of sequences can be seen in Table 1.
We split the problem up in two cases $p > t$ and $p \leq t$. For $p > t$ we get the following result:

Proposition 2. *Let $p > t$ and let $\Delta_{t,p}$ the set of all (approximate) designs with t treatments and p periods.*
Let d^* the design which consists of sequence s_3 with proportion

$$\pi_1 = 1 - \frac{(r-1)(t-r)+p(p-1)}{pt(p-t)}$$

and of sequence s_4 with proportion $\pi_2 = 1-\pi_1$. It holds:

1. *d^* is universally optimal for estimating the treatment effects in the model with additive carryover effects and in the model without it.*
2. *$\forall d \in \Delta_{t,p} : MSE(d) \geq MSE(d^*)$.*

Proof. 1. The design d^* fulfills the requirements of theorem 3 in Kushner (1998).
2. Because d^* is a generalized Youden design, it is universally optimal in the row-column model. (Shah and Sinha, 1989). Therefore $tr(M_{d11})$ is maximized and $\max_{d\in\Delta_{t,p}} q_{d11} = q_{d^*11}$.
Noting that $q_{d^*12} = 0$, the proposition follows.

□

We show that this design is not only universally optimal but also bias-optimal in the sense of Azaïs and Druilhet (1997). In the case $p \leq t$ the bias-optimal design is even simplier.

Proposition 3. *Let $d^* \in \Omega_{t,n,p}$. In case $p \leq t$ let d^* consist only of sequence s_2. In case $p > t$ let d^* be as in proposition 2. It follows: d^* is universally bias-optimal in the sense of Azaïs and Druilhet (1997).*

Proof. We observe for symmetric designs: $M_d = \frac{q_{d12}}{q_{d11}} H_t$ and therefore $tr(M_d) = (t-1)\frac{q_{d12}}{q_{d11}}$. Because M_d and especially M_{d^*} are c.s. and $q_{d^*12} = 0$ in both cases, the proposition follows. □

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

Proposition 4. *Let $p \leq t$ and let B_s be the number of periods, in which the treatment of the period right before was repeated. Then it holds:*

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

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

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

From $\sum_{m=1}^{p} a_m = 0$ we get $\sum_{m \in M^*} a_m = - \sum_{m \notin M^*} a_m$. We further know $\sum_{m \in M^*} a_m^2 \geq \sum_{m \in M^*} a_m \geq B_s$ and thus $\sum_{m \notin M^*} a_m^2 \geq \sum_{m \notin M^*} (-a_m) \geq B_s$.
With that we get

$$\begin{aligned}
\sum_{m \in M^*} (a_m + 1)^2 + \sum_{m \notin M^*} (a_m + 1)^2 &= \sum_{m \in M^*} a_m^2 + 2 \sum_{m \in M^*} a_m + \sum_{m \in M^*} 1 \\
&\quad + \sum_{m \notin M^*} a_m^2 + 2 \sum_{m \notin M^*} a_m + \sum_{m \notin M^*} 1 \\
&= p + \sum_{m \in M^*} a_m^2 + \sum_{m \notin M^*} a_m^2 \\
&\geq p + 2B_s.
\end{aligned}$$

□

With that we can conclude:

Corollary 1. *Let B_s the number of periods, in which the treatment of the period right before was repeated. It holds for every sequence s:*

$$\begin{aligned}
q_{11}(s) &\leq p - \frac{1}{p}(p + 2B_s) = p - 1 - \frac{2}{p}B_s \\
q_{12}(s) &\leq (B_s - 1)\frac{p-1}{p}.
\end{aligned}$$

If we define $B(d)$ of a design d as the weighted mean $B(d) = \sum_s B_s \pi(s)$ this also holds for design d.

With this preliminary work we can calculate 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 let $\Delta_{t,2}$ the set of all (approximate) designs with t treatments and 2 periods.*
Define d_{s_1} as a symmetric design which only consists of sequences $s_1 = [1, 2]$. It holds $\min_{d \in \Delta_{t,2}} MSE(d) = MSE(d_{s_1})$.

