

Parallel, AA/BB, AB/BA and Balaam's design: efficient and maximin choices when testing the treatment effect in a mixed effects linear regression

Citation for published version (APA):

Candel, M. J. J. M. (2012). Parallel, AA/BB, AB/BA and Balaam's design: efficient and maximin choices when testing the treatment effect in a mixed effects linear regression. *Pharmaceutical Statistics*, 11(2), 97-106. <https://doi.org/10.1002/pst.502>

Document status and date:

Published: 01/01/2012

DOI:

[10.1002/pst.502](https://doi.org/10.1002/pst.502)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 23 Mar. 2025

Parallel, AA/BB, AB/BA and Balaam's design: efficient and maximin choices when testing the treatment effect in a mixed effects linear regression[†]

Math J. J. M. Candel*

When examining the effect of treatment *A* versus *B*, there may be a choice between a parallel group design, an AA/BB design, an AB/BA cross-over and Balaam's design. In case of a linear mixed effects regression, it is examined, starting from a flexible function of the costs involved and allowing for subject dropout, which design is most efficient in estimating this effect. For no carry-over, the AB/BA cross-over design is most efficient as long as the dropout rate at the second measurement does not exceed $2\rho/(1+\rho)$, ρ being the intraclass correlation. For steady-state carry-over, depending on the costs involved, the dropout rate and ρ , either a parallel design or an AA/BB design is most efficient. For types of carry-over that allow for self carry-over, interest is in the direct treatment effect plus the self carry-over effect, with either an AA/BB or Balaam's design being most efficient. In case of insufficient knowledge on the dropout rate or ρ , a maximin strategy is devised: choose the design that minimizes the maximum variance of the treatment estimator. Such maximin designs are derived for each type of carry-over. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: AB/BA cross-over design; Balaam's design; efficient design; maximin design; (extended) parallel design

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 cross-over trial in which subjects receive both treatments, *A* and *B*, but the sequencing of the treatments differs between two randomly allocated groups [1, 2]. Such a design is considered most suited when examining treatments for chronic or ongoing diseases, such as asthma, rheumatism, migraine or (frequent) heartburn. In these cases, there is no real possibility that the disease gets cured, but the aim is to moderate the effects of the disease [2]. If there is no differential carry-over and the sample sizes are equal, an AB/BA design yields more efficient estimates of the treatment effect than a parallel design and, consequently, has more power to test this effect. However, if differential carry-over is possible, accounting for such an effect in the analysis reduces power, because the treatment effect can only be estimated for data obtained at the first measurement [2]. A third design, not used as often as the first two, is Balaam's design. Here, subjects are randomly allocated to one of four treatment sequences: AA, BB, AB and BA. This design allows for simultaneously estimating differential carry-over and treatment effects also making use of data at the second measurement [3]. A fourth design that we will consider is a special case of Balaam's design, only involving the treatment sequences AA and BB. This design extends the parallel design across two treatment periods.

If the outcome variable is continuous and (approximately) normally distributed, the data obtained in previous designs can be

analyzed by mixed effects linear regression [3]. Of primary interest is testing the treatment effect of, for instance, a new medication for heartburn. A relevant issue is which design is most efficient in estimating the treatment effect, thereby yielding maximum power for testing this effect. Such optimality has already been examined by Senn [2] when comparing cross-over and parallel designs, and by Berger and Wong [4] when comparing all four designs introduced before. The present study extends these studies in the sense that, compared with Berger and Wong [4], we will use a more flexible cost function, and compared with Senn [2], we will also consider Balaam's design and the AA/BB design. Both studies are extended in that we will allow for dropout of subjects and additionally derive maximin designs [5]. Robustness of cross-over designs to dropout has been examined for four-treatment four-period cross-over designs [6]. In this study, we will systematically examine the effect of dropout on the efficiencies of designs involving only two treatments and at most two periods.

The paper is structured as follows. Section 2 will discuss important features of the designs considered and will motivate the delimitations of the present study. Section 3 will present linear mixed model formulations for analyzing data from each of these

[†]Supporting information may be found in the online version of this article.

Department Methodology and Statistics, Maastricht University, Maastricht, The Netherlands

*Correspondence to: Math J. J. M. Candel, Department Methodology and Statistics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.
E-mail: Math.Candel@maastrichtuniversity.nl

designs. Next, section 4 will introduce the relative efficiency criterion for comparing the four designs and will provide asymptotic expressions for this criterion in case of maximum likelihood estimation of the treatment effect, also allowing for dropout. Section 5 will extend the relative efficiencies, starting from a fixed budget and a flexible cost function. Because the efficiency ordering of designs depends on the intraclass correlation and dropout rate, in Section 6, we will derive, employing a maximin strategy, optimal designs in case knowledge concerning these issues is insufficient. Section 7 will give an empirical application of the results and the paper will close with some issues for further research.

