

## Research Article

# Maximin Efficiencies under Treatment-Dependent Costs and Outcome Variances for Parallel, AA/BB, and AB/BA Designs

Math J. J. M. Candel 

Department Methodology and Statistics, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

Correspondence should be addressed to Math J. J. M. Candel; [math.candel@maastrichtuniversity.nl](mailto:math.candel@maastrichtuniversity.nl)

Received 26 January 2018; Accepted 19 July 2018; Published 1 October 2018

Academic Editor: Reinoud Maex

Copyright © 2018 Math J. J. M. Candel. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

If there are no carryover effects, AB/BA crossover designs are more efficient than parallel (A/B) and extended parallel (AA/BB) group designs. This study extends these results in that (a) optimal instead of equal treatment allocation is examined, (b) allowance for treatment-dependent outcome variances is made, and (c) next to treatment effects, also treatment by period interaction effects are examined. Starting from a linear mixed model analysis, the optimal allocation requires knowledge on intraclass correlations in A and B, which typically is rather vague. To solve this, maximin versions of the designs are derived, which guarantee a power level across plausible ranges of the intraclass correlations at the lowest research costs. For the *treatment effect*, an extensive numerical evaluation shows that if the treatment costs of A and B are equal, or if the sum of the costs of one treatment and measurement per person is less than the remaining subject-specific costs (e.g., recruitment costs), the maximin crossover design is most efficient for ranges of intraclass correlations starting at 0.15 or higher. For other cost scenarios, the maximin parallel or extended parallel design can also become most efficient. For the *treatment by period interaction*, the maximin AA/BB design can be proven to be the most efficient. A simulation study supports these asymptotic results for small samples.

## 1. Introduction

The standard design of a randomized clinical trial is the parallel group design: subjects are randomly assigned to one of two treatments, say A or B. An alternative, well-known design is the AB/BA crossover trial in which subjects receive both treatments, A and B, but the sequencing of the treatments is opposite for two randomly allocated groups [1, 2]. An AB/BA crossover trial is considered most suited when examining treatments for chronic or ongoing diseases, such as rheumatism, chronic obstructive pulmonary disease, or (frequent) heartburn. In these cases, there is no real possibility that the disease gets cured, and the aim is to moderate the effects of the disease [2]. A third design that we will consider involves treatment sequences AA and BB. This design extends the parallel design across two treatment periods, allows for testing treatment by time interaction effects, and is a realistic alternative for the AB/BA design in case the treatment regime should continue.

If the outcome variable is continuous and (approximately) normally distributed, the data can be analyzed by mixed effects regression [3]. Of primary interest is testing the treatment effect of, for instance, a new medication for chronic obstructive pulmonary disease. A relevant issue then is which design is the most efficient in estimating the treatment effect, thereby yielding maximum power for testing this effect. Such optimality has already been examined when comparing crossover and parallel designs [2] and when comparing all three designs introduced before [4, 5]. If there are no carryover effects and no dropouts, the sample sizes are equal and equally allocated to the treatments, an AB/BA design yields more efficient estimates of the treatment effect than a parallel and extended parallel design and consequently, will yield more power to test this effect.

The present study extends results on the relative efficiencies of these designs in that (a) optimal instead of equal treatment allocation is examined, (b) allowance is made for treatment-dependent outcome variances, and (c) next to treatment effects, also treatment by period interaction effects are examined.

Outcome variances may differ between treatments [6, 7]. This also is to be expected if treatments differ in terms of their effectiveness. Furthermore, since research costs and outcome variances may differ between treatments, equal allocation to treatments may not be the most efficient. The issue then is how to allocate subjects to treatments such that a design's efficiency is optimized, and how different designs relate in terms of efficiency under such optimal allocation. Optimal allocation requires a priori knowledge on parameters of the analysis model, that is, intraclass correlations for the mixed effects model that we consider. Since this knowledge typically is rather vague, optimal allocations and corresponding efficiencies for maximin versions of the (extended) parallel design and crossover design will be derived. These maximin designs guarantee a power level across plausible ranges of the intraclass correlations at the lowest research costs.

In designs where treatments are successively given to the same group of subjects, carryover may occur. For the AB/BA trial, it may be that, in the AB sequence treatment, A still has an effect on the outcome, when B has been given and the second measurement is done. When in the BA sequence, the effect of B is present, once A has been administered and this effect differs from the carryover effect for the AB sequence, differential carryover occurs. The present study assumes that differential carryover can be safely excluded or is negligible and that this effect does not need to be estimated in analyzing the data.

The paper is structured as follows. Section 2 will present the linear mixed model for analyzing data from each of the three designs. Section 3 will introduce the efficiency criterion and will provide asymptotic expressions for this criterion in the case of maximum likelihood estimation of the treatment effect. Starting from a flexible cost function, optimal allocations to treatments will be derived as well as resulting design efficiencies. Since the efficiencies depend on the intraclass correlations and knowledge on these parameters is often limited, in Section 4, we will derive maximin designs. Section 5 will show to what extent the asymptotic efficiencies translate into desired power levels for small sample sizes. Section 6 will give an application of the results, and Section 7 will discuss some issues for further research.

## 2. Linear Mixed Effects Models

In the case of a parallel design, an extended parallel design, and a crossover design, the subjects are randomly allocated to one of the two arms. In a parallel design to treatment A or treatment B, in an extended parallel design, they are allocated to treatment sequence AA or BB, and in a crossover trial to treatment sequence AB or BA. We consider a quantitative outcome variable, denoted as  $y_{ij}$  for person  $j$  ( $j = 1, \dots, N$ ) at measurement occasion  $i$ , and assume  $y_{ij}$  is (approximately) normally distributed.

For a parallel design and outcome variances that differ between treatments A and B, simple linear regression with heterogeneous variances may be an adequate tool for data analysis:

$$y_{ij} = \beta_0 + \beta_1 \text{treat}_{ij} + \delta_{ij}^A (1 - \text{treat}_{ij}) + \delta_{ij}^B \text{treat}_{ij}, \quad (1)$$

where treatment is coded 0 for persons having treatment A and coded 1 for persons having treatment B, and  $\delta_{ij}^A$  and  $\delta_{ij}^B$  are normally distributed, with mean 0 and variances  $\sigma_A^2$  and  $\sigma_B^2$ , respectively. The random terms  $\delta_{ij}^A$  and  $\delta_{ij}^B$  can be thought of as consisting of a random person (between-subject) effect,  $u_{0j}$ , and a treatment-dependent random error (within-subject) effect,  $\varepsilon_{ij}^A$  and  $\varepsilon_{ij}^B$ . In formula,  $\delta_{ij}^A = u_{0j} + \varepsilon_{ij}^A$ , and  $\delta_{ij}^B = u_{0j} + \varepsilon_{ij}^B$ . These two sources of random variation cannot be separated in a single-period parallel trial.

For a crossover AB/BA design and an extended, two-period, parallel AA/BB design, however, the variances of  $u_{0j}$  and of  $\varepsilon_{ij}^A$  and  $\varepsilon_{ij}^B$  can be identified. The linear regression model can then be extended with a random intercept as well as a fixed effect of time, yielding the following mixed effects model:

$$y_{ij} = \beta_0 + \beta_1 \text{treat}_{ij} + \beta_2 \text{time}_{ij} + u_{0j} + \varepsilon_{ij}^A (1 - \text{treat}_{ij}) + \varepsilon_{ij}^B \text{treat}_{ij}. \quad (2)$$

In (2), time is coded 0 for observations at the first measurement and coded 1 for observations at the second measurement. The random terms  $u_{0j}$ ,  $\varepsilon_{ij}^A$ , and  $\varepsilon_{ij}^B$  are independently normally distributed, with mean 0 and variances  $\sigma_0^2$ ,  $\sigma_{\varepsilon_A}^2$ , and  $\sigma_{\varepsilon_B}^2$ , respectively. Their relation with the variances in (1) is  $\sigma_A^2 = \sigma_0^2 + \sigma_{\varepsilon_A}^2$  for treatment A and  $\sigma_B^2 = \sigma_0^2 + \sigma_{\varepsilon_B}^2$  for treatment B.

In the case we want to examine whether there is an interaction between treatment and period, the model in (2) is extended as follows:

$$y_{ij} = \beta_0 + \beta_1 \text{treat}_{ij} + \beta_2 \text{time}_{ij} + \beta_3 \text{treat}_{ij} \times \text{time}_{ij} + u_{0j} + \varepsilon_{ij}^A (1 - \text{treat}_{ij}) + \varepsilon_{ij}^B \text{treat}_{ij}, \quad (3)$$

where  $\beta_3$  represents the treatment by period interaction effect. The parameters in (1)–(3) can be estimated through maximum likelihood (ML). In what follows, we are interested in optimally estimating  $\beta_1$  in (1) and (2), which will be denoted as  $\beta_{\text{treat}}$ , and in optimally estimating  $\beta_3$  in (3), which will be denoted as  $\beta_{\text{treat} \times \text{time}}$ . A relevant concept is the intraclass correlation, which is between-subject variation on the outcome as compared to the total outcome variation. For the models in (2) and (3), this can be expressed as  $\rho_A = \sigma_0^2 / (\sigma_0^2 + \sigma_{\varepsilon_A}^2)$  and  $\rho_B = \sigma_0^2 / (\sigma_0^2 + \sigma_{\varepsilon_B}^2)$  for treatments A and B, respectively. The larger the person (between-subject) variance as compared to the error (within-subject) variance, the larger the intraclass correlations. Note that we assume a common between-subject variance, but allow for treatment-dependent within-subject variances, leading to treatment-dependent within-subject correlations. We also define a variance ratio  $\phi = (\sigma_0^2 + \sigma_{\varepsilon_A}^2) / (\sigma_0^2 + \sigma_{\varepsilon_B}^2) = \sigma_A^2 / \sigma_B^2$ , which can be expressed as a ratio of the intraclass correlations,  $\phi = \rho_B / \rho_A$ .

## 3. Optimal Allocations and Corresponding Design Efficiencies

Let  $\text{Var}(\hat{\beta}_x | \xi)$  denote the variance of the estimator of the treatment effect  $\beta_1$  in (1) or (2) or the variance of the

treatment by the period interaction effect  $\beta_3$  in (3), given a design  $\xi$ . The efficiency of an estimator of  $\beta_x$  is defined as the inverse of its variance, that is,  $(\text{Var}(\widehat{\beta}_x|\xi))^{-1}$ . In the sequel, we will consider the efficiency of one design,  $\xi_1$ , versus another design,  $\xi_2$ , which is defined as  $\text{Var}(\widehat{\beta}_x|\xi_2)/\text{Var}(\widehat{\beta}_x|\xi_1)$  and denoted as the relative efficiency. Since no closed-form expressions are available for the variances of the maximum likelihood (ML) estimator, asymptotic variances of the ML estimator were derived (Appendices A.1 and A.2).

The optimal allocation to treatments minimizes the variance of the estimator of  $\beta_{\text{treat}}$  in (1) or (2) and of  $\beta_{\text{treat} \times \text{time}}$  in (3), given a fixed research budget. Note that changing the coding of the treatment factor or the time factor in (1)–(3), for instance into 1 versus  $-1$  instead of 1 versus 0, will not affect the optimal allocation. Such a change of coding leads to a linear transformation of  $\beta_{\text{treat}}$  or  $\beta_{\text{treat} \times \text{time}}$ , and this will change the variance of their estimators only by a multiplicative constant. This implies that allocations that minimize the variance of the estimators do not depend on the coding of treatment and time.

To derive the optimal allocations under a budget restriction, we need to define a budget function. Let the costs involved with each subject in the parallel design be  $c_{\text{sp}}$  euros, in an extended parallel design be  $c_{\text{sep}}$  euros, and in a crossover design be  $c_{\text{sc}}$  euros. These costs may represent financial rewards given to subjects for participating in the trial but also the (average) costs of recruiting a subject. Furthermore, for treatments A and B there are, for each subject, costs  $c_A$  and  $c_B$ , respectively, and each measurement may involve  $c_t$  euros. Finally, attached to each treatment sequence, there may be administration costs  $c_{\text{ts}}$ .