Proof. We get that

$$MSE(d) \geq \frac{(t-1)^2}{n\,\pi} + \left(\frac{-\pi}{2\pi}\right)^2 \delta = \frac{(t-1)^2}{n\,\pi} + \frac{1}{4}\delta$$

with π the proportion of sequence s_1. From that the proposition follows. ☐

Even in the case of $p \geq 3$ we can restrict our examination on designs that only consist of s_1 and s_2. This is shown in the following proposition.

Proposition 6. *Let $3 \leq p \leq t$ and let $\Delta_{t,p}$ the set of all (approximate) designs with t treatments and p periods.*
Let $B(d) = \sum_s \pi(s) B_s$, the weighted mean of the number of consecutive treatments in the design. It holds:

$$MSE(d) \geq \frac{(t-1)^2}{n q_{d11}} + \left(\frac{q_{d12}}{q_{d11}}\right)^2 \delta \geq \frac{(t-1)^2}{n\left(p - 1 - \frac{2}{p}B(d)\right)} + \left(\frac{\frac{p-1}{p}(B(d)-1)}{p - 1 - \frac{2}{p}B(d)}\right)^2 \delta$$

with equality if the design is symmetric and consists only of sequences s_1 and s_2.

Proof. From corollary 1 we know that $q_{d11} \leq p - 1 - \frac{2}{p}B(d)$ and $q_{d12} \leq (B(d) - 1)\frac{p-1}{p}$.

As long as $B(d) \leq 1$ we have $q_{d12} \leq 0$ and we get a minimal $MSE(d)$ if q_{d12} is as large as possible.

If $B(d) > 1$, we see that $q_{d11} < p - 1 - \frac{2}{p}$ and $(q_{d12})^2 \geq 0$. This means for any design d with $B(d) > 1$ we get $MSE(d) \geq \frac{(t-1)^2}{n(p-1-\frac{2}{p})} = MSE(d_{s_2})$, with d_{s_2} the design that only consists of s_2. This completes the proof. ☐

We know that the design d_{s_1} is universally optimal if $\delta = 0$ so it has to be optimal in a (small) domain around it. The limit is given by the following proposition.

Proposition 7. *Let $3 \leq p \leq t$ and let $\Delta_{t,p}$ the set of all (approximate) symmetric designs with t treatments and p periods.*
Define d_{s_1} as the symmetric design which only consists of sequence s_1.
Then $\forall d \in \Delta_{t,p}$ it holds: If $\delta \leq \frac{(t-1)^2 p^2}{n(p-1)(p+1)(p-2)}$ then $MSE\left(d_{s_1}\right) \leq MSE\left(d\right)$.

Proof. Let d a design with proportion π for sequence s_1 and proportion $(1-\pi)$ for sequence s_2. Define $v = \frac{(t-1)^2}{n}$. We can find an optimal mixture by differentiating the MSE with respect to π. For given δ we get the minimal MSE, if

$$\pi = \frac{vp(p+1)(p-2)}{-2vp + \delta(p-1)^2(p+1)(p-2)}.$$

We calculate the derivative of π with respect to δ and get:

$$\frac{\partial \pi}{\partial \delta} = \frac{-(vp((p-1)(p+1)(p-2))^2)}{(2vp - \delta(p-1)^2(p+1)(p-2))^2}.$$

Noting that the derivation is negative and the optimal π decreases in δ.
We determine the point δ for that $\pi = 1$ holds and get:

$$\delta_1 = \frac{vp^2}{(p-1)(p+1)(p-2)}.$$

Note that:

$$\frac{\partial MSE\,(d)}{\partial \pi}\Big|_{\pi=0} = \frac{-(2vp)}{(-p^2+p+2)^2} \leq 0.$$

Furthermore the derivative has exactly one root at the optimal π. Thus for $\delta \leq \delta_1$ we get that $MSE(d)$ decreases if π increases and $\pi = 1$ gives the minimal MSE for all mixtures. □

For larger δ the optimal design is a mixture of s_1 and s_2.