2. DELIMITATIONS OF THE STUDY

In designs where treatments are successively given to the same group of subjects, carry-over may occur. For an AB/BA cross-over trial, it may be that in the AB sequence A still has an effect on the outcome, once B has been given and the second measurement is carried out. When in the BA sequence, the effect of B is still present once A has been administered, and this effect differs from the carry-over effect for the AB sequence, differential carry-over occurs. In a cross-over trial differential carry-over may lead to biased estimates of the treatment effect [2]. For this reason, a two-stage approach has been proposed [7]: first perform a test on differential carry-over, and then choose, depending on the outcome of this test, the appropriate test for the treatment effect. If the test for differential carry-over is significant, only the data for the first measurement are used to obtain an unbiased estimate of the treatment effect, otherwise the data from both measurements are used. However, if it has been concluded incorrectly that there is differential carry-over, this two-stage procedure can be shown to seriously inflate the type I error for the test of the treatment effect [8]. Furthermore, once it has been concluded that differential carry-over occurs, only the data for the first period are examined. If sample sizes are chosen to guarantee a certain power level, having to omit the data from the second period will lead to an underpowered study.

Carry-over preferably should be avoided in designing studies. However, in some cases it is not known how long a washout period should be to be effective, in other cases sufficiently long washout periods are not possible because of time constraints, an increased probability of dropout of subjects or other practical constraints (cf. [9]). Also, although the present paper focuses on carry-over in a pharmacological context, it is possible that a washout period that is effective in a pharmacological sense, might not prevent carry-over because of lingering psychological effects [10].

We will assume that a priori, a decision is made as to whether and what type of carry-over is expected to occur. If carry-over is possible and sizeable, this effect should be taken care of in modeling the data and in sample size calculation. If carry-over can be safely excluded or is negligible, this effect does not need to be estimated in analyzing the data. Such a strategy prevents the problems attached to the two-stage analysis procedure.

In the literature, different kinds of carry-over have been identified and defined. In *simple carry-over*, the carry-over only depends on the engendering treatment, the treatment immediately preceding a treatment in the treatment sequence [2, 11, 12]. This type of carry-over has been studied extensively but can be considered rather unrealistic in a pharmacological setting [2, 12] and will, therefore, not be examined in what follows. Carry-over may

also depend on the perturbed treatment, the treatment following a treatment [13]. An example is *steady-state carry-over* in which there is carry-over in case treatments are switched, but no carry-over in case the same treatment is continued [11, 12, 14]. This might occur in multiple dose studies aimed at treating a chronic disease. A treatment regime consisting of multiple doses will lead to some steady state for the patient, and when followed by the same treatment regime, the same steady state will be reached within the second treatment period. This need not be true when the treatment regimes are switched. In this case, there might be uncertainty concerning the rates at which drugs are eliminated from the body or, because of practical constraints, the washout period might be too short.

A special case of steady-state carry-over applies when one of the treatments is a placebo. When a placebo is the engendering treatment, from a pharmacological point of view there can be no effect and thus also no carry-over effect. The carry-over effects with placebo as engendering treatment can then be set to zero [2, 11]. We will denote this as *no placebo carry-over*, in which case neither placebo carry-over nor self carry-over occurs.

If carry-over cannot be excluded and not one specific form of carry-over applies, it is advisable to take the *saturated carry-over* model [12]. This model also involves carry-over of a treatment into itself, which implies that the observation period is too short to reach a steady state [13, 14]. In such a case, one might not be interested in the direct effect, but rather in the total effect of treatment A versus B , that is, the direct effect plus self carry-over effect [2]. Again, if one of the treatments is a placebo, a special case arises where one of the treatments has no effect and thus also no carry-over effect. We will denote this as *no placebo carry-over with self carry-over*. Here, also self carry-over is involved, implying that also for this type of carry-over interest will be in the total effect, that is, the direct effect plus self carry-over effect, of the treatment versus the placebo.

In case of self carry-over, none of the designs, as examined in the present study, is suitable for estimating a treatment by period interaction, because interest is then in the total treatment effect in the second period. In case of no placebo carry-over and steady state carry-over, both Balaam's design and an extended parallel design allow for the estimation of a treatment by period interaction, whereas for no carry-over, all three two-period designs are suitable. Although estimating and testing for a treatment by period interaction might be worthwhile, in the present study we will assume that there is no such interaction that limits the design choice.