In the case of allocation proportions  $p_A$  for treatment A and  $p_B = 1 - p_A$  for treatment B in a parallel design having  $n_p$  subjects, the following budget  $C^*$  is required:

$$C^* = 2c_{\text{ts}} + n_p(c_{\text{sp}} + c_t) + p_A n_p c_A + (1 - p_A) n_p c_B. \quad (4)$$

For the designs that we consider, this budget function can be reparametrized such that it is the same as the cost function given by Yuan and Zhou [8], thereby generalizing the cost function proposed by Brown [9] and Berger and Wong [4].

For an AB/BA crossover design, involving  $n_c$  subjects and allocation proportions  $p_{\text{AB}}$  for treatment sequence AB and  $p_{\text{BA}} = 1 - p_{\text{AB}}$  for treatment sequence BA, noting that each subject receives both treatment A and B and is measured twice, the following budget is required:

$$C^* = 2c_{\text{ts}} + n_c c_{\text{sc}} + n_c c_A + n_c c_B + 2n_c c_t. \quad (5)$$

Finally, the required budget for an AA/BB design, involving  $n_{\text{ep}}$  subjects, with allocation proportions  $p_{\text{AA}}$  and  $p_{\text{BB}} = 1 - p_{\text{AA}}$  for the treatment sequences AA and BB, respectively, is as follows:

$$C^* = 2c_{\text{ts}} + n_{\text{ep}} c_{\text{sep}} + 2n_{\text{ep}} p_{\text{AA}} c_A + 2n_{\text{ep}} (1 - p_{\text{AA}}) c_B + 2n_{\text{ep}} c_t. \quad (6)$$

Note that, for the functions in (4)–(6), the budget may simply be the total number of observations involved in

a study, by setting  $c_t = 1$  and the other costs to 0. It can also represent the total number of subjects involved, by setting  $c_{\text{sp}} = c_{\text{sep}} = c_{\text{sc}} = 1$  and the remaining costs to 0.

In what follows, we will assume that the subject-specific costs of the two-period designs are the same; that is,  $c_{\text{sep}} = c_{\text{sc}} = c_{\text{s}_{2p}}$ . Since subjects in these designs receive two treatments and a washout period may be involved, these costs are very likely larger than those of a parallel design. We also assume that the subject-specific costs for the two-period designs will not exceed 2 times the subject-specific costs for the parallel design, so that  $c_{\text{sp}} \leq c_{\text{s}_{2p}} \leq 2c_{\text{sp}}$ . Finally, since each design involves two treatment sequences, the administration costs are the same for each of the three designs considered, and thus, the budgets that are available for remaining costs are identical; that is, the budget  $C = C^* - 2c_{\text{st}}$  is the same for each design.

**3.1. Treatment Effect.** For treatment effect estimation, the optimal allocations to the treatment sequences are derived in Appendix B. The optimal allocations and corresponding (asymptotic) variances of the treatment effect estimators are shown in the second and third column of Table 1, respectively. The optimal allocation ratios of the parallel and the extended parallel design depend on the costs and intraclass correlations: the more the expensive treatment A (or the cheaper treatment B) and the larger the intraclass correlation in treatment A (or the smaller the intraclass correlation in treatment B), the more the subjects have to be assigned to treatment B. The optimal allocation ratio for a crossover design is 1, which may be expected, since both groups receive both treatment A and B.

**3.2. Treatment by Period Interaction Effect.** In the case the treatment by period interaction effect is of primary interest, the optimal allocations can be derived along lines similar to the derivations for the treatment effect (Appendix B). The allocations and corresponding optimal variances are displayed in Table 1. Note that, similar to treatment effect estimation, the allocation ratio for a crossover design is 1, whereas the allocation ratio for an extended parallel design depends on the treatment costs and intraclass correlations, such that more persons are allocated to treatment sequence AA if the intraclass correlation of A decreases, the intraclass correlation of B increases, the costs of treatment A decrease, or the costs of treatment B increase.

## 4. Maximin Designs

Choosing the optimal allocation requires knowledge on the intraclass correlations  $\rho_A$  and  $\rho_B$  (remember that the variance ratio  $\phi$  is fixed if  $\rho_A$  and  $\rho_B$  are given). Commonly, there is only limited knowledge on these parameters. A possible solution is the maximin strategy [4], consisting of 2 steps: (1) for each design determine the minimum efficiency of the effect estimator across the plausible ranges for the intraclass correlations  $\rho_A$  and  $\rho_B$  and (2) choose that design which maximizes this minimum efficiency. Such a design optimizes a worst case scenario and is called a *maximin design*. The maximin strategy implies choosing the design

TABLE 1: Optimal allocation ratios and corresponding variances of the ML estimator of the treatment effect and the treatment by period interaction for different designs under heterogeneity of outcome variances and costs.

Treatment effect		
Design	Allocation ratio ( $n_1/n_2$ )	$\text{Var}(\hat{\beta}_{\text{treat}})$
Crossover design	1	$(c_A + c_B + 2c_t + c_{s_{2p}})(1 - ((2\rho_A\rho_B)/(\rho_A + \rho_B)))\sigma_y^2/C$
Parallel design	$\sqrt{(\rho_B/\rho_A)((c_B + c_t + c_{sp})/(c_A + c_t + c_{sp}))}$	$\frac{(\sqrt{(c_A + c_t + c_{sp})(\rho_B/(\rho_A + \rho_B))} + \sqrt{(c_B + c_t + c_{sp})(\rho_A/(\rho_A + \rho_B))})^2(\sigma_y^2/C)$
Extended parallel design	$\sqrt{(\rho_B(1 + \rho_A)/(\rho_A(1 + \rho_B))) \times ((2(c_B + c_t) + c_{s_{2p}})/(2(c_A + c_t) + c_{s_{2p}}))}$	$\frac{(\sqrt{2(c_A + c_t) + c_{s_{2p}}})(\rho_B(1 + \rho_A)/(\rho_A + \rho_B)) + \sqrt{2(c_B + c_t) + c_{s_{2p}}})(\rho_A(1 + \rho_B)/(\rho_A + \rho_B))^2(\sigma_y^2/C)$
Treatment by period interaction		
Design	Allocation ratio ( $n_1/n_2$ )	$\text{Var}(\hat{\beta}_{\text{treat} \times \text{time}})$
Crossover design	1	$4(c_A + c_B + 2c_t + c_{s_{2p}})(1 + 2\rho_A\rho_B/(\rho_A + \rho_B))(\sigma_y^2/C)$
Extended parallel design	$\sqrt{(\rho_B(1 - \rho_A)/(\rho_A(1 - \rho_B))) \times (2(c_B + c_t) + c_{s_{2p}})/(2(c_A + c_t) + c_{s_{2p}})}$	$\frac{2(\sqrt{2(c_A + c_t) + c_{s_{2p}}})(\rho_B(1 - \rho_A)/(\rho_A + \rho_B)) + \sqrt{2(c_B + c_t) + c_{s_{2p}}})(\rho_A(1 - \rho_B)/(\rho_A + \rho_B))^2(\sigma_y^2/C)$

Note.  $n_1$ : sample size for A (parallel design), AB (crossover design), or AA (extended parallel design) sequence;  $n_2$ : sample size for B (parallel design), BA (crossover design), or BB (extended parallel design) sequence;  $\sigma_y^2 = \sigma_A^2 + \sigma_B^2$ .

that minimizes the maximum variance of the estimator of the effect of interest. In determining sample sizes, choosing values for the intraclass correlations  $\rho_A$  and  $\rho_B$  within their plausible ranges (and thus a variance ratio  $\phi$  within its plausible range) for which the variance is maximum will guarantee the desired power level also for all other values of these parameters. Moreover, the maximin design guarantees this power level at the lowest research costs. In what follows, we will refer to ranges of  $\rho_A$  and  $\rho_B$  that have lower bounds  $\rho_A^L$  and  $\rho_B^L$  and upper bounds  $\rho_A^U$  and  $\rho_B^U$ , respectively.

**4.1. Treatment Effect.** From the asymptotic variances in Table 1, one can derive for which values of  $\rho_A$  and  $\rho_B$  (and thus for which value of the variance ratio  $\phi$ ), the variance of the treatment effect estimator is maximized. These derivations are given in Appendix C. The maximin parameter values and corresponding variances for the treatment effect estimator under optimal allocation to the treatments are shown in Table 2. The corresponding optimal allocations for the maximin designs are obtained by substituting the maximin parameter values of Table 2 into the allocation ratios as given in Table 1.

If for a parallel design the maximin value for the variance ratio  $\phi^* = (c_A + c_t + c_{sp})/(c_B + c_t + c_{sp})$  is within the plausible range for  $\phi$ , that is,  $[\rho_B^L/\rho_A^U, \rho_B^U/\rho_A^L]$  and  $c_{s_{2p}} \leq 2c_{sp}$ , then a parallel design is always less efficient than a maximin crossover design. If for an extended parallel design the maximin value for one of the intraclass correlations is within the plausible range for the corresponding intraclass correlation, then also this design is less efficient than a maximin crossover design. For other scenarios, the relations between the maximin designs are more complicated, depending on the ranges for  $\rho_A$  and  $\rho_B$ , the costs of treatments, subject recruitment, and measurement.

A systematic numerical evaluation was done to examine under what conditions the crossover design is the best choice in terms of efficiency. For  $\rho_A$  and  $\rho_B$ , we consider ranges of

width 0.10 (small), 0.30 (medium), and 0.60 (large). The lower bounds were  $\{0.01, 0.05, 0.10, 0.15, 0.20, \dots\}$ , where the largest possible lower bound was determined by the width of the range under consideration. For instance, if the range is 0.30 (medium), the largest lower bound for the intraclass correlation is 0.70. All combinations of small, medium, and large ranges for  $\rho_A$  and  $\rho_B$  were considered. The values of the variance ratio  $\phi$  thus considered vary from 1/100 to 100. Since in most crossover trials, the intraclass correlation exceeds 0.30 [1–3, 10, 11], ranges with lower bounds of 0.30 or higher are empirically most relevant. The empirical evidence on the costs  $c_A$ ,  $c_B$ ,  $c_t$ ,  $c_{sp}$ , and  $c_{s_{2p}}$  is scarce, and we thus choose costs covering a wide range of scenarios. Let  $CR_A = (c_A + c_t)/c_{sp}$ ,  $CR_B = (c_B + c_t)/c_{sp}$ , and  $CR_p = c_{s_{2p}}/c_{sp}$  (note that the relative efficiencies of the maximin designs depend only on these cost ratios).  $CR_A$  and  $CR_B$  take on the values 100, 20, 10, 1, 0.1, 0.05, and 0.01. For  $CR_p$ , we consider 1 and 2.

If the costs of treatments are identical between the treatment arms, that is,  $CR_A = CR_B$ , for most scenarios examined, the crossover maximin design turns out to be most efficient. For  $CR_p = 1$  and  $CR_A = CR_B \leq 1$ , the crossover design always is the most efficient. For  $CR_p = 1$  and  $CR_A = CR_B > 1$ , or  $CR_p = 2$ , only if the lower bound of one of the intraclass correlations is 0.05 or lower and the ranges of the intraclass correlations do not overlap, the parallel design can become most efficient. Since in most empirical studies the intraclass correlations will exceed 0.05, this implies that, for equal costs of treatments, the crossover maximin design will almost always be the most efficient design.

In the case the treatment costs differ and  $CR_A \leq 1$  and  $CR_B \leq 1$ , only in the case the lower bound of one of the intraclass correlations is 0.10 or smaller, the parallel or the extended parallel maximin design can become most efficient. The extended parallel design can only become most efficient if  $CR_p = 1$ . Hence, in all scenarios with unequal treatment costs and  $CR_A \leq 1$  and  $CR_B \leq 1$ , for intraclass correlations of 0.15 or higher, the maximin crossover design is most efficient.



TABLE 2: Values for the parameters and variances of the treatment effect estimator for each of the maximin designs.