Proposition 8. *Let $3 \leq p \leq t$ and let $\Delta_{t,p}$ the set of all (approximate) designs with t treatments and p periods.*
Define d^* as a symmetric design having sequence s_1 with proportion

$$\pi^* = \frac{(t-1)^2 p\,(p\,(1-p)+2)}{2(t-1)^2 p - n\delta\,(p-1)^2\,(p+1)\,(p-2)}$$

and sequence s_2 with proportion $1-\pi^*$.
Then $\forall d \in \Delta_{t,p}$ it holds: If

$$\delta > \frac{(t-1)^2 p^2}{n(p-1)(p+1)(p-2)} = \delta_1 : MSE\,(d^*) \leq MSE\,(d)\,.$$

Proof. We only need to prove that the mixture of π^* sequences with s_1 and $(1-\pi^*)$ with s_2 is the best under all mixtures of s_1 and s_2. Let π the proportion of sequence s_1 in an arbitrary design and let $(1-\pi)$ be the proportion of s_2. Define $v := \frac{(t-1)^2}{n}$. We differentiate the MSE of the design with the respect to π and get:

$$\frac{\partial MSE}{\partial \pi} = \frac{2(\pi(p-1)^2(p+1)(p-2)\delta - vp^2(p-1) + 2v(1-\pi)p)}{(p^2-p+2\pi-2)^3}.$$

Solving for a root with respect to π gives π^*.
We see that π^* decreases with δ but stays greater than 0.
At $\delta = \frac{vp^2}{(p-1)(p+1)(p-2)}$ we observe $\pi = 1$ and because π^* can not be smaller than zero, a valid design always exists. □

4. Optimal Designs for the model with period effects

We extend the model and include period effects, i.e. 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 the model become (with an argument of Kunert and Martin (2000))

$$\begin{aligned}
\tilde{M}_{d11} &= T_d^T \omega^{\perp}\left([U,\, P]\right) T_d) \\
\tilde{M}_{d12} &= H_t F_d^T \omega^{\perp}\left([U,\, P]\right) T_d \\
\tilde{M}_{d22} &= H_t F_d^T \omega^{\perp}\left([U,\, P]\right) F_d H_t
\end{aligned}$$

The calculations of $tr\left(\tilde{M}_{d11}\right)$ and $tr\left(\tilde{M}_{d12}\right)$ are different from the calculations before. Especially it is not possible to express the traces as weighted means of the traces of the sequences. Following Cheng and Wu (1980) these traces can be written as:

$$\begin{aligned}
tr(\tilde{M}_{d11}) &= n\,\tilde{q}_{d11} = n\,q_{d11} - \frac{1}{n}\sum_{i=1}^{t}\sum_{k=1}^{p} l_{dik}^2 + \frac{1}{np}\sum_{i=1}^{t} r_{di}^2 \\
tr(\tilde{M}_{d12}) &= n\,\tilde{q}_{d12} = n\,q_{d12} - \frac{1}{n}\sum_{i=1}^{t}\sum_{k=1}^{p} l_{dik}\tilde{l}_{dik} + \frac{1}{np}\sum_{i=1}^{t} r_{di}\tilde{r}_{di},
\end{aligned}$$

where l_{dik} is the number on appearances of treatment i in period k, $\tilde{l}_{dik}$ the number on appearances of treatment i in period $k-1$ with $\tilde{l}_{di1} = 0$, r_{di} the number on appearances of treatment i and $\tilde{r}_{di}$ the number on appearances of treatment i in the first $p-1$ periods and q_{d11} and q_{d12} are as in section 3. Example 4.6 in Kunert (1983) shows that there are (non-symmetric) desgins such that $\tilde{M}_{d12} = 0$, while $M_{d12} \neq 0$. However, $tr(\tilde{M}_{d12}) \neq tr(M_{d12})$ can only be achieved if $tr(\tilde{M}_{d11}) < tr(M_{d11})$.