In a mixed effects analysis, subjects are conceived of as a random factor, whereas in a fixed effects analysis, subjects are considered a fixed factor. The fixed effects approach is known to yield unbiased estimates of the standard error of the treatment effect, whereas the mixed effects approach can yield biased estimates, the bias being non-negligible for small samples [3]. The mixed effects approach, however, is in line with aiming at generalizing one's results to a population of subjects. Furthermore, a mixed effects analysis is expected to result in a higher efficiency of the treatment estimate in case extra treatment information can be recovered from the subject totals. This will occur, irrespective of the type of two-period design, when subjects have missing observations [3]. As illustrated for an AB/BA cross-over trial [3, p.274], the efficiency of the mixed effects approach versus the fixed effects approach increases with an increasing number of missing values and with a decreasing between-subject variance of the outcome variable as compared to the within-subject variance. These

features motivate employing a mixed effects analysis. The next section specifies the different models.

3. LINEAR MIXED EFFECTS MODELS

In case of a parallel design, an extended parallel design and a cross-over design, the subjects are randomly and evenly allocated to one of two arms: in a parallel design to treatment A or treatment B, in an extended parallel design to treatment sequence AA or BB, and in a cross-over trial to treatment sequence AB or BA. In Balaam's design, there are four arms corresponding to four treatment sequences, AA, BB, AB, and BA, respectively, with the subjects being evenly and randomly distributed across these four arms.

The dependent variable is a quantitative outcome, denoted as y_{ij} for person j ($j = 1, \dots, N$) at measurement occasion i . If y_{ij} is (approximately) normally distributed, for a parallel design simple linear regression may be an adequate tool for data analysis:

$$y_{ij} = \beta_0 + \beta_1 \text{treat}_{ij} + \delta_{ij}. \quad (1)$$

Here, *treat* is coded 0 for persons having treatment A and coded 1 for persons having treatment B. The random term δ_{ij} is assumed to be normally distributed, with mean 0 and variance σ_δ^2 . The random term can be thought of as consisting of a random person (between-subject) effect, u_{0j} , and a random error (within-subject) effect, ϵ_{ij} . In formula: $\delta_{ij} = u_{0j} + \epsilon_{ij}$. These two sources of random variation can not be separated in a parallel trial.

For a cross-over AB/BA design and an extended parallel AA/BB design however, the variances of u_{0j} and ϵ_{ij} can be identified. For these designs, we can extend the linear regression model with a random intercept, yielding the following mixed effects model:

$$y_{ij} = \beta_0 + \beta_1 \text{treat}_{ij} + \beta_2 \text{time}_{ij} + \beta_3 \text{carry}_{ij} + u_{0j} + \epsilon_{ij}. \quad (2)$$

The random terms u_{0j} and ϵ_{ij} are independently normally distributed, with mean 0 and variances σ_0^2 and σ_ϵ^2 , respectively. Their relation with the variance in Equation (1) is $\sigma_\delta^2 = \sigma_0^2 + \sigma_\epsilon^2$. In Equation (2), *time* is coded 0 for observations at the first measurement and coded 1 for observations at the second measurement. *Carry* represents the difference in carry-over between the two treatment sequences, the differential carry-over effect. When B is the first treatment, *carry* is coded 1 at the second measurement, and coded 0 otherwise. Note that the carry-over effect for the AB sequence (for a cross-over trial) and for the AA sequence (for an extended parallel design), is indistinguishable from the time effect captured by β_2 . With these codings, β_1 represents the treatment effect, β_2 represents the time effect (including carry-over in

the AB or AA sequence) and β_3 represents the differential carry-over effect of the BA or BB sequence. In an AB/BA cross-over trial for each type of carry-over as distinguished in Section 2, we have $\beta_3 \neq 0$. In an AA/BB parallel trial, however, $\beta_3 \neq 0$ only if there is self carry-over. Remember that in this case (i.e., for saturated carry-over and no placebo carry-over with self-carry-over) the effect of interest is not β_1 , but the total effect of the treatments at the second measurement, corresponding to $\beta_t = \beta_1 + \beta_3$. This total effect can not be estimated in a parallel or in an AB/BA design.