Design	Maximin parameter values	Var( $\hat{\beta}_{\text{treat}}$ ) for the maximin design
Crossover design	$\rho_A^L, \rho_B^L$	$(c_A + c_B + 2c_t + c_{s_{2p}})(1 - 2\rho_A^L\rho_B^L/(\rho_A^L + \rho_B^L))(\sigma_y^2/C)$
Parallel design	If $(\rho_B^L/\rho_A^L) \leq (c_A + c_t + c_{sp})/(c_B + c_t + c_{sp}) \leq (\rho_B^U/\rho_A^L)$ , choose $(\rho_B/\rho_A) = (c_A + c_t + c_{sp})/(c_B + c_t + c_{sp})$	$(c_A + c_B + 2(c_t + c_{sp}))(\sigma_y^2/C)$
	If $(c_A + c_t + c_{sp})/(c_B + c_t + c_{sp}) > (\rho_B^U/\rho_A^L)$ , then $\rho_A^L$ and $\rho_B^U$	$(\sqrt{(c_A + c_t + c_{sp})(\rho_B^U/(\rho_A^L + \rho_B^U))})$ $+ \sqrt{(c_B + c_t + c_{sp})(\rho_A^L/(\rho_A^L + \rho_B^U))}(\sigma_y^2/C)$
Extended parallel design	If $(c_A + c_t + c_{sp})/(c_B + c_t + c_{sp}) < (\rho_B^L/\rho_A^U)$ , then $\rho_A^U$ and $\rho_B^L$	$(\sqrt{(c_A + c_t + c_{sp})(\rho_B^L/(\rho_A^U + \rho_B^L))})$ $+ \sqrt{(c_B + c_t + c_{sp})(\rho_A^U/(\rho_A^U + \rho_B^L))}(\sigma_y^2/C)$
	If $(1/\rho_A^U) + 1 \leq ((\lambda(1 - \rho_B^U)^2)/((1 + \rho_B^U)\rho_B^U)) \leq (1/\rho_A^L) + 1$ , then $\rho_A^* = (\rho_B^U(1 + \rho_B^U))/(\lambda(1 - \rho_B^U)^2 - \rho_B^U(1 + \rho_B^U)), \rho_B^U$ else if $(1/\rho_B^U) + 1 \leq (((1 - \rho_A^U)^2)/(\lambda(1 + \rho_A^U)\rho_A^U)) \leq (1/\rho_B^L) + 1$ , then $\rho_A^U, \rho_B^* = (\lambda\rho_A^U(1 + \rho_A^U))/((1 - \rho_A^U)^2 - \lambda\rho_A^U(1 + \rho_A^U))$ , else if $(\lambda(1 - \rho_B^U)^2)/((1 + \rho_B^U)\rho_B^U) < (1/\rho_A^U) + 1$ , and $(1 - \rho_A^U)^2/(\lambda(1 + \rho_A^U)\rho_A^U) < (1/\rho_B^U) + 1$ , then $\rho_A^U, \rho_B^U$	$(c_A + c_B((1 + \rho_B^U)/(1 - \rho_B^U))$ $+ ((2c_t + c_{s_{2p}})/(1 - \rho_B^U)))(\sigma_y^2/C)$ $(c_A((1 + \rho_A^U)/(1 - \rho_A^U)) + c_B + ((2c_t +$ $c_{s_{2p}})/(1 - \rho_A^U)))(\sigma_y^2/C)$ $(\sqrt{(2(c_A + c_t) + c_{s_{2p}})(\rho_B^U(1 + \rho_B^U)/(\rho_A^U + \rho_B^U))})$ $+ \sqrt{(2(c_B + c_t) + c_{s_{2p}})(\rho_A^U(1 + \rho_A^U)/(\rho_A^U + \rho_B^U))}(\sigma_y^2/(2C))$
	else if $((\lambda(1 - \rho_B^U)^2)/((1 + \rho_B^U)\rho_B^U)) < (1/\rho_A^U) + 1$ , then $\rho_A^U, \rho_B^L$	$(\sqrt{(2(c_A + c_t) + c_{s_{2p}})(\rho_B^L(1 + \rho_A^U)/(\rho_A^U + \rho_B^L))})$ $+ \sqrt{(2(c_B + c_t) + c_{s_{2p}})(\rho_A^U(1 + \rho_B^L)/(\rho_A^U + \rho_B^L))}(\sigma_y^2/(2C))$
	else $\rho_A^L, \rho_B^U$	$(\sqrt{(2(c_A + c_t) + c_{s_{2p}})(\rho_B^U(1 + \rho_A^L)/(\rho_A^L + \rho_B^U))})$ $+ \sqrt{(2(c_B + c_t) + c_{s_{2p}})(\rho_A^L(1 + \rho_B^U)/(\rho_A^L + \rho_B^U))}(\sigma_y^2/(2C))$

Note.  $\sigma_y^2 = \sigma_A^2 + \sigma_B^2; \lambda = (2c_A + 2c_t + c_{s_{2p}})/(2c_B + 2c_t + c_{s_{2p}})$ .

In the case the treatment costs differ and  $CR_A > 1$  or  $CR_B > 1$ , the maximin crossover design is less often most efficient. For these cost scenarios, also for ranges of intraclass correlations exceeding 0.15, the maximin parallel and extended parallel design may become more efficient. This especially occurs if the costs of treatment A and lower bound of the range of  $\rho_A$  are both larger (or smaller) than the costs of treatment B and lower bound of  $\rho_B$ , respectively. The efficiency improvement is large if treatment A is much more expensive than treatment B and if the costs of treatments and measurements are large compared to the subject-related costs. This is illustrated in Figure 1. The top row shows that if the costs of treatment A are larger than the costs of treatment B and the lower bound of  $\rho_A$  is larger than the lower bound of  $\rho_B$ , a parallel design is most efficient, even up to an upper bound 1 of  $\rho_B$  if  $CR_A = 100$ . As can also be seen, the upper bound of  $\rho_A$  is not very relevant in terms of the relative efficiencies. The left plot of the middle row of Figure 1 shows that if the lower bounds of  $\rho_A$  and  $\rho_B$  are equal, then for almost all upper bounds of  $\rho_B$ , the crossover design is most efficient. Again, as can be seen in the rightmost plot of the middle row, if the lower bound of  $\rho_A$  is higher than the lower bound of  $\rho_B$ , then for higher upper bounds of  $\rho_B$ , the parallel design is most efficient but to a lesser extent as compared to a smaller lower bound of  $\rho_B$ . As is evident from the four subplots in the top and middle row, when increasing the ratio  $CR_A/CR_B$ , the crossover design becomes less efficient as compared to the other two designs. The subplots of the bottom row furthermore show that the crossover design also becomes less

efficient compared to the other designs if  $CR_A$  and  $CR_B$  increase while the ratio  $CR_A/CR_B$  remains constant. This illustrates that the efficiency of the other designs relative to the crossover design becomes larger if the costs of treatments and measurements are large compared to the subject-related costs. However, to summarize, if the treatment costs differ and  $CR_A > 1$  or  $CR_B > 1$ , no simple rules of the thumb emerge and the most solid way to choose the most efficient design is just to calculate the maximin variances as given in Table 2.

Finally, if  $CR_p = 2$ , the maximin parallel design is consistently more efficient than the maximin extended parallel design (as is illustrated in Figure 1). If  $CR_p = 1$ , the maximin extended parallel design can also become more efficient than the maximin parallel design.

**4.2. Treatment by Period Interaction Effect.** The maximin parameter values and corresponding variances of the estimator of the treatment by period interaction effect are shown in Table 3. The derivations of these results can be done along lines similar to the derivations for the treatment effect estimator (Appendix C). The optimal allocation for the extended parallel design is obtained by substituting the maximin parameter values in the expression for the allocation ratio in Table 1. For a crossover design, the allocation ratio is 1. The maximin efficiency of an extended parallel design is always higher than that of a crossover design if the maximin value  $\rho_A^*$  is within the plausible range for  $\rho_A$ . This follows from

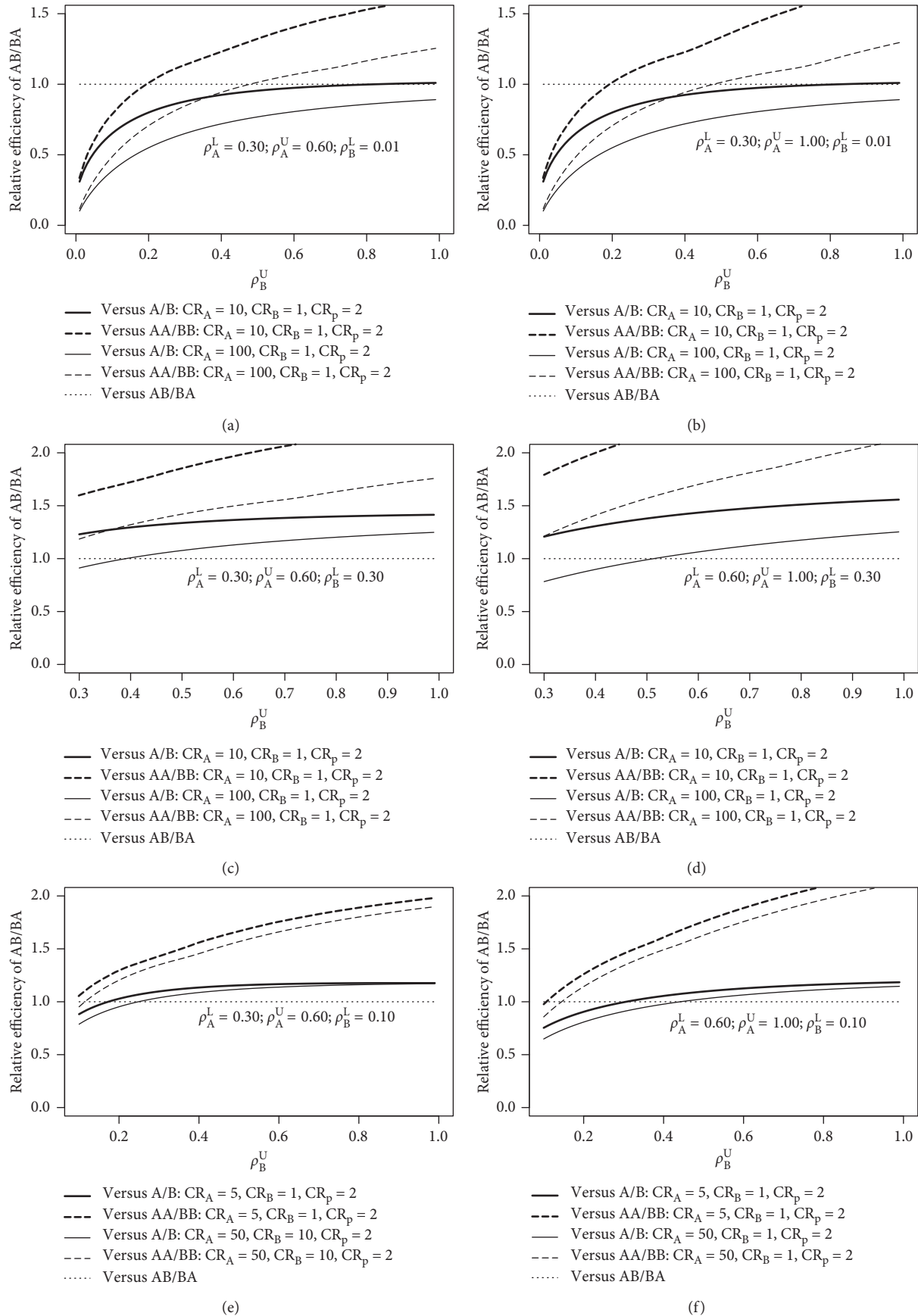


FIGURE 1: Relative efficiency for the treatment estimator of the AB/BA design versus the A/B and AA/BB designs as a function of the ranges for the intra-class correlations in treatment A (lower bound =  $\rho_A^L$ , upper bound =  $\rho_A^U$ ) and B (lower bound =  $\rho_B^L$ , upper bound =  $\rho_B^U$ ), in case of minimizing the research costs.

TABLE 3: Values for the parameters and variances of the treatment by period interaction effect estimator for each of the maximin designs.