Note that for symmetric designs it holds that $tr(\tilde{M}_{d11}) = n\,q_{d11}$ and $tr(\tilde{M}_{d12}) = n\,q_{d12}$.Therefore in the case of $p > t$ the results of Proposition 2 extend to the model with period effects.

For $t \leq p$, we start by showing that for a wide class of designs the loss in $tr(\tilde{M}_{d11})$ is higher than the possible gain in $|tr(\tilde{M}_{d12})|$.

Proposition 9. *Let $\tilde{\Omega}_{t,n,p}$ the set of all designs with t treatments, n units and p periods, where all treatments appear equally often, i.e. $r_{di} = \frac{np}{t}\,, 1 \leq i \leq t$. If $d \in \tilde{\Omega}_{t,n,p}$, it holds that*

$$tr\left(T_d^T \omega\left(\omega^{\perp}\left(U\right)P\right)T_d\right) \geq |tr(H_t F_d^T \omega\left(\omega^{\perp}\left(U\right)P\right)T_d)|.$$

Proof. From the Cauchy-Schwarz inequality it is known that $|\tilde{q}_{d12}| \leq \sqrt{\tilde{q}_{d11}\tilde{q}_{d22}}$. Similarly it holds for $Q = \frac{1}{n}PP^T - \frac{1}{np}1_{np}1_{np}^T = \omega\left(\omega^{\perp}\left(U\right)P\right)$:

$$tr\left(T_d^T Q T_d\right) tr\left(H_t F_d^T Q F_d H_t\right) \geq |tr(H_t F_d^T Q T_d)|^2.$$

We know that while $tr\left(T_d^T Q T_d\right)$ is the loss in $\tilde{M}_{d11}$, $|tr(H_t F_d^T Q T_d)|$ is the gain in $\tilde{M}_{d12}$. Thus it satisfies to show that $tr\left(T_d^T Q T_d\right) \geq tr\left(H_t F_d^T Q F_d H_t\right)$. Therefore we calculate:

$$
\begin{aligned}
tr\left(T_d^T Q T_d\right) =& \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p}\left(l_{dij}-\bar{l}_{d\cdot j}\right)^2 - \frac{1}{np}\sum_{i=1}^{t}\left(r_{di}-\bar{r}_{d\cdot}\right)^2 \\
tr\left(H_t F_d^T Q F_d H_t\right) =& \frac{1}{n}\sum_{i=1}^{t}\sum_{j=o}^{p-1}\left(l_{dij}-\bar{l}_{d\cdot j}\right)^2 - \frac{1}{np}\sum_{i=1}^{t}\left(\tilde{r}_{di}-\bar{\tilde{r}}_{d\cdot}\right)^2 \\
=& \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{p}\left(l_{dij}-\bar{l}_{d\cdot j}\right)^2 - \frac{1}{np}\sum_{i=1}^{t}\left(r_{di}-\bar{r}_{d\cdot}\right)^2 \\
& -\frac{1}{n}\sum_{i=1}^{t}\left(l_{dip}-\bar{l}_{d\cdot p}\right)^2 + \frac{2}{np}\sum_{i=1}^{t}\left(r_{di}-\bar{r}_{d\cdot}\right)\left(l_{dip}-\bar{l}_{d\cdot p}\right) \\
& -\frac{1}{np}\sum_{i=1}^{t}\left(l_{dip}-\bar{l}_{d\cdot p}\right)^2 \\
=& tr\left(T_d^T Q T_d\right) - \frac{1}{np}\sum_{i=1}^{t}(p+1)\left(l_{dip}-\bar{l}_{d\cdot p}\right)^2 \\
& +\frac{2}{np}\sum_{i=1}^{t}(r_{di}-\bar{r}_{d\cdot})(l_{dip}-\bar{l}_{d}\cdot p)
\end{aligned}
$$

If $r_{di} = \frac{np}{t}$, $1 \leq i \leq t$, it holds:

$$
tr\left(H_t F_d^T Q F_d H_t\right) - tr\left(T_d^T Q T_d\right) = -\frac{1}{np}\sum_{i=1}^{t}(p+1)\left(l_{dip}-\bar{l}_{d\cdot p}\right)^2 \leq 0.
$$

□

With that result we are now able to show that for every design where each treatment appears equally often there is a symmetric design with lower or equal bias.