In Balaam's design, three differential carry-over effects can be modelled:

$$y_{ij} = \beta_0 + \beta_1 \text{treat}_{ij} + \beta_2 \text{time}_{ij} + \beta_{AB} \text{carry}_{ABij} + \beta_{BB} \text{carry}_{BBij} + \beta_{BA} \text{carry}_{BAij} + u_{0j} + \epsilon_{ij}. \quad (3)$$

Here, β_1 represents the treatment effect, β_2 represents the period effect and, when present, the carry-over effect for the AA sequence. The variable *carry*AB is coded 0 except for the second measurement in the AB sequence, where the coding is 1. Similarly, *carry*BB is coded 0 except for the second measurement in the BB sequence, where the coding is 1, and *carry*BA is coded 0 except for the second measurement in the BA sequence, where the coding is 1. *carry*AB, *carry*BB and *carry*BA allow the carry-over effect for each corresponding treatment sequence to differ from the carry-over effect in the AA sequence. If β_{AB} , β_{BB} and β_{BA} in Equation (3) are nonzero, we have a saturated carry-over model in which no treatment by period interaction can be identified. No differential carry-over corresponds to $\beta_{AB} = \beta_{BB} = \beta_{BA} = 0$ in Equation (3). Steady-state carry-over corresponds to $\beta_{BB} = 0$, implying that carry-over only occurs when there is a switch of treatments. No placebo carry-over, with A being the placebo, corresponds to $\beta_{AB} = 0$ and $\beta_{BB} = 0$. The case of no carry-over for placebo A with self carry-over corresponds to $\beta_{AB} = 0$, thus allowing for carry-over effects only in the BA and BB sequences, the sequences that start with the active treatment. Table I summarizes the values of the regression coefficients. Again, for saturated carry-over or no placebo carry-over with self carry-over, the effect of interest is not β_1 , but the total effect of treatment A versus B at the second measurement, that is, $\beta_t = \beta_1 + \beta_{BB}$.

The parameters in Equations (1)–(3) can be estimated through maximum likelihood (ML). For instance, for the model in Equation (3), eight parameters have to be estimated: six fixed regression weights, β_0 , β_1 , β_2 , β_{AB} , β_{BB} , and β_{BA} , and two variance components, σ_0^2 and σ_ϵ^2 . A relevant concept is the intraclass correlation, which is the correlation between the two measurements obtained for the same person. For the models in Equations (2) and (3), this can be expressed as $\rho = \sigma_0^2 / (\sigma_0^2 + \sigma_\epsilon^2)$. The larger the

Table I. Constraints on the carry-over parameters for different types of carry-over in case of Balaam's design.

Carry-over Parameters ¹	Saturated carry-over	No placebo carry-over & self carry-over ²	No placebo carry-over ²	Steady-state carry-over	No carry-over
β_{AB}	no	0	0	no	0
β_{BB}	no	no	0	0	0
β_{BA}	no	no	no	no	0

¹ The parameters correspond to the regression coefficients in Equation (3), where β_{XY} represents the differential carry-over effect of the first treatment X on the second treatment Y, when compared with the carry-over effect in the AA sequence.

² For this type of carry-over, treatment A corresponds to the placebo.

person (between-subject) variance as compared with the error (within-subject) variance, the larger the intraclass correlation.

4. RELATIVE EFFICIENCIES UNDER DIFFERENT SCENARIOS OF CARRY-OVER

To compare the efficiencies of the designs when testing the treatment effect, we will consider the relative efficiency of two designs. Let $Var(\hat{\gamma}|\xi)$ denote the variance of the estimator of the treatment effect γ given a design ξ . Let ξ_1 and ξ_2 denote two different designs, say a parallel design and an AB/BA cross-over design, respectively. The relative efficiency of ξ_1 versus ξ_2 for the estimator of the treatment effect γ is defined as $RE_{\xi_1 vs \xi_2}(\hat{\gamma}) = Var(\hat{\gamma}|\xi_2) / Var(\hat{\gamma}|\xi_1)$. Because no closed-form expressions are available for the variances of the ML estimator of the treatment effect, in what follows, we will derive asymptotic variances of the ML estimator, and the resulting relative efficiencies are therefore asymptotic relative efficiencies.

4.1. Steady-state carry-over

First, we introduce some additional notation. Let n_p^*, n_{ep}^*, n_c^* and n_B^* denote the numbers of subjects that are recruited for a parallel design, an extended parallel design, a cross-over design and Balaam's design, respectively. Let p denote the dropout rate before the first measurement. After this dropout $n_p^*(1-p) = n_p, n_{ep}^*(1-p) = n_{ep}, n_c^*(1-p) = n_c$ and $n_B^*(1-p) = n_B$, subjects remain that have at least one observation and thus can be included in the analysis. As will be seen in what follows, because dropout before the first measurement affects all designs to the same extent, the dropout rate p does not enter into the relative efficiencies. Dropout may also occur after the first but before the second measurement. Such dropout does not occur in a parallel design, and its rate q , therefore, only affects the efficiencies of the two-period designs.