Design	Maximin parameter values	Var( $\widehat{\beta}_{\text{treat} \times \text{time}}$ )
Crossover design	$\rho_A^U, \rho_B^U$	$(4(c_A + c_B + 2c_t + c_{s_{2p}})(1 + ((2\rho_A^U \rho_B^U)/(\rho_A^U + \rho_B^U))))(\sigma_y^2/C)$
Extended parallel design	If $(1/\rho_A^U) - 1 \leq ((\lambda(1 + \rho_B^L)^2)/((1 - \rho_B^L)\rho_B^L)) \leq (1/\rho_A^L) - 1$ , then $\rho_A^* = (\rho_B^L(1 - \rho_B^L))/(\lambda(1 + \rho_B^L)^2 + \rho_B^L(1 - \rho_B^L)), \rho_B^L$ ,	$((4(c_A(1 + \rho_B^L) + c_B(1 - \rho_B^L) + 2c_t + c_{s_{2p}}))/(1 + \rho_B^L))(\sigma_y^2/C)$
	else if $(1/\rho_B^U) - 1 \leq (((1 + \rho_A^L)^2)/(\lambda(1 - \rho_A^L)\rho_A^L)) \leq (1/\rho_B^L) - 1$ , then $\rho_A^L, \rho_B^* = (\lambda\rho_A^L(1 - \rho_A^L))/((1 + \rho_A^L)^2 + \lambda\rho_A^L(1 - \rho_A^L))$ ,	$((4(c_A(1 - \rho_A^L) + c_B(1 + \rho_A^L) + 2c_t + c_{s_{2p}}))/(1 + \rho_A^L))(\sigma_y^2/C)$
	else if $(\lambda(1 + \rho_B^L)^2)/((1 - \rho_B^L)\rho_B^L) > (1/\rho_A^L) - 1$ and $(1 + \rho_A^L)^2/(\lambda(1 - \rho_A^L)\rho_A^L) > (1/\rho_B^L) - 1$ , then $\rho_A^L, \rho_B^L$	$2(\sqrt{((2(c_A + c_t) + c_{s_{2p}})((\rho_B^L(1 - \rho_A^L))/(\rho_A^L + \rho_B^L)))}) + \sqrt{((2(c_B + c_t) + c_{s_{2p}})(\rho_A^L(1 - \rho_B^L)/(\rho_A^L + \rho_B^L)))})^2(\sigma_y^2/C)$
	else if $((\lambda(1 + \rho_B^L)^2)/((1 - \rho_B^L)\rho_B^L)) < (1/\rho_A^U) - 1$ , then $\rho_A^U, \rho_B^L$	$2(\sqrt{((2(c_A + c_t) + c_{s_{2p}})((\rho_B^L(1 - \rho_A^U))/(\rho_A^U + \rho_B^L)))}) + \sqrt{((2(c_B + c_t) + c_{s_{2p}})(\rho_A^U(1 - \rho_B^L)/(\rho_A^U + \rho_B^L)))})^2(\sigma_y^2/C)$
	else $\rho_A^L, \rho_B^U$	$2(\sqrt{((2(c_A + c_t) + c_{s_{2p}})((\rho_B^U(1 - \rho_A^L))/(\rho_A^L + \rho_B^U)))}) + \sqrt{((2(c_B + c_t) + c_{s_{2p}})(\rho_A^L(1 - \rho_B^U)/(\rho_A^L + \rho_B^U)))})^2(\sigma_y^2/C)$

Note.  $\sigma_y^2 = \sigma_A^2 + \sigma_B^2; \lambda = (2c_A + 2c_t + c_{s_{2p}})/(2c_B + 2c_t + c_{s_{2p}})$ .

$$\frac{4(c_A(1 + \rho_B^L) + c_B(1 - \rho_B^L) + 2c_t + c_{s_{2p}})}{(1 + \rho_B^L)} \left( \frac{\sigma_y^2}{C} \right) \leq 4(c_A + c_B + 2c_t + c_{s_{2p}}) \left( \frac{\sigma_y^2}{C} \right), \quad (7)$$

where the right-hand side of the inequality in turn is smaller than the variance of a maximin crossover design (Table 3). The higher maximin efficiency of the extended parallel design can also be shown to hold if the maximin value  $\rho_B^*$  is within the plausible range for  $\rho_B$ . Furthermore, if the variance maximizing values  $\rho_A^*$  and  $\rho_B^*$  are outside the plausible ranges for  $\rho_A$  and  $\rho_B$ , respectively, then values for  $\rho_A$  and  $\rho_B$  that coincide with one of the borders of their corresponding ranges should be chosen as values that maximize the variance. But in that case, even smaller variances result for the extended parallel design.

**4.3. Maximin Designs That Minimize the Number of Subjects and Number of Measurements.** As noted in Section 3, by setting  $c_{sp} = c_{sep} = c_{sc} = 1$  and the remaining costs to 0 in (4)–(6), the budget is simply the total number of subjects involved in a study, and by setting  $c_t = 1$  and the other costs to 0, the budget reduces to the total number of measurements involved. When the budget is the total sample size and interest is in estimating the treatment effect, it can be proven, based on the formulas in Table 2, that a maximin crossover design requires less subjects than a maximin parallel design. From an extensive numerical evaluation analogous to the one of Section 4.1, a maximin crossover design also appears to require less subjects than a maximin extended parallel design.

When minimizing the number of measurements, the numerical evaluation shows again that the maximin crossover design is the best choice provided the lower bounds of both intraclass correlations are 0.10 or higher. In other cases

also a maximin parallel design may minimize the total number of measurements. Since in most crossover trials the intraclass correlation exceeds 0.30 [1–3, 10, 11], in practice, this implies that the maximin crossover trial also is the best choice when minimizing the number of measurements.

In the case, interest is in the treatment by period interaction, Section 4.2 showed a maximin extended parallel design to be more efficient and thus also to require less budget than a maximin crossover trial. In the special case where the number of subjects or the total number of measurements are minimized, the maximin extended parallel design will therefore also outperform the maximin crossover design.

## 5. Monte Carlo Evaluation of the Power of Maximin Designs

The efficiencies as derived for the maximin designs are based on the asymptotic variance of the ML estimator,  $\text{Var}(\widehat{\beta}_x | \xi)$ . For sufficiently large numbers of subjects, the relation between the asymptotic variance of the ML estimator and the power level  $1 - \gamma$  to detect a treatment effect in a two-tailed test with type I error rate  $\alpha$  can be approximated as follows:

$$\text{Var}(\widehat{\beta}_x | \xi) = \left( \frac{\beta_x}{z_{1-\alpha/2} + z_{1-\gamma}} \right)^2, \quad (8)$$

where  $z_{1-\alpha/2}$  and  $z_{1-\gamma}$  are the 100  $(1 - \alpha/2)$  and 100  $(1 - \gamma)$  percentiles of the standard normal distribution. For small sample sizes calculated by (8), corrections are needed [12, 13]. For each of the three designs, these corrections will be applied. We will examine to what extent the differences between designs in asymptotic efficiencies translate into corresponding differences in power levels for small samples. Also, when planning sample sizes based on the asymptotic

variances, we can check whether the commonly used power levels of 80% or 90% are realized in the case of small samples.

For the treatment effect estimator, the following expression for the required number of subjects results for a crossover design with optimal allocation:

$$n_c = (z_{1-\alpha/2} + z_{1-\gamma})^2 \left( 1 - \frac{2\rho_A\rho_B}{\rho_A + \rho_B} \right) \left( \frac{\sigma_A^2 + \sigma_B^2}{\beta_{\text{treat}}^2} \right). \quad (9)$$

If we let  $ES = \beta_{\text{treat}} / \sqrt{0.5(\sigma_A^2 + \sigma_B^2)}$  be the effect size based on the outcome variances in the treatment and control arm (cf. [14]), then (9) can be rewritten as follows:

$$n_c = 2 \left( \frac{z_{1-\alpha/2} + z_{1-\gamma}}{ES} \right)^2 \left( 1 - \frac{2\rho_A\rho_B}{\rho_A + \rho_B} \right). \quad (10)$$

Note that, in the case of a maximin design, the expression is the same as (10), however, with  $\rho_A^L$  and  $\rho_B^L$  being substituted for  $\rho_A$  and  $\rho_B$ , respectively. Similar rewritings of the sample sizes in terms of the effect size are possible for the parallel and extended parallel design, respectively:

$$n_p = 2 \left( \frac{z_{1-\alpha/2} + z_{1-\gamma}}{ES} \right)^2 \cdot \left( 1 + \left( \frac{c_A + c_B + 2c_t + 2c_{sp}}{\sqrt{(c_A + c_t + c_{sp})(c_B + c_t + c_{sp})}} \right) \cdot \left( \sqrt{\frac{\rho_B}{\rho_A + \rho_B}} \sqrt{\frac{\rho_A}{\rho_A + \rho_B}} \right) \right), \quad (11)$$

$$n_{ep} = \left( \frac{z_{1-\alpha/2} + z_{1-\gamma}}{ES} \right)^2 \left( 1 + \frac{2\rho_A\rho_B}{\rho_A + \rho_B} + \left( \frac{2c_A + 2c_B + 4c_t + 2c_{s_{2p}}}{\sqrt{(2c_A + 2c_t + c_{s_{2p}})(2c_B + 2c_t + c_{s_{2p}})}} \right) \cdot \left( \sqrt{\frac{\rho_B(1 + \rho_A)}{\rho_A + \rho_B}} \sqrt{\frac{\rho_A(1 + \rho_B)}{\rho_A + \rho_B}} \right) \right).$$

The choices to be made for  $\rho_A$  and  $\rho_B$  in the case of maximin versions of the parallel and extended parallel design are determined by the conditions as formulated in Table 2.

In the case of the treatment by period interaction effect, the following expression for the required number of subjects can be derived for a crossover design with optimal allocation:

$$n_c = 8 \left( \frac{z_{1-\alpha/2} + z_{1-\gamma}}{ES} \right)^2 \left( 1 + \frac{2\rho_A\rho_B}{\rho_A + \rho_B} \right), \quad (12)$$

where  $ES = \beta_{\text{treat} \times \text{time}} / \sqrt{0.5(\sigma_A^2 + \sigma_B^2)}$ . In the case of a maximin design, the expression is the same as (12), however, with  $\rho_A^U$  and  $\rho_B^U$  being substituted for  $\rho_A$  and  $\rho_B$ , respectively. The expression for the sample size of an extended parallel design, when allocating optimally, can be written as follows:

$$n_{ep} = 4 \left( \frac{z_{1-\alpha/2} + z_{1-\gamma}}{ES} \right)^2 \left( 1 - \frac{2\rho_A\rho_B}{\rho_A + \rho_B} + \left( \frac{2c_A + 2c_B + 4c_t + 2c_{s_{2p}}}{\sqrt{(2c_A + 2c_t + c_{s_{2p}})(2c_B + 2c_t + c_{s_{2p}})}} \right) \cdot \left( \sqrt{\frac{\rho_B(1 - \rho_A)}{\rho_A + \rho_B}} \sqrt{\frac{\rho_A(1 - \rho_B)}{\rho_A + \rho_B}} \right) \right). \quad (13)$$

The choices to be made for  $\rho_A$  and  $\rho_B$  in the case of a maximin extended parallel design are determined by the conditions formulated in Table 3.

Since maximin designs only require information on plausible ranges of model parameters, they are more practical than optimal designs. In what follows, we will therefore examine through a Monte Carlo simulation the power for maximin designs in the case of small sample sizes. First, we will discuss the factors that are varied and motivate the choices made for these factors in determining the simulation scenarios.

### 5.1. Choice of Ranges/Values for Relevant Factors

**5.1.1. Effect Sizes.** The effect size, ES, is commonly categorized into small (0.2), medium (0.5), and large (0.8) [14]. Being primarily interested in the small sample performance of (9)–(13), we only will consider  $ES = 0.8$ , leading to the smallest sample sizes.