Proposition 10. *Let d be an arbitrary design with $p \leq t$ and let $\tilde{\Delta}_{t,p}$ the set of all symmetric (approximate) designs with t treatments, $p \geq 3$ periods and $r_{di} = \frac{np}{t}$, $1 \leq i \leq t$.*
It holds: $\forall d \in \tilde{\Omega}_{t,n,p}\ \exists d^* \in \tilde{\Delta}_{t,p}$ with $\left(\frac{\tilde{q}_{d12}}{\tilde{q}_{d11}}\right)^2 \geq \left(\frac{q_{d^*12}}{q_{d^*11}}\right)^2$ and $q_{d^*11} \geq \tilde{q}_{d11}$.

Proof. Case 1: $\left(\frac{\tilde{q}_{d12}}{\tilde{q}_{d11}}\right)^2 \geq \left(\frac{1}{p}\right)^2$.
Define d_{s_1} as the design which only consists of sequence s_1. It holds:

$$
\left(\frac{q_{d_{s_1}12}}{q_{d_{s_1}11}}\right)^2 = \left(\frac{1}{p}\right)^2 \leq \left(\frac{\tilde{q}_{d12}}{\tilde{q}_{d11}}\right)^2 .
$$

With $\tilde{q}_{d11} \leq q_{d11} \leq q_{d_{s_1}11}$ the proposition follows.

Case 2: $\left(\frac{\tilde{q}_{d12}}{\tilde{q}_{d11}}\right)^2 < \left(\frac{1}{p}\right)^2$.

Let d_2 the symmetric design that consists of sequences s_1 with proportion $\pi = |\tilde{q}_{d12}|\frac{p}{p-1}$ and sequences s_2 with proportion $1-\pi$.
It follows that $|\tilde{q}_{d12}| = |q_{d_212}|$ and $q_{d_211} = p - 1 - \frac{2}{p} + \frac{2|\tilde{q}_{d12}|}{p-1}$.
One can split up the traces $\tilde{q}_{d11}$ and $|\tilde{q}_{d12}|$ in the parts one gets in the model without period effects (the 'old' traces) and the gain respectively the loss towards the period effects and get: $\tilde{q}_{d11} = q_{d11} - q_{d11,diff}$ and $|\tilde{q}_{d12}| \geq |q_{d12}| - |q_{d12,diff}|$.
We need to show that $\tilde{q}_{d11} < \tilde{q}_{d_211}$.We get

$$\begin{aligned}
q_{d_211} - \tilde{q}_{d11} =& p - 1 - \frac{2}{p} + \frac{2|\tilde{q}_{d12}|}{p-1} - \tilde{q}_{d11} \\
\geq& p - 1 - \frac{2}{p} + \frac{2}{p-1}\left(|q_{d12}| - |q_{d12,diff}|\right) \\
&- q_{d11}(d) + q_{d11,diff} \\
=& p - 1 - \frac{2}{p} + \frac{2}{p-1}|q_{d12}| - q_{d11}(d) + \frac{p-3}{p-1}|q_{d12,diff}| \\
\geq& p - 1 - \frac{2}{p} + \frac{2}{p-1}|q_{d12}| - q_{d11}
\end{aligned}$$

If $\tilde{d}$ is a mixture of s_1 and s_2 we get

$$\frac{2}{p-1}|q_{\tilde{d}12}| - q_{\tilde{d}11} = \frac{2}{p} - p + 1$$

and with corollary 1 it follows:

$$\frac{2}{p-1}|q_{d12}| - q_{d11} \geq \frac{2}{p} - p + 1$$

for (an arbitrary) design d. Thus we get

$$q_{d_211} - \tilde{q}_{d11} \geq 0$$

and the proposition follows. □

Thus the calculations of the sections before hold for a wide class of designs even in the model with period effects.