For steady-state carry-over there only is carry-over in treatment sequences where there is a switch of treatments. The corresponding asymptotic variances of the ML estimator of the treatment effect are derived in appendix A (available online as supporting information) and displayed in Table II. From these variances,

the following expressions for the asymptotic relative efficiencies result :

$$RE_{p vs c}(\hat{\beta}_1) = \frac{n_p}{n_c}, RE_{p vs B}(\hat{\beta}_1) = \frac{n_p}{n_B} \times \frac{2(1+\rho)}{(3-q)+\rho(1+q)} \text{ and } RE_{B vs ep}(\hat{\beta}_1) = \frac{n_B}{n_{ep}} \times \frac{(3-q)+\rho(1+q)}{2(2-q(1-\rho))}. \tag{4}$$

With equal numbers of subjects recruited for each design (i.e., $n_p^* = n_{ep}^* = n_c^* = n_B^*$ and thus also $n_p = n_{ep} = n_c = n_B$), we have $RE_{p vs c}(\hat{\beta}_1) = 1$. For differential carry-over, only the data from the first period of a cross-over trial can be used in estimating the treatment effect (e.g., [2]), and thus for identical sample sizes this design will be as efficient as a parallel trial. Furthermore, $RE_{p vs c}(\hat{\beta}_1) \leq 1$ and $RE_{B vs ep}(\hat{\beta}_1) \leq 1$: a parallel design and a cross-over trial are both less efficient than Balaam's design, and Balaam's design, in turn, is less efficient than an extended parallel design. The AA/BB design thus is most efficient (see Table III).

The degree to which the extended parallel design is more efficient than Balaam's design depends on the intraclass correlation ρ and the dropout rate q after the first measurement. The extended parallel design can be shown to be at most 1.33 times more efficient than Balaam's design, the efficiency gain becoming less the larger q and ρ become.

4.2. No placebo carry-over

If there is no carry-over in case the first treatment is a placebo and also no self carry-over, then $RE_{p vs c}(\hat{\beta}_1)$ is also given by Equation (4). Other relative efficiencies are (see Table II):

$$RE_{p vs B}(\hat{\beta}_1) = \frac{n_p}{n_B} \times \frac{3(1-\rho^2)}{(5-2q)-2(1-q)\rho-(2q+1)\rho^2} \text{ and } RE_{p vs ep}(\hat{\beta}_1) = \frac{n_p}{n_{ep}} \times \frac{1+\rho}{2-q(1-\rho)}. \tag{5}$$

With equal sample sizes (i.e., $n_p = n_{ep} = n_c = n_B$), we have $RE_{p vs c}(\hat{\beta}_1) = 1$ (see Equation (4)) and $RE_{p vs B}(\hat{\beta}_1) \leq 1$ and $RE_{p vs ep}(\hat{\beta}_1) \leq 1$ (see Equation (5)). So, similar to steady-state

Table II. Asymptotic variance of the ML estimator of the treatment effect for different designs and types of carry-over.

	Parallel design	Extended parallel design	Cross-over design	Balaam's design
Saturated carry-over ¹	not suitable	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_{ep}} \times \frac{1-q\rho^2}{1-q}$	not suitable	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_B} \times \frac{2-(q+1)\rho^2}{1-q}$
No placebo carry-over & self carry-over ¹	not suitable	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_{ep}} \times \frac{1-q\rho^2}{1-q}$	not suitable	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_B} \times \frac{5-2q+2(1-q)\rho-6\rho^2-2(1-q)\rho^3+(2q+1)\rho^4}{(1-q)(3-q)-2(1-q)\rho^2}$
No placebo carry-over	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_p}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_{ep}} \times \frac{(1+\rho)}{2-q(1-\rho)}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_c}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_B} \times \frac{3(1-\rho^2)}{5-2q-2(1-q)\rho-(2q+1)\rho^2}$
Steady-state carry-over	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_p}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_{ep}} \times \frac{(1+\rho)}{2-q(1-\rho)}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_c}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_B} \times \frac{2(1+\rho)}{(3-q)+\rho(1+q)}$
No carry-over	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_p}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_{ep}} \times \frac{(1+\rho)}{2-q(1-\rho)}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_c} \times \frac{(1-\rho)}{2-q(1+\rho)}$	$\frac{4(\sigma_0^2 + \sigma_\epsilon^2)}{n_B} \times \frac{(1-\rho^2)}{2-q(1+\rho^2)}$

¹For these types of carry-over interest is in the total effect of A versus B, that is, the direct effect plus carry-over effect onto a treatment itself.