**5.1.2. Costs.** The empirical evidence on costs is rather scarce, but we will choose the costs such that they imply minimizing the sample size of a study (i.e.,  $c_A = c_B = c_t = 0$  and  $c_{sp} = c_{s_{2p}} = 1$ ).

**5.1.3. Intraclass Correlations.** The ranges for  $\rho_A$  and  $\rho_B$  are identical to the ranges of the numerical evaluation of Section 4.1. Since we are interested in the small sample performance, for each design, we consider that pair of ranges across all combinations of ranges for the intraclass correlations (i.e., small-small, medium-medium, large-large, small-medium, small-large, and medium-large) that lead to the smallest sample sizes. Since this each time turns out to be a pair from the small-small category, the same was done for all pairs of medium and large ranges, which will be used more often in practice. For each design, the two resulting pairs of ranges of intraclass correlation are displayed in the two leftmost columns of Table 4 and the Table in Appendix D.

**5.1.4. Power Level and Type I Error Rate.** In sample size planning commonly used power levels are 80% and 90% in a two-tailed test with either a 5% or a 1% type I error rate. Focusing on the small sample performance, we will consider 80% power in a two-tailed test with a 5% type I error rate. For small sample sizes derived from the standard normal



TABLE 4: Powers from the Monte Carlo simulations for maximin designs in the case of the treatment effect. For each pair of ranges of the intraclass correlations, the asymptotic efficiency of each design versus the most efficient design is given within brackets.

		Treatment effect			
		N	Crossover design	Parallel design	Extended parallel design
<i>Type I error rate = 0.05</i>					
$[\rho_A^L, \rho_A^U]$	$[\rho_B^L, \rho_B^U]$				
[0.01,0.10]	[0.90,1.00]	30	0.857 (1)	0.648 (0.61)	<b>0.836</b> (0.96)
[0.01,0.10]	[0.90,1.00]	44	0.959 (1)	<b>0.822</b> (0.61)	0.950 (0.96)
[0.01,0.30]	[0.01,0.30]	36	0.916 (1)	0.805 (0.50)	<b>0.823</b> (0.76)
[0.01,0.30]	[0.70,1.00]	52	0.982 (1)	<b>0.819</b> (0.51)	0.922 (0.70)
[0.70,1.00]	[0.70,1.00]	10	<b>0.813</b> (1)	0.180 (0.15)	0.182 (0.15)
[0.90,1.00]	[0.90,1.00]	6	<b>0.902</b> (1)	0.088 (0.05)	0.087 (0.05)
<i>Type I error rate = 0.01</i>					
$[\rho_A^L, \rho_A^U]$	$[\rho_B^L, \rho_B^U]$				
[0.01,0.10]	[0.90,1.00]	45	0.854 (1)	0.618 (0.61)	<b>0.848</b> (0.96)
[0.01,0.10]	[0.90,1.00]	66	0.974 (1)	<b>0.829</b> (0.61)	0.967 (0.96)
[0.01,0.30]	[0.01,0.30]	55	0.931 (1)	0.817 (0.50)	<b>0.838</b> (0.76)
[0.01,0.30]	[0.70,1.00]	77	0.989 (1)	<b>0.822</b> (0.51)	0.931 (0.70)
[0.70,1.00]	[0.70,1.00]	16	<b>0.856</b> (1)	0.112 (0.15)	0.119 (0.15)
[0.90,1.00]	[0.90,1.00]	8	<b>0.849</b> (1)	0.029 (0.05)	0.030 (0.05)

Note. The power printed in bold indicates the design for which the sample calculation should yield a power of 80%. N: total sample size.

distribution (as in (9)–(13)), corrections are needed that turn out to depend on the type I error rate [12, 13]. For this reason, we will also study a 1% type I error rate.

5.2. *Simulation Procedure and Testing Methods.* For each of the 20 simulation scenarios, 25,000 data sets were generated. To distinguish chance deviations of the simulated power from systematic deviations, a 95% predictive interval was calculated. For a nominal power  $\pi$ , this is defined as  $[\pi + z_{0.025} \sqrt{\pi(1-\pi)/N\text{sim}}, \pi + z_{0.975} \sqrt{\pi(1-\pi)/N\text{sim}}]$ , where  $z_{0.025}$  and  $z_{0.975}$  are the 2.5 and 97.5 percentiles of the standard normal distribution and  $N\text{sim}$  is the number of simulations. Since the nominal power  $\pi = 0.80$  and  $N\text{sim} = 25,000$ , the 95% predictive interval is [0.795, 0.805].

For a test of the treatment effect, the data generated for the crossover design were analyzed with a two-sample  $t$ -test on the difference scores obtained by subtracting the two measurements for each subject. The model in (2) implies homogeneity of variances for these difference scores, so that a pooled variance  $t$ -test was applied. The data generated for the parallel design were simply analyzed by a two-sample  $t$ -test on the original scores, whereas the data for the extended parallel design were analyzed with a two-sample  $t$ -test on the scores averaged across both measurements. For these parallel designs, (1) and (2) imply that the analyzed scores may have variances differing between groups, so that an unpooled variance  $t$ -test was applied. For the treatment by period interaction effect, the data generated for the crossover design were analyzed with a two-sample (pooled variance)  $t$ -test on the scores averaged for each subject across both measurements, whereas the data generated for the extended parallel design were analyzed with a two-sample (unpooled variance)  $t$ -test on the differences between the two measurements (3). These different  $t$ -tests follow for each of the designs (involving equal numbers of measurements per subject) from the analysis models in (1)–(3) and do not require asymptotic assumptions.

In calculating the required sample size, (9)–(11) were used, when interest is in testing the treatment effect, and (12) and (13) were used, when interest is in testing the treatment by period interaction. The optimal allocations for each design are given in Table 1, taking the maximin values for  $\rho_A$  and  $\rho_B$  as determined from Tables 2 and 3 for the treatment main and treatment by period interaction effect, respectively. Since the outcome variances are unknown, the calculated sample sizes were subsequently corrected. For an unpooled variance  $t$ -test, if the number of persons per arm is 8 or more, sufficient corrections for two-tailed tests are 2 extra persons per arm if  $\alpha = 0.05$  and 4 extra persons per arm if  $\alpha = 0.01$ . With less than 8 persons in one or both arms, sufficient corrections are 3 extra persons per arm if  $\alpha = 0.05$  and 4 extra persons per arm if  $\alpha = 0.01$  ([13], p. 568). For a pooled variance  $t$ -test, we only need to add 1 person per arm if  $\alpha = 0.05$  (two-tailed) and 2 persons per arm if  $\alpha = 0.01$  (two-tailed) ([12], p. 1216-1217). These are sufficient corrections for planned powers of 80% and 90%. The simulations, statistical tests, and power calculations were done in R, version 3.1.3 [15].

5.3. *Results.* As can be seen in Table 4 and Appendix D, for all sample size-design combinations that should yield 80% power, the simulated powers were either within or above the 95% predictive intervals. This indicates that the asymptotic results, supplemented with simple correction rules for using the standard normal distribution, yield sample sizes that guarantee the desired level of power. The realized power levels generally are higher than 80%, since the small sample corrections are sufficient and in some cases smaller corrections would have been appropriate [12, 13].

The power differences between the designs can become rather large and are in line with the asymptotic relative efficiencies. For the examples of Table 4, the crossover design always is most efficient and in the simulation also has the

highest power. Additional simulations show that similar conclusions can be drawn for ranges of intraclass correlations for which the crossover design is not most efficient. As expected, when testing the treatment by period interaction, the extended parallel design has more power than the crossover design (Appendix D).

## 6. Application in Planning a Trial

Suppose one would like to perform a randomized trial on the effectiveness of indacaterol versus tiotropium, among subjects suffering from chronic obstructive pulmonary disease, similar to Donohue et al. [16]. After 12 weeks of treatment, one plans to evaluate the effect of 18  $\mu\text{g}$  of tiotropium versus 300  $\mu\text{g}$  of indacaterol on the bronchodilator efficacy of 24 h postdose forced expiratory volume in 1 s (FEV1 in mL). The variance of FEV1 in the indacaterol and the tiotropium conditions in the study by Donohue et al. [16] differed significantly from each other, their ratio being 1.58. No information was available on the intraclass correlations and the research costs. Suppose reasonable guesses on the intraclass correlations are  $\rho_A \in [0.10, 0.70]$  for indacaterol and  $\rho_B \in [0.30, 0.90]$  for tiotropium. If one aims at minimizing the total sample size (so we set  $c_A = c_B = c_t = 0$  and  $c_{sp} = c_{s\_2p} = 1$ ), the maximin crossover design is more efficient than the AA/BB and A/B designs requiring only 48% and 43%, respectively, of the number of subjects of these designs. To be able to detect a medium effect ( $ES = 0.5$ , see below (9)) with 80% power in a two-tailed test with a 5% type I error rate, taking as maximin parameter values  $\rho_A^* = 0.10$  and  $\rho_B^* = 0.30$  in (10), 54 subjects are needed. Since the sample size calculation in (10) is based on the standard normal, whereas the test statistic follows a  $t$ -distribution, we add 1 subject to each treatment sequence [12], yielding a total sample size of 56 subjects with 28 subjects being allocated to each of the two treatment sequences of the crossover design.

## 7. Conclusion and Discussion

We examined the asymptotic efficiency of the ML estimator of the treatment and the treatment by period interaction effect for three two-treatment designs: a parallel, an extended parallel, and a crossover design. For a flexible cost function, the optimal allocations to the treatment sequences and corresponding optimal efficiencies were derived. Since commonly the intraclass correlations for each of the treatments and the ratio of treatment-dependent variances are not precisely known, also maximin designs were derived, which guarantee a power level across plausible ranges of values for the intraclass correlations at the lowest costs.

When interested in testing the main effects of the treatments, the relations between the efficiencies of the maximin versions of the A/B, AB/BA, and AA/BB designs depend on assumed ranges of the intraclass correlations, on the costs of the treatments and the costs of recruiting and measuring subjects. A numerical investigation shows that if A and B are equally expensive or the sum of the costs of one treatment and measurement per person are less than the remaining subject-specific costs (such as recruitment costs), then the crossover design is most efficient for ranges of

intraclass correlations starting at 0.15 or higher. In other cost scenarios, also for ranges of the intraclass correlations above 0.15, the parallel design or its extended version may become most efficient. Then, the efficiency relations are complicated, and the most efficient design is best determined by the results in Table 2. For the treatment by period interaction, however, the maximin AA/BB trial is proven to be more efficient than the maximin AB/BA design.

Since the efficiency comparisons of the maximin designs were based on asymptotic variances, a Monte Carlo simulation study was done for small samples. After applying correction factors in sample size planning based on the standard normal distribution, it was shown that (a) the asymptotic relative efficiencies translate into corresponding relative power levels and (b) power levels targeted in sample size planning are realized. This illustrates the practical utility of these results for sample size calculation.

If prerandomization measurements of the outcome variable are available, these could be included as covariates in the analysis [1]. Adding covariates in a randomized trial will not change the treatment effect of interest but will lead to a reduction of the intercept variance and thus of the intraclass correlations [3]. Provided that the costs of prerandomization measurements are the same for all designs (and there are no missing values on these prerandomization covariates), the results of the present study also apply.

The present study did not consider carry over in deriving optimal and maximin designs. If there is *self carry over*, that is, carry over from a treatment onto itself, this implies that steady-state did not yet occur in the first period, and then the total treatment effect would be the relevant effect, that is, the direct effect in the first period plus the carry over effect in the second period [5]. If there is self carry over, one-period designs and the AB/BA design are not suitable, as they do not allow for estimating the total treatment effect, leaving only the AA/BB design as a suitable option. There may also be *steady-state carry over*, which can only occur if there is a switch of treatments [17]. Such carry over would affect the efficiency of the crossover design. Although one commonly tries to avoid carry over, examining to what extent steady-state carry over affects the relative efficiency of the maximin crossover design would be an interesting issue for further research.