5. Efficiency in terms of MSE

As one usually is not choosing the model for MSE-optimality reasons but for reasons like optimal estimation of the main effects it is useful to take a look at the efficiency, i.e. the ratio between the optimal and the actual MSE. We take the ratio

$$Eff(d) = MSE_{opt}/MSE(d)$$

because our goal is to minimize the MSE. We examine the efficiency of three different designs d_{s_1} respectively d_{s_3}, d_{s_2} respectively d_{s_4} and d_{opt}, the optimal design in the model with carryover effects. As before we define $v = \frac{(t-1)^2}{n}$.

Again, we split the problem up in two cases. First we examine the case $p > t$. Obviously we get:

$$Eff(d_{opt}) = 1$$

as the MSE-optimal and the optimal design for estimating direct effects fall together.

For the design d_{s_3} we get:

$$Eff\,(d_{s_3}) = \frac{\overbrace{pt(t-1)^2((t-1)p^2 + r(r-t))}^{e_d}}{e_d + \delta n((p-r)(p+r-1) + t(r-1))^2}.$$

At $\delta = 0$ the efficiency is 1 (due to its optimality) and falls towards zero with growing δ.

For the design d_{s_4} we get:

$$Eff\,(d_{s_4}) = \frac{e_d}{e_d + \delta n((t-1)p^2 + (1-t^2)p + (r-t)(r-1))^2}.$$

We observe similar behavior for the efficiency as for d_{s_3}.

Now focus on the case: $p \leq t$. First let $\delta \in [0, \delta_1)$. Obviously

$$Eff(d_{s_1}) = 1.$$

For design d_{s_2} we get:

$$Eff(d_{s_2}) = \frac{(\delta(p-1) + p^2 v)(p^2 - p - 2)}{vp^3(p-1)}.$$

The efficiency increases in δ and is $1 - 2/(p(p-1))$ at $\delta = 0$. For higher p we get a relatively high efficiency even in the worst case.

For the optimal design in the model with carryover effects we get:

$$Eff(d_{opt}) = \frac{\left(vp^2 + \delta\,(p-1)\right)\left(p\,(p-1)^2\,t - 2\right)^2}{p^2\,(p-1)^2\left((p-1)\,((p-1)\,t-1)^2\,\delta + vpt\left(p\,(p-1)^2\,t-2\right)\right)}.$$

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 assume that $\delta > \delta_1$. For design d_{s_1} we get:

$$Eff(d_{s_1}) = \frac{vp^3\left((p-1)^2\left(p^2-p-2\right)\delta - vp\right)}{\delta\,(p-1)\,(p+1)^2\,(p-2)^2\,(vp^2 + \delta\,(p-1))}.$$

which is rapidly decreasing towards 0 when δ increases.
The design d_{s_2} is the other way round (because in the optimal design the proportion of sequences s_2 increases). Here the efficiency calculates to:

$$Eff(d_{s_2}) = 1 - \frac{vp}{\delta(p-1)^2(p+1)(p-2)}.$$

The efficiency tends to 1 for bigger δ.
The last but most important efficiency is the efficiency of the universally optimal design. In this case we get:

$$\begin{aligned} Eff\left(d_{opt}\right) = & \frac{\left(\left(p^3-2p^2+p\right)t-2\right)^2}{\delta\left(p-1\right)^3\left(p-2\right)^2\left(p+1\right)^2} \\ & \times \frac{vp\left(vp-\left(p-1\right)^2\left(p-2\right)\left(p+1\right)\delta\right)}{\left(1-p\right)\left(t\left(1-p\right)+1\right)^2\delta - vpt\left(p\left(p-1\right)^2 t-2\right)}. \end{aligned}$$

The efficiency of d_{opt} first increases until the δ where the model is MSE-optimal. Afterwards it increases but not that rapidly as d_{s_1} does.
We can summarize that the model that are optimal for estimating treatment effects are highly efficient in the MSE if the δ is small.

6. Example

In this section we discuss two examples. First we examine the case $p > t$.