Table III. Designs yielding the most efficient maximum likelihood estimator of the treatment effect ¹.

	Saturated carry-over ²	No placebo carry-over & self carry-over ^{2,3}	No placebo carry-over ³	Steady state carry-over	No carry-over
<i>Equal sample sizes</i>					
$\rho \leq 2 - \sqrt{3}$	EP	EP	EP	EP	CO
$2 - \sqrt{3} < \rho \leq 0.865$	EP	EP	BA	EP	CO
$0.865 < \rho$	EP	$\left\{ \begin{array}{l} q < q_k: \text{BA} \\ q_k \leq q < 1: \text{EP} \end{array} \right.$	BA	EP	CO
<i>Equal budgets:⁴</i>					
$\rho \leq 2 - \sqrt{3}$	EP	EP	$\left\{ \begin{array}{l} q < 1: \text{EP or PA} \\ q = 1: \text{PA} \\ q < 1: \text{BA or PA} \\ q = 1: \text{PA} \\ q < q_i: \text{BA} \\ q_i \leq q < 1: \text{BA or PA} \\ q = 1: \text{PA} \end{array} \right.$	$\left\{ \begin{array}{l} q < 1: \text{EP or PA} \\ q = 1: \text{PA} \end{array} \right.$	$\left\{ \begin{array}{l} q < \frac{2\rho}{1+\rho}: \text{CO} \\ \frac{2\rho}{1+\rho} \leq q < 1: \text{CO or PA} \\ q = 1: \text{PA} \end{array} \right.$
$2 - \sqrt{3} < \rho \leq 1/5(1 + \sqrt{6})$	EP	EP			
$1/5(1 + \sqrt{6}) < \rho \leq 0.865$	EP	EP			
$0.865 < \rho$	EP	$\left\{ \begin{array}{l} q < q_k: \text{BA} \\ q_k \leq q < 1: \text{EP} \end{array} \right.$			

¹ PA, Parallel design; EP, Extended parallel design; BA, Balaam's design; CO, AB/BA cross-over design.
² For these types of carry-over, interest is in the total effect of A versus B (i.e., direct effect + carry-over effect onto a treatment itself). If $q = 1$, then the total treatment effect cannot be estimated in either of the designs.
³ $q_k = \frac{(2\rho^3 + 3\rho^2 - 2\rho - 1 - (1 - \rho^2)\sqrt{8\rho^2 + 4\rho + 1})/2\rho^2; \quad q_i = \frac{(5\rho^2 - 2\rho - 1)}{2(\rho^2 - \rho + 1)}$.
⁴ It is assumed that the research costs in Equations (11)–(13) satisfy $c_{sp} \leq c_{sep} = c_{sc} = c_{sb} \leq 2c_{sp}$ and $c_{ts} = 0$.

carry-over, a parallel design and a cross-over trial are equally efficient, and are less efficient than Balaam's design and an extended parallel design. We, furthermore, have:

$$RE_{ep vs B}(\hat{\beta}_1) = \frac{n_{ep}}{n_B} \times \frac{3(1-\rho)(2-q(1-\rho))}{(5-2q)-2(1-q)\rho-(2q+1)\rho^2} \quad (6)$$

If $\rho \leq 2 - \sqrt{3}$ ($= 0.268$), we have $RE_{ep vs B}(\hat{\beta}_1) \geq 1$. If $q = 0$ and $\rho = 0$, the extended parallel design maximally improves the efficiency of Balaam's design, its efficiency being 1.2 times as large. If $\rho > 2 - \sqrt{3}$, then Balaam's design improves the extended parallel design. Examining several cross-over trials, in most cases ρ easily exceeds 0.30 (e.g., [1, p.17-20, p.69], [2, p. 268], [3, p.275, p.277], [15], [16, p.5]) and thus in many cases, Balaam's design will be most efficient for this type of carry-over. The gain in efficiency increases with a decreasing dropout rate q and an increasing ρ . Table III summarizes these results.