## Appendix

### A.1. Asymptotic Variance for the ML Estimator of the Treatment Effect

Let the vector of observed scores on the dependent variable for person  $j (= 1, \dots, N)$  in a two-period design be denoted as  $\mathbf{y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \end{bmatrix}$ . The linear mixed model for the scores  $\mathbf{y}_j$  can be expressed as follows:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}u_{0j} + \mathbf{Q}\boldsymbol{\epsilon}_j^A + (\mathbf{I} - \mathbf{Q})\boldsymbol{\epsilon}_j^B, \quad (\text{A.1})$$

where  $\mathbf{X}_j$  is the design matrix for subject  $j$ ,  $\boldsymbol{\beta}$  is the vector of regression coefficients,  $u_{0j}$  is the random person effect,

$\boldsymbol{\varepsilon}_j^A = \begin{bmatrix} \varepsilon_{1j}^A \\ \varepsilon_{2j}^A \end{bmatrix}$  is the vector of residual scores under treatment

A, and  $\boldsymbol{\varepsilon}_j^B = \begin{bmatrix} \varepsilon_{1j}^B \\ \varepsilon_{2j}^B \end{bmatrix}$  is the vector of residual scores under

treatment B. In (A.1),  $\mathbf{I} = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$ ,  $\mathbf{I}$  is the identity matrix, and  $\mathbf{Q}$

is a matrix which indicates whether treatment A has been given or not in a particular period, so for a person  $j$  with

treatment sequence AB, we have  $\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ , indicating that

A has been given in period 1 and treatment B in period 2.

Let  $\mathbf{J}$  be a matrix with only ones (of order 2 by 2). The variance-covariance matrix of  $\mathbf{y}_j$  can be derived as follows:

$$\mathbf{V}_j = \mathbf{Q}^T \mathbf{Q} \sigma_{\varepsilon A}^2 + (\mathbf{I} - \mathbf{Q})^T (\mathbf{I} - \mathbf{Q}) \sigma_{\varepsilon B}^2 + \sigma_0^2 \mathbf{J}. \quad (\text{A.2})$$

For the AA sequence, we have  $\mathbf{Q} = \mathbf{I}$ , and (A.2) can be rewritten as follows:

$$\mathbf{V}_j = \sigma_{\varepsilon A}^2 \mathbf{I} + \sigma_0^2 \mathbf{J}. \quad (\text{A.3})$$

By applying the result that, for an  $n \times n$  matrix,  $\mathbf{X} = a\mathbf{I} + b\mathbf{J}$ , with  $a \neq 0$  and  $a \neq -nb$ , the inverse is  $\mathbf{X}^{-1} = (1/a)(\mathbf{I} - \mathbf{J}(b/(nb + a)))$  ([18], p. 443), the inverse of the matrix in (A.3) can be written as follows:

$$\begin{aligned} \mathbf{V}_j^{-1} &= \frac{1}{\sigma_{\varepsilon A}^2} \left( \mathbf{I} - \mathbf{J} \left( \frac{\sigma_0^2}{2\sigma_0^2 + \sigma_{\varepsilon A}^2} \right) \right) \\ &= \frac{1}{\sigma_{\varepsilon A}^2} \times \frac{1}{2\sigma_0^2 + \sigma_{\varepsilon A}^2} \times \begin{pmatrix} \sigma_0^2 + \sigma_{\varepsilon A}^2 & -\sigma_0^2 \\ -\sigma_0^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 \end{pmatrix}. \end{aligned} \quad (\text{A.4})$$

For the BB sequence, we have  $\mathbf{I} - \mathbf{Q} = \mathbf{I}$ , and in a similar way, we obtain

$$\begin{aligned} \mathbf{V}_j^{-1} &= \frac{1}{\sigma_{\varepsilon B}^2} \left( \mathbf{I} - \mathbf{J} \left( \frac{\sigma_0^2}{2\sigma_0^2 + \sigma_{\varepsilon B}^2} \right) \right) \\ &= \frac{1}{\sigma_{\varepsilon B}^2} \times \frac{1}{2\sigma_0^2 + \sigma_{\varepsilon A}^2} \times \begin{pmatrix} \sigma_0^2 + \sigma_{\varepsilon B}^2 & -\sigma_0^2 \\ -\sigma_0^2 & \sigma_0^2 + \sigma_{\varepsilon B}^2 \end{pmatrix}. \end{aligned} \quad (\text{A.5})$$

Finally, for the AB sequence and BA sequence, we can derive the following equations, respectively:

$$\begin{aligned} \mathbf{V}_j^{-1} &= \frac{1}{(\sigma_0^2 + \sigma_{\varepsilon A}^2)(\sigma_0^2 + \sigma_{\varepsilon B}^2) - \sigma_0^4} \times \begin{pmatrix} \sigma_0^2 + \sigma_{\varepsilon B}^2 & -\sigma_0^2 \\ -\sigma_0^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 \end{pmatrix}, \\ \mathbf{V}_j^{-1} &= \frac{1}{(\sigma_0^2 + \sigma_{\varepsilon A}^2)(\sigma_0^2 + \sigma_{\varepsilon B}^2) - \sigma_0^4} \times \begin{pmatrix} \sigma_0^2 + \sigma_{\varepsilon A}^2 & -\sigma_0^2 \\ -\sigma_0^2 & \sigma_0^2 + \sigma_{\varepsilon B}^2 \end{pmatrix}. \end{aligned} \quad (\text{A.6})$$

The information matrix of the ML estimator of  $\boldsymbol{\beta}$  can be written as

$$\text{Inf } \hat{\boldsymbol{\beta}} = \sum_{j=1}^N \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j. \quad (\text{A.7})$$

Taking the inverse of the information matrix yields the asymptotic variance-covariance matrix of the ML estimators of  $\boldsymbol{\beta}$ .

Different treatment sequences not only lead to different matrices  $\mathbf{V}_j$  but also to different  $\mathbf{X}_j$  matrices. As an example, we consider the crossover design. In total, we have  $N$  subjects, and  $p$  is the proportion of persons being allocated to the AB sequence. Let  $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2)$ , where the regression coefficients represent the intercept, the treatment effect, and the time effect, respectively. For persons

allocated to the AB sequence, we have  $\mathbf{X}_j = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ ,

where  $j = 1, \dots, Np$ . For persons allocated to the BA sequence, we have  $\mathbf{X}_j = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ , where  $j = Np + 1, \dots, N$ .

We can now elaborate the matrix formulation in (A.7). For persons in the AB sequence, we have

$$\begin{aligned} \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j &= \frac{1}{(\sigma_0^2 + \sigma_{\varepsilon A}^2)\sigma_{\varepsilon B}^2 + \sigma_0^2\sigma_{\varepsilon A}^2} \\ &\cdot \begin{bmatrix} \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon B}^2 & \sigma_{\varepsilon A}^2 & \sigma_{\varepsilon A}^2 \\ \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 \\ \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 \end{bmatrix}. \end{aligned} \quad (\text{A.8})$$

and for persons in the BA sequence, we obtain

$$\begin{aligned} \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j &= \frac{1}{(\sigma_0^2 + \sigma_{\varepsilon A}^2)\sigma_{\varepsilon B}^2 + \sigma_0^2\sigma_{\varepsilon A}^2} \\ &\cdot \begin{bmatrix} \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon B}^2 & \sigma_{\varepsilon A}^2 & \sigma_{\varepsilon B}^2 \\ \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & -\sigma_0^2 \\ \sigma_{\varepsilon B}^2 & -\sigma_0^2 & \sigma_0^2 + \sigma_{\varepsilon B}^2 \end{bmatrix}. \end{aligned} \quad (\text{A.9})$$

The information matrix in (A.7) can be obtained by summing the expressions in (A.8) and (A.9) across all persons in each of the treatment sequences, and the asymptotic variance-covariance matrix of the ML estimators then results by taking the inverse of this matrix. We are interested in the variance of the treatment effect estimator, which is the entry in row 2 and column 2 of the resulting variance-covariance matrix:

$$\begin{aligned} \text{Var}(\hat{\beta}_1 | \xi_{\text{crossover}}) &= \frac{(\sigma_{\varepsilon B}^2 - \sigma_{\varepsilon A}^2)^2 p(1-p) + (\sigma_0^2 + \sigma_{\varepsilon A}^2)\sigma_{\varepsilon B}^2 + \sigma_0^2 \cdot \sigma_{\varepsilon A}^2}{(4\sigma_0^2 + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon B}^2)p(1-p)N}, \end{aligned} \quad (\text{A.10})$$

which, noting that  $\rho_A = \sigma_0^2/(\sigma_0^2 + \sigma_{\varepsilon A}^2)$ ,  $\rho_B = \sigma_0^2/(\sigma_0^2 + \sigma_{\varepsilon B}^2)$ , and  $\sigma_y^2 = 2\sigma_0^2 + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon B}^2$ , can be written as the expression given in Table 5. Along similar lines, the variance of the treatment effect estimator for a parallel and an extended parallel design, as shown in Table 5, can be derived.

TABLE 5: Asymptotic variances of the ML estimator of the treatment effect and treatment by period interaction effect for different designs under heterogeneity of outcome variances.

Design	$\text{Var}(\widehat{\beta}_{\text{treat}})$	$\text{Var}(\widehat{\beta}_{\text{treatment}})$
Crossover design	$\frac{((\rho_A - \rho_B)^2 p(1-p) + \rho_A \rho_B (1 - \rho_A \rho_B)) / ((\rho_A + \rho_B)(\rho_A + \rho_B + 2\rho_A \rho_B)) (\sigma_y^2 / ((1-p)pN))}{((\rho_B / (\rho_A + \rho_B))(1/p) + (\rho_A / (\rho_A + \rho_B))1 / (1-p)) (\sigma_y^2 / N)}$	$(1 + 2\rho_A \rho_B / (\rho_A + \rho_B)) (\sigma_y^2 / ((1-p)pN))$
Parallel design	$\frac{((\rho_B / (\rho_A + \rho_B))(1/p) + (\rho_A / (\rho_A + \rho_B))1 / (1-p)) (\sigma_y^2 / N)}{((\rho_B / (\rho_A + \rho_B))(1/p) + (\rho_A / (\rho_A + \rho_B))1 / (1-p)) + ((\rho_A \rho_B) / (\rho_A + \rho_B)) (1 / ((1-p)p)) (\sigma_y^2 / (2N))}$	na
Extended parallel design	$\frac{((\rho_B / (\rho_A + \rho_B))(1/p) + (\rho_A / (\rho_A + \rho_B))1 / (1-p)) + ((\rho_A \rho_B) / (\rho_A + \rho_B)) (1 / ((1-p)p)) (\sigma_y^2 / (2N))}{((\rho_B (1 - \rho_A)) / (\rho_A + \rho_B)) (1/p) + ((\rho_A (1 - \rho_B)) / (\rho_A + \rho_B)) (1 / (1-p))} (\sigma_y^2 / N)$	$((\rho_B (1 - \rho_A)) / (\rho_A + \rho_B)) (1/p) + ((\rho_A (1 - \rho_B)) / (\rho_A + \rho_B)) (1 / (1-p)) (\sigma_y^2 / N)$

Note.  $\sigma_y^2 = \sigma_A^2 + \sigma_B^2$ ; na: not applicable.