With Proposition 2 we know that the MSE-optimal is an optimal design for estimating treatment effects. As in example 4.6.4 of Bose and Dey (2009) let $p = 6, t = 3$. We can calculate an exact optimal design for $n = 54$. With Proposition 2 we get $\pi_1 = 4/9$ so that there are 24 sequences of $s_3 = [1, 2, 3, 1, 2, 3]$ and 30 sequences of $s_4 = [1, 1, 2, 2, 3, 3]$.
In Figure 1 the efficiencies for the three different sequences are shown. We see that the efficiencies of d_{s_3} and d_{s_4} fall rapidly when δ increases but that the efficiency of d_{s_3} is lower than that of d_{s_4}. While the MSE-optimal and the optimal design fall together the efficiency is 1. Note that all designs have efficiency 1 at the point $\delta = 0$.
There is the strong recommendation to use d_{opt} as it is optimal in the models with and without carryover effects and further is MSE- and bias-optimal.

Now, take a look at the case $p \le t$. Let $p = 3, t = 4$. Bose and Dey (2009) give in example 4.6.2 an exact design for $n = 48$ which consists of the sequences $s_1 = [1, 2, 3]$ and $s_2 = [1, 2, 2]$. We know from the propositions 7 and 8 that the optimal proportion of sequence s_1 depends on δ and t. We calculate the boundary δ_1 as $\frac{1 \cdot 9 \cdot 9}{48 \cdot 1 \cdot 4 \cdot 2} = \frac{27}{128} \approx 0.21$. The optimal proportion of s_1 as a function of δ can be seen in figure 2. We see that the proportion falls rapidly right after δ_1 but it will never reach zero. At the point $\delta = 1$ the proportion is still 0.15.
With that knowledge we are now able to calculate the efficiencies of the three

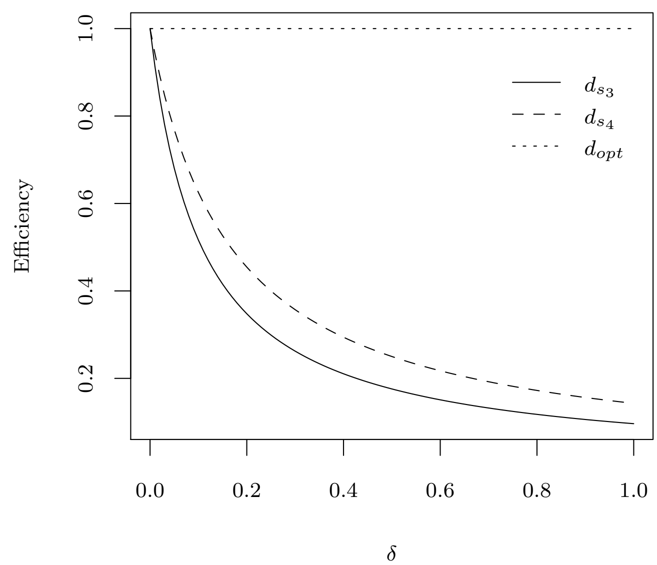

FIG 1. *MSE-efficiencies for designs d_{s_3} (solid line), d_{s_4} (dashed line) and d_{opt} (dotted line) for parameters $p = 6, t = 3, n = 54$ and $\delta \in [0, 1]$.*

competing designs. We plot them in figure 3. It is not surprising that the efficiency of design d_{s_2} increases with δ and the efficiency of d_{s_1} decreases. The most interesting curve is the one of the optimal design for estimating treatment effects. The efficiency of that design increases a short time and then falls towards zero but not that strong as the design d_{s_1} does. Since the optimal proportion π is in $(0, 1]$ there is a δ for that the designs d_{s_1} and d_{opt} have efficiency 1 but there is no δ such that the efficiency of $d_{s_2} = 1$.