4.3. No differential carry-over

In case there is no differential carry-over, from the variances in Table II, it follows that

$$RE_{p vs ep}(\hat{\beta}_1) = \frac{n_p}{n_{ep}} \times \frac{(1+\rho)}{2-q(1-\rho)} \text{ and } RE_{ep vs B}(\hat{\beta}_1) = \frac{n_{ep}}{n_B} \times \frac{(2-q(1-\rho))(1-\rho)}{2-q(1+\rho^2)} \quad (7)$$

In case of equal sample sizes, we have $RE_{p vs ep}(\hat{\beta}_1) \leq 1$ and $RE_{ep vs B}(\hat{\beta}_1) \leq 1$. An extended parallel trial is at least as efficient as a parallel trial, and Balaam's design in turn is at least as efficient as an extended parallel trial. To examine whether a cross-over trial is more efficient than Balaam's design, we consider

$$RE_{c vs B}(\hat{\beta}_1) = \frac{n_c}{n_B} \times \frac{(2-q(1+\rho))(1+\rho)}{2-q(1+\rho^2)}, \quad (8)$$

which, for equal sample sizes, exceeds 1. The latter result was also established by Lasserre [17, p.913], however, assuming there was no dropout.

4.4. Saturated carry-over

For saturated carry-over, also self carry-over is involved, implying, as we argued before, that the total effect of A versus B, β_t , is of interest. This only can be estimated in designs having the AA and BB treatment sequences, so that a single asymptotic relative efficiency needs to be considered (see Table II):

$$RE_{ep vs B}(\hat{\beta}_t) = \frac{n_{ep}}{n_B} \times \frac{2-(1+q)\rho^2}{1-q\rho^2} \quad (9)$$

For $n_{ep} = n_B$, an extended parallel design is at least as efficient as Balaam's design. The extended parallel design is at most twice as efficient as Balaam's design (when $q = 1$ or $\rho = 0$).

4.5. No placebo carry-over with self carry-over

Because also for this type of carry-over, self carry-over is involved, we are interested in the total effect of A versus B. The AA/BB and Balaam’s design being the only candidates for estimating this effect, a single relative efficiency needs to be considered (see Table II):

$$RE_{ep\ vs\ B}(\hat{\beta}_t) = \frac{n_{ep}}{n_B} \times \frac{5 - 2q + 2(1-q)\rho - 6\rho^2 - 2(1-q)\rho^3 + (2q+1)\rho^4}{(3-q-2\rho^2)(1-q\rho^2)} \tag{10}$$

As long as the dropout rate q exceeds $(2\rho^3 + 3\rho^2 - 2\rho - 1 - (1 - \rho^2)\sqrt{8\rho^2 + 4\rho + 1})/2\rho^2$, we have $RE_{ep\ vs\ B}(\hat{\beta}_t) \geq 1$. Because $(2\rho^3 + 3\rho^2 - 2\rho - 1 - (1 - \rho^2)\sqrt{8\rho^2 + 4\rho + 1})/2\rho^2$ is negative for $\rho \leq 0.865$, this implies that for $\rho \leq 0.865$ the extended parallel design is most efficient. If $\rho > 0.865$, then also Balaam’s design can be most efficient.

When the budget is fixed, the number of subjects involved with each of the designs will not be the same. For a given budget the number of persons in the two-period designs will be smaller than the number of subjects in a parallel design, because of the costs of repeated measurements, prolonged participation of the subjects and possibly longer recruitment times [2]. Hence, the efficiency ordering of the designs will possibly be different when, instead of the number of subjects, the budgets are identical. This extension will be considered in the next section.

5. RELATIVE EFFICIENCIES GIVEN A FIXED BUDGET

First, we will specify a cost function. Let the costs involved with each subject in the parallel design be c_{sp} euros and the costs involved with each subject in an extended parallel design, a cross-over design and Balaam’s design be c_{sep} , c_{sc} and c_{sB} euros, respectively. These costs may represent the financial rewards given to subjects participating in a trial, but also the costs of recruiting a single subject. For treatment A and B there are 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_{ts} .

Commonly, the number of subjects is calculated such that a particular power in testing the intervention effect is guaranteed. To be sure of a certain power level, this number may be based on some pessimistic estimate of dropout, not necessarily equal to the actual dropout rates. On the other hand, we assume that the budget that is required to realize this power, is calculated assuming that none of the recruited subjects dropout during the study. Speculating on possible dropout, which would allow for lowering the budget given a particular sample size at the start of the trial, may bear the risk of exceeding the budget.

In case of equal allocation ratios for treatment A and B in a parallel design, having initially n_p^* subjects, the following budget C is then required:

$$C = 2c_{ts} + n_p^*c_{sp} + \frac{n_p^*}{2}c_A + \frac{n_p^*}{2}c_B + n_p^*c_T = 2c_{ts} + n_p^* \left(c_{sp} + \frac{c_\lambda}{2} \right), \tag{11}$$

where $c_\lambda = c_A + c_B + 2c_T$. For the designs that we consider, this budget function can be reparameterized such that it is the same as the cost function given by Yuan and Zhou [18], thereby generalizing the cost function proposed by Brown [19], also discussed by Berger and Wong [4].