## A.2. Asymptotic Variance for the ML Estimator of the Treatment by Period Interaction Effect

The derivations of the variance of treatment by period interaction estimator are also similar, but start from another model and corresponding design matrices for the fixed regression coefficients. Now  $\beta^T = (\beta_0, \beta_1, \beta_2, \beta_3)$ , where the regression coefficients represent the intercept, the treatment effect, the time effect, and the treatment by period interaction effect, respectively. When considering again a crossover trial as an example, for persons assigned to the AB sequence, we have

$$\mathbf{X}_j = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \text{ where } j = 1, \dots, Np. \text{ For persons assigned}$$

to the BA sequence, we have  $\mathbf{X}_j = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$ , where  $j = Np + 1, \dots, N$ . For persons in the AB sequence, we obtain

$$\mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j = \frac{1}{(\sigma_0^2 + \sigma_{\varepsilon A}^2) \sigma_{\varepsilon B}^2 + \sigma_0^2 \sigma_{\varepsilon A}^2} \begin{bmatrix} \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon B}^2 & \sigma_{\varepsilon A}^2 & \sigma_{\varepsilon A}^2 & \sigma_{\varepsilon A}^2 \\ \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 \\ \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 \\ \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 \end{bmatrix}, \quad (\text{A.11})$$

and for persons in the BA sequence, we obtain

$$\mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j = \frac{1}{(\sigma_0^2 + \sigma_{\varepsilon A}^2) \sigma_{\varepsilon B}^2 + \sigma_0^2 \sigma_{\varepsilon A}^2} \begin{bmatrix} \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon B}^2 & \sigma_{\varepsilon A}^2 & \sigma_{\varepsilon B}^2 & 0 \\ \sigma_{\varepsilon A}^2 & \sigma_0^2 + \sigma_{\varepsilon A}^2 & -\sigma_0^2 & 0 \\ \sigma_{\varepsilon B}^2 & -\sigma_0^2 & \sigma_0^2 + \sigma_{\varepsilon B}^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (\text{A.12})$$

The information matrix in (A.7) is obtained by summing the expressions in (A.11) and (A.12) across all persons in both treatment sequences, and the asymptotic variance-covariance matrix of the ML estimators then results by taking the inverse of this matrix. We are interested in the variance of the treatment by period interaction effect estimator, which is the entry in row 4 and column 4 of the variance-covariance matrix:

$$\text{Var}(\widehat{\beta}_3 | \xi_{\text{cross-over}}) = \frac{(4\sigma_0^2 + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon B}^2)}{p(1-p)N}, \quad (\text{A.13})$$

which can be rewritten in terms of  $\rho_A$ ,  $\rho_B$ , and  $\sigma_y^2$ , as the expression in Table 5. Along similar lines, the variance of the treatment by period interaction effect estimator for an extended parallel design, as shown in Table 5, can be derived.

## B. Derivation of Optimal Allocations under a Budget Constraint

The asymptotic variance of the treatment effect estimator is minimized as a function of  $p$ , the allocation proportion, given a fixed budget  $C$ . For a crossover, parallel, and extended parallel design,  $p$  is the proportion allocated to the sequence AB, A, and AA, respectively. The variance for a crossover design as given in Table 5 can be rewritten in terms of the research costs and budget  $C$ , employing the cost function in (5) and noting that  $c_{sc} = c_{s,2p}$ :

$$\begin{aligned} \text{Var}(\widehat{\beta}_1 | \xi_{\text{crossover}}) &= \left( \frac{(\rho_A - \rho_B)^2 p(1-p) + \rho_A \rho_B (1 - \rho_A \rho_B)}{(\rho_A + \rho_B)(\rho_A + \rho_B + 2\rho_A \rho_B)} \right) \\ &\quad \cdot \left( \frac{\sigma_y^2}{(1-p)pC} \right) (c_A + c_B + 2c_t + c_{s,2p}) \\ &= \left( (\rho_A - \rho_B)^2 + \frac{\rho_A \rho_B (1 - \rho_A \rho_B)}{(1-p)p} \right) \\ &\quad \cdot \left( \frac{\sigma_y^2 (c_A + c_B + 2c_t + c_{s,2p})}{(\rho_A + \rho_B)(\rho_A + \rho_B + 2\rho_A \rho_B)C} \right). \end{aligned} \quad (\text{B.1})$$

It is easy to see that this expression is minimized if  $p(1-p)$  is maximized, which is the case if  $p = 0.5$ .

For a parallel design, the variance of the treatment effect estimator as given in Table 5 can be rewritten in terms of the cost function in (4) as follows:

$$\begin{aligned} \text{Var}(\widehat{\beta}_1 | \xi_{\text{parallel}}) &= \left( \left( \frac{\rho_B}{\rho_A + \rho_B} \right) \frac{1}{p} + \left( \frac{\rho_A}{\rho_A + \rho_B} \right) \frac{1}{(1-p)} \right) \frac{\sigma_y^2}{C} \\ &\quad \cdot (c_{sp} + c_t + pc_A + (1-p)c_B). \end{aligned} \quad (\text{B.2})$$

Taking the derivative of (B.2) with respect to  $p$  yields the following expression:

$$\begin{aligned} & \left( \left( (2(c_{sp} + c_t) + c_A + c_B) \left( \frac{\rho_A}{\rho_A + \rho_B} \right) - (c_{sp} + c_t + c_B) \right) p^2 \right. \\ & \left. + 2(c_{sp} + c_t + c_B) \left( \frac{\rho_B}{\rho_A + \rho_B} \right) p - (c_{sp} + c_t + c_B) \left( \frac{\rho_B}{\rho_A + \rho_B} \right) \right) \\ & \cdot \frac{\sigma_y^2}{Cp^2(1-p)^2} \end{aligned} \quad (B.3)$$

Solving the expression for  $p$  gives two solutions, one of which turns out to give a minimum (the second derivative of the variance as a function of  $p$  is positive for this particular value):

$$p = \frac{\sqrt{(c_{sp} + c_t + c_B)\rho_B}}{\sqrt{(c_{sp} + c_t + c_B)\rho_B} + \sqrt{(c_{sp} + c_t + c_A)\rho_A}} \quad (B.4)$$

Finally, for an extended parallel design, the variance of the treatment effect estimator as given in Table 5 can be rewritten in terms of the cost function in (6) noting that  $c_{sep} = c_{s_{2p}}$ :

$$\begin{aligned} \text{Var}(\hat{\beta}_1 | \xi_{\text{extended parallel}}) &= \left( \left( \frac{\rho_B}{\rho_A + \rho_B} \right) \frac{1}{p} + \left( \frac{\rho_A}{\rho_A + \rho_B} \right) \frac{1}{(1-p)} \right. \\ & \left. + \left( \frac{\rho_A \rho_B}{\rho_A + \rho_B} \right) \frac{1}{(1-p)p} \right) \\ & \cdot \frac{\sigma_y^2 (c_{s_{2p}} + 2pc_A + 2(1-p)c_B + 2c_t)}{2C} \end{aligned} \quad (B.5)$$

Taking the derivative with respect to  $p$  yields the following expression:

$$\begin{aligned} & \left( \left( (c_{s_{2p}} + 2c_t) \left( \frac{\rho_A - \rho_B}{\rho_A + \rho_B} \right) - 2c_B \left( \frac{(1 + \rho_A)\rho_B}{\rho_A + \rho_B} \right) \right. \right. \\ & \left. \left. + 2c_A \left( \frac{(1 + \rho_B)\rho_A}{\rho_A + \rho_B} \right) \right) p^2 + 2(c_{s_{2p}} + 2c_t + 2c_B) \right. \\ & \left. \cdot \left( \frac{\rho_B(1 + \rho_A)}{\rho_A + \rho_B} \right) p - (c_{s_{2p}} + 2c_t + 2c_B) \left( \frac{\rho_B(1 + \rho_A)}{\rho_A + \rho_B} \right) \right) \\ & \cdot \frac{\sigma_y^2}{2Cp^2(1-p)^2} \end{aligned} \quad (B.6)$$

Solving (B.6) for  $p$  gives two solutions, one of which turns out to give a minimum (the second derivative of the variance as a function of  $p$  is positive for his particular value):

$$p = \frac{\sqrt{(c_{s_{2p}} + 2(c_t + c_B))(1 + \rho_A)\rho_B}}{\sqrt{(c_{s_{2p}} + 2(c_t + c_B))(1 + \rho_A)\rho_B} + \sqrt{(c_{s_{2p}} + 2(c_t + c_A))(1 + \rho_B)\rho_A}} \quad (B.7)$$

Substituting the optimal allocation  $p = 0.5$ , and the allocations in (B.4) and (B.7) into the corresponding

expressions for the variance of the treatment effect estimator yields the optimal variances as given in Table 1 of the main text. Derivations along similar lines can be done if interest is in minimizing the variance of the treatment by period effect estimator. These result in the allocations and variances, as also presented in Table 1 of the main text.

### C. Derivation of Maximin Designs and Variances of the Effect Estimators

Maximin designs optimize the allocation to the treatment sequences under the worst case, that is, the maximum variance of the treatment effect estimator across plausible ranges of the model parameters, here the intraclass correlations  $\rho_A$  and  $\rho_B$ .

For crossover designs, the variance of the treatment effect estimator under optimal allocation to the treatments is (see Table 1 of the main text)

$$\text{Var}(\hat{\beta}_1 | \xi_{\text{crossover}}) = (c_A + c_B + 2c_t + c_{s_{2p}}) \left( 1 - \frac{2\rho_A\rho_B}{\rho_A + \rho_B} \right) \frac{\sigma_y^2}{C} \quad (C.1)$$

By taking the derivatives with respect to  $\rho_A$  and  $\rho_B$ , it can be shown that this expression decreases as a function of these parameters. This implies that the worst case occurs for the lower bounds of the ranges for  $\rho_A$  and  $\rho_B$ . This yields the expression for the variance of the treatment effect estimator in the case of a maximin crossover design in Table 2 of the main text.

Similarly, for a parallel design, we have to examine for which values of  $\rho_A$  and  $\rho_B$  the following expression is maximized:

$$\begin{aligned} \text{Var}(\hat{\beta}_1 | \xi_{\text{parallel}}) &= \left( \sqrt{(c_A + c_t + c_{sp}) \left( \frac{\rho_B}{\rho_A + \rho_B} \right)} \right. \\ & \left. + \sqrt{(c_B + c_t + c_{sp}) \left( \frac{\rho_A}{\rho_A + \rho_B} \right)} \right)^2 \frac{\sigma_y^2}{C} \end{aligned} \quad (C.2)$$

For derivations purposes, it is more convenient to rewrite (C.2) in terms of  $\phi = \rho_B/\rho_A$ :

$$\begin{aligned} \text{Var}(\hat{\beta}_1 | \xi_{\text{parallel}}) &= \left( \sqrt{(c_A + c_t + c_{sp}) \left( \frac{\phi}{1 + \phi} \right)} \right. \\ & \left. + \sqrt{(c_B + c_t + c_{sp}) \left( \frac{1}{1 + \phi} \right)} \right)^2 \frac{\sigma_y^2}{C} \end{aligned} \quad (C.3)$$

Taking the derivative of (C.3) with respect to  $\phi$ , we find that the variance increases as a function of  $\phi$  as long as  $\phi \leq (c_A + c_t + c_{sp})/(c_B + c_t + c_{sp})$  and decreases as a function of  $\phi$  if  $\phi > (c_A + c_t + c_{sp})/(c_B + c_t + c_{sp})$ . So, if we can choose  $\rho_A$  and  $\rho_B$  from their plausible ranges such that  $\rho_B/\rho_A = (c_A + c_t + c_{sp})/(c_B + c_t + c_{sp})$ , this maximizes (C.2). On the other hand, if even the lower bound of  $\rho_A$ ,  $\rho_A^L$ , and the upper bound of  $\rho_B$ ,  $\rho_B^U$ , yield  $\rho_B^U/\rho_A^L < (c_A + c_t + c_{sp})/(c_B + c_t + c_{sp})$ , then choose  $\rho_A^L$  and  $\rho_B^U$ . Furthermore, if



$(c_A + c_t + c_{sp}) / (c_B + c_t + c_{sp}) < \rho_B^L / \rho_A^U$ , then choose  $\rho_A^U$  and  $\rho_B^L$ . Substituting the maximin values of  $\rho_A$  and  $\rho_B$  into (C.2) will result in the variances of the treatment effect estimator as displayed in Table 2 of the main text.