7. Discussion

We examine the MSE-optimality of crossover designs when neglecting the carryover effects of the design. Besides its theoretical justification there is a reason for this designs in practice even if a MSE-optimal design may not be the design one chooses. We found that the balanced block design with no self-adjacencies gives the estimators with the lowest MSE in a domain around 0. We also found that the optimal design in the bigger model gives highly efficient estimators in terms of MSE. The MSE-optimal design consists of sequences that are known to deliver good designs for other criterion like Φ_p-optimality. As the optimal design

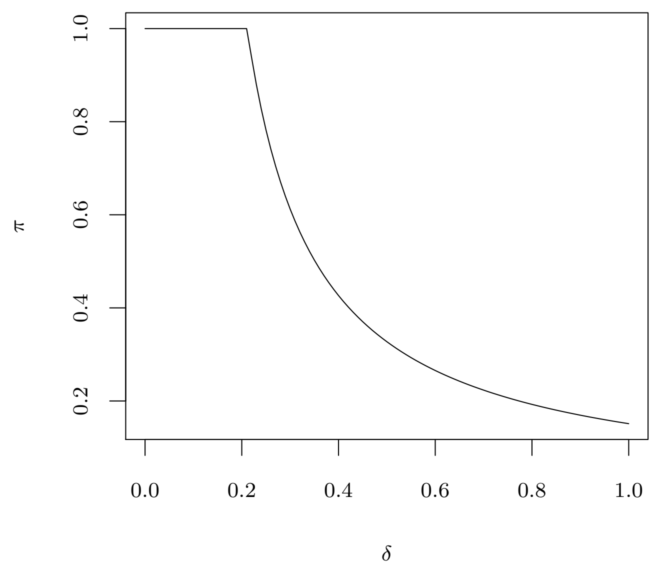

FIG 2. *Proportion π of s_1 as a function of δ.*

in the bigger model is highly efficient we recommend to plan an experiment as if there are carryover effects, try to avoid these effects and analyze the experiment without any carryover effects. In case of the non-existence of carryover effects there will be a higher chance of finding treatments differences. This also confirms the findings of David et al. (2001).

Acknowledgements

The nancial support by the Deutsche Forschungsgemeinschaft (SFB 823, Statistik nichtlinearer dynamischer Prozesse) is gratefully acknowledged.

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.

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

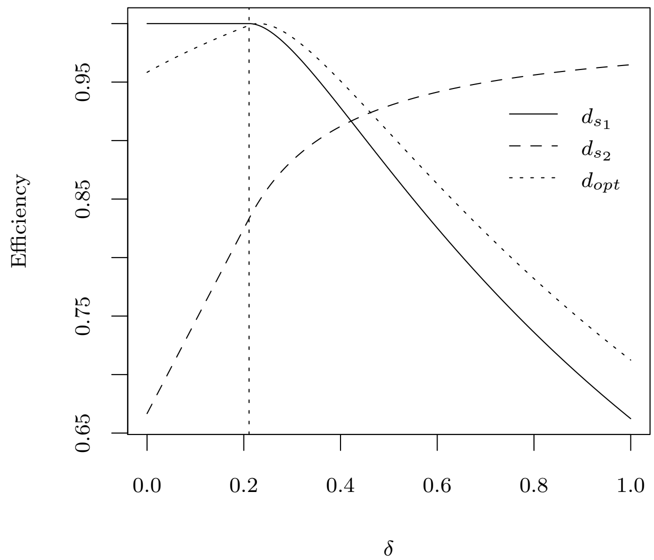

FIG 3. *MSE-efficiencies for designs d_{s_1} (solid line), d_{s_2} (dashed line) and d_{opt} (dotted line) for parameters $p = 3, t = 4, n = 48$ and $\delta \in [0, 1]$. The vertical dashed line marks δ_1.*

CHENG, C.-S. and WU, C.-F. (1980). Balanced Repeated Measurements Designs. *Ann. Statist.* **8** 1272–1283.

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

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.

KUNERT, J. and MARTIN, R. J. (2000). On the Determination of optimal Designs for an Inference Model. *Ann. Statist.* **28** 1728–1742.

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

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

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

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.

Senn, S. (2002). *Cross-over trials in clinical research*, 2 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.