For an AB/BA cross-over design, involving n_c^* recruited subjects and equal allocation ratios for each treatment sequence, noting that each subject receives both treatment A and B and is measured twice, the following budget is required:

$$C = 2c_{ts} + n_c^*c_{sc} + n_c^*c_A + n_c^*c_B + 2n_c^*c_T = 2c_{ts} + n_c^* (c_{sc} + c_\lambda). \tag{12}$$

In the case of equal allocation ratios, the required budgets for an AA/BB design and Balaam’s design, involving n_{ep}^* and n_B^* recruited subjects, respectively, are

$$\begin{aligned} C &= 2c_{ts} + n_{ep}^*c_{sep} + n_{ep}^*c_A + n_{ep}^*c_B + 2n_{ep}^*c_T \\ &= 2c_{ts} + n_{ep}^* (c_{sep} + c_\lambda), \text{ and} \\ C &= 4c_{ts} + n_B^*c_{sB} + n_B^*c_A + n_B^*c_B + 2n_B^*c_T = 4c_{ts} + n_B^* (c_{sB} + c_\lambda). \end{aligned} \tag{13}$$

Note that for the functions in Equations (11)–(13), the budget may simply be the total number of observations involved in a study by setting $c_T = 1$ and setting the other costs to 0. It could also represent the total number of subjects involved, by setting $c_{ts} = c_\lambda = 0$ and $c_{sp} = c_{sep} = c_{sc} = c_{sB} = 1$.

In what follows, we will assume that the subject-specific costs of the two-period designs are identical, that is, $c_{sep} = c_{sc} = c_{sB}$. Because subjects in these designs receive two treatments and a washout period may be involved, these costs very likely will be larger than those of a parallel design $c_{sep} \geq c_{sp}$, $c_{sc} \geq c_{sp}$ and $c_{sB} \geq c_{sp}$. It is also assumed that the subject specific costs for the two-period designs will never exceed two times the subject specific costs for the parallel designs c_{sep} , c_{sc} , $c_{sB} \leq 2c_{sp}$. Finally, administration costs per treatment sequence are assumed to be negligible compared with the other costs, so that $c_{ts} = 0$.

5.1. Steady-state carry-over

Assuming equal budgets, identical subject specific costs for the two-period designs, that is, $c_{sep} = c_{sc} = c_{sB}$, and $c_{ts} = 0$, it follows from Equations (11)–(13) that the sample sizes for these designs are the same. Consequently, the efficiency ordering of these designs equals the ordering established in section 4.1: An AA/BB design is more efficient than Balaam’s design, and Balaam’s design is more efficient than a cross-over design.

For equal budgets, the asymptotic relative efficiency of a parallel design versus a cross-over trial as given in Equation (4), becomes

$$RE_{p\ vs\ c}(\hat{\beta}_1) = \frac{c_\lambda + c_{sc}}{1/2c_\lambda + c_{sp}}. \tag{14}$$

Because very likely $c_{sc} \geq c_{sp}$, $RE_{p\ vs\ c}(\hat{\beta}_1) \geq 1$ and thus a parallel design is more efficient than a cross-over design. Rewriting the second expression in Equation (4) in terms of costs, gives

$$RE_{p\ vs\ B}(\hat{\beta}_1) = \left(\frac{c_\lambda + c_{sB}}{c_\lambda/2 + c_{sp}} \right) \times \frac{2(1+\rho)}{(3-q) + \rho(1+q)}. \tag{15}$$

From Equation (15) it follows that Balaam’s design is more efficient than a parallel trial whenever

$$q \leq \frac{(\rho + 3) - 2(\rho + 1)(c_\lambda + c_{sB}) / (c_\lambda/2 + c_{sp})}{(1 - \rho)}. \tag{16}$$

A reasonable upper bound for the subject-specific costs in Balaam's design may be twice the subject-specific costs in a parallel trial: $c_{SB} = 2c_{sp}$. Since the upper bound for the dropout rate in Equation (16) decreases in c_{SB} , $c_{SB} = 2c_{sp}$ yields the lowest upper bound $q \leq -(1 + 3\rho)/(1 - \rho)$, which, because $0 < \rho < 1$, is never true. So no lowest upper bound for q can be defined. To determine when a parallel design is more efficient than Balaam's