For the extended parallel design, the variance for optimal allocation to the treatment sequences is given in Table 1 of the main text:

$$\begin{aligned} \text{Var}(\hat{\beta}_1 | \xi_{\text{extended parallel}}) &= \left( \sqrt{(2(c_A + c_t) + c_{s_{2p}}) \left( \frac{\rho_B(1 + \rho_A)}{\rho_A + \rho_B} \right)} \right. \\ &\quad \left. + \sqrt{(2(c_B + c_t) + c_{s_{2p}}) \left( \frac{\rho_A(1 + \rho_B)}{\rho_A + \rho_B} \right)} \right)^2 \\ &\quad \cdot \frac{\sigma_y^2}{2C}. \end{aligned} \quad (\text{C.4})$$

Taking the derivative of (C.4) with respect to  $\rho_A$  shows that the variance of the treatment effect estimator increases as a function of  $\rho_A$  as long as  $\lambda(1 - \rho_B)^2 / (\rho_B(1 + \rho_B)) \leq (1 + \rho_A) / \rho_A$  where  $\lambda = (2(c_A + c_t) + c_{s_{2p}}) / (2(c_B + c_t) + c_{s_{2p}})$  and decreases if  $\lambda(1 - \rho_B)^2 / (\rho_B(1 + \rho_B)) > (1 + \rho_A) / \rho_A$ . Taking into account a feasible range for  $\rho_A$ , the value for  $\rho_A$  maximizing the variance in (C.4),  $\rho_A^*$ , therefore is

$$\rho_A^* = \frac{\rho_B(1 + \rho_B)}{\lambda(1 - \rho_B)^2 - \rho_B(1 + \rho_B)} \quad \text{if} \quad \rho_A^L \leq \rho_A^* \leq \rho_A^U \quad (\text{C.5})$$

$$\text{or} \quad \frac{1}{\rho_A^U} + 1 \leq \frac{\lambda(1 - \rho_B)^2}{(1 + \rho_B)\rho_B} \leq \frac{1}{\rho_A^L} + 1,$$

$$\rho_A^* = \rho_A^U \quad \text{if} \quad \rho_A^U \leq \rho_A^* \quad \text{or} \quad \frac{\lambda(1 - \rho_B)^2}{(1 + \rho_B)\rho_B} < \frac{1}{\rho_A^U} + 1, \quad (\text{C.6})$$

$$\rho_A^* = \rho_A^L \quad \text{if} \quad \rho_A^* < \rho_A^L \quad \text{or} \quad \frac{1}{\rho_A^L} + 1 < \frac{\lambda(1 - \rho_B)^2}{(1 + \rho_B)\rho_B}. \quad (\text{C.7})$$

Taking the derivative of (C.4) with respect to  $\rho_B$  shows that the variance increases as a function of  $\rho_B$  as long as  $(1 - \rho_A)^2 / (\lambda\rho_A(1 + \rho_A)) \leq (1 + \rho_B) / \rho_B$  and decreases if  $(1 - \rho_A)^2 / (\lambda\rho_A(1 + \rho_A)) > (1 + \rho_B) / \rho_B$ . So, also taking into account a feasible range for  $\rho_B$ , we have as maximin value for  $\rho_B$ :

$$\rho_B^* = \frac{\lambda\rho_A(1 + \rho_A)}{(1 - \rho_A)^2 - \lambda\rho_A(1 + \rho_A)} \quad \text{if} \quad \rho_B^L \leq \rho_B^* \leq \rho_B^U \quad (\text{C.8})$$

$$\text{or} \quad \frac{1}{\rho_B^U} + 1 \leq \frac{(1 - \rho_A)^2}{\lambda(1 + \rho_A)\rho_A} \leq \frac{1}{\rho_B^L} + 1,$$

$$\rho_B^* = \rho_B^U \quad \text{if} \quad \rho_B^U < \rho_B^* \quad \text{or} \quad \frac{(1 - \rho_A)^2}{\lambda(1 + \rho_A)\rho_A} < \frac{1}{\rho_B^U} + 1, \quad (\text{C.9})$$

$$\rho_B^* = \rho_B^L \quad \text{if} \quad \rho_B^* < \rho_B^L \quad \text{or} \quad \frac{1}{\rho_B^L} + 1 < \frac{(1 - \rho_A)^2}{\lambda(1 + \rho_A)\rho_A}. \quad (\text{C.10})$$

For each of the intraclass correlations,  $\rho_A$  and  $\rho_B$ , there are three possible values that maximize the variance of the treatment effect estimator. Not all values can cooccur.

TABLE 6

Treatment by period interaction				
	N	Crossover design	Extended parallel design	
Type I error rate = 0.05				
$[\rho_A^L, \rho_A^U]$		$[\rho_B^L, \rho_B^U]$		
[0.01, 0.10]	[0.01, 0.10]	110	<b>0.799</b> (0.90)	0.840 (1)
[0.01, 0.30]	[0.01, 0.30]	130	<b>0.801</b> (0.76)	0.899 (1)
[0.70, 1.00]	[0.70, 1.00]	34	0.206 (0.15)	<b>0.830</b> (1)
[0.90, 1.00]	[0.90, 1.00]	15	0.111 (0.05)	<b>0.883</b> (1)
Type I error rate = 0.01				
$[\rho_A^L, \rho_A^U]$		$[\rho_B^L, \rho_B^U]$		
[0.01, 0.10]	[0.01, 0.10]	166	<b>0.808</b> (0.90)	0.848 (1)
[0.01, 0.30]	[0.01, 0.30]	194	<b>0.800</b> (0.76)	0.912 (1)
[0.70, 1.00]	[0.70, 1.00]	51	0.117 (0.15)	<b>0.840</b> (1)
[0.90, 1.00]	[0.90, 1.00]	23	0.048 (0.05)	<b>0.911</b> (1)

Note. The power printed in bold indicates the design for which the sample calculation should yield a power of 80%. N: total sample size.

(C.5) and (C.8) imply that  $(1 + \rho_B^*)^2(1 + \rho_A^*)^2 = (1 - \rho_B^*)^2(1 - \rho_A^*)^2$ , which in turn is true only if  $\rho_A^* = \rho_B^* = 0$ , an unrealistic condition that we will not consider. Furthermore, if  $\rho_B^* = \rho_B^L$  then  $(1/\rho_B^L) + 1 < ((1 - \rho_A)^2) / (\lambda(1 + \rho_A)\rho_A)$  (C.10), which implies that  $\lambda < (((1 - \rho_A)^2) / ((1 + \rho_A)\rho_A)) \times (\rho_B^L / (\rho_B^L + 1))$ . If at the same time  $\rho_A^* = \rho_B(1 + \rho_B) / (\lambda(1 - \rho_B)^2 - \rho_B(1 + \rho_B))$ , then also  $\lambda = ((1 + \rho_A) / \rho_A) \times (\rho_B^L(\rho_B^L + 1) / (1 - \rho_B^L)^2)$ . The latter two conditions for  $\lambda$  imply that  $(1 + \rho_A)^2(1 + \rho_B^L)^2 < (1 - \rho_A)^2(1 - \rho_B^L)^2$ , which is never true. On the other hand, if  $\rho_A^* = \rho_B(1 + \rho_B) / (\lambda(1 - \rho_B)^2 - \rho_B(1 + \rho_B))$ , then it is possible that  $\rho_B^* = \rho_B^U$ . It can also be shown that if  $\rho_B^* = \lambda\rho_A(1 + \rho_A) / ((1 - \rho_A)^2 - \lambda\rho_A(1 + \rho_A))$ , then it is only possible that  $\rho_A^* = \rho_A^U$ . So, all this implies that we can rewrite the “if” conditions in (C.5)–(C.7), with  $\rho_B$  replaced by  $\rho_B^U$ , and the “if” conditions in (C.8)–(C.10), with  $\rho_A$  replaced by  $\rho_A^U$ .

Finally, we can show that the combination  $\rho_A^* = \rho_A^L$  and  $\rho_B^* = \rho_B^L$  cannot occur. This implies the inequalities in (C.7) and (C.10), which in turn imply that  $(1/\rho_A^L) + 1 < (\lambda(1 - \rho_B^U)^2) / ((1 + \rho_B^U)\rho_B^U)$  and  $(1/\rho_B^L) + 1 < ((1 - \rho_A^U)^2) / (\lambda(1 + \rho_A^U)\rho_A^U)$ , respectively, which imply  $\lambda > ((1 + \rho_A^L) / \rho_A^L) \times ((1 + \rho_B^U) / \rho_B^U) / (1 - \rho_B^U)^2$  and  $\lambda < ((1 - \rho_A^U)^2) / ((1 + \rho_A^U)\rho_A^U) \times (\rho_B^L / (\rho_B^L + 1))$ , respectively, which are incompatible. All other combinations of the maximin values  $\rho_A^*$  and  $\rho_B^*$  can be shown to be possible. These results lead to the procedure for determining the maximin design as delineated in Table 2 of the main text for an extended parallel design.

Similar derivations can be given for the maximin versions of a crossover and extended parallel design in the case interest is in the treatment by period interaction effect. These derivations are, upon request, available from the author.

#### D. Powers from the Monte Carlo Simulations for Maximin Designs in the Case of the Treatment by Period Interaction

Table 6 provides the simulated powers for maximin designs in the case of the treatment by period interaction. For each

pair of ranges of the intraclass correlations, the asymptotic efficiency of each design versus the most efficient design is given within brackets.

### Data Availability

This study is not based on empirical data. However, the R programs that are used in this paper are available upon request from the corresponding author.

### Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

### References

- [1] B. Jones and M. G. Kenward, *Design and Analysis of Cross-Over Trials*, Chapman & Hall, London, UK, 1989.
- [2] S. Senn, *Cross-Over Trials in Clinical Research*, Wiley, Chichester, UK, 2002.
- [3] H. Brown and R. Prescott, *Applied Mixed Models in Medicine*, Wiley, Chichester, UK, 2006.
- [4] M. P. F. Berger and W. K. Wong, *An Introduction to Optimal Designs for Social and Biomedical Research*, Wiley, Chichester, UK, 2009.
- [5] M. J. J. M. Candel, "Parallel, AA/BB, AB/BA and Balaam's design: efficient and maximin choices when testing the treatment effect through mixed effects linear regression," *Pharmaceutical Statistics*, vol. 11, no. 2, pp. 97–106, 2012.
- [6] R. J. Grissom, "Heterogeneity of variance in clinical data," *Journal of Consulting and Clinical Psychology*, vol. 68, no. 1, pp. 155–165, 2000.
- [7] L.-C. Huang, M.-J. Wen, and S. H. Cheung, "Noninferiority studies with multiple new treatments and heterogeneous variances," *Journal of Biopharmaceutical Statistics*, vol. 25, no. 5, pp. 958–971, 2015.
- [8] Y. Yuan and J. Zhou, "Cost-efficient higher-order crossover designs for two-treatment clinical trials," *Pharmaceutical Statistics*, vol. 4, no. 4, pp. 245–252, 2005.
- [9] B. W. Brown Jr., "The crossover experiment for clinical trials," *Biometrics*, vol. 36, no. 1, pp. 69–79, 1980.
- [10] R. Garcia, M. Benet, C. Arnau, and E. Cobo, "Efficiency of the cross-over design: an empirical evaluation," *Statistics in Medicine*, vol. 23, no. 24, pp. 3773–3780, 2004.
- [11] M. G. Kenward and J. H. Roger, "The use of baseline covariates in crossover studies," *Biostatistics*, vol. 11, no. 1, pp. 1–17, 2010.
- [12] G. J. P. Van Breukelen and M. J. J. M. Candel, "Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient!," *Journal of Clinical Epidemiology*, vol. 65, no. 11, pp. 1212–1218, 2012.
- [13] M. J. J. M. Candel and G. J. P. Van Breukelen, "Sample size calculation for treatment effects in randomized trials with fixed cluster sizes and heterogeneous intraclass correlations and variances," *Statistical Methods in Medical Research*, vol. 24, no. 5, pp. 557–573, 2015.
- [14] J. Cohen, "A power primer," *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992.
- [15] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013, <http://www.R-project.org/>.
- [16] J. F. Donohue, C. Fogarty, J. Lötval, D. A. Mahler, and H. Worth, "Once-daily bronchodilators for chronic obstructive pulmonary disease: indacaterol versus tiotropium," *American Journal of Respiratory and Critical Care Medicine*, vol. 182, no. 2, pp. 155–162, 2010.
- [17] S. Senn and D. Lambrou, "Robust and realistic approaches to carry-over," *Statistics in Medicine*, vol. 17, no. 24, pp. 2849–2864, 1998.
- [18] S. R. Searle, G. Casella, and C. E. McCulloch, *Variance Components*, Wiley, New York, NY, USA, 1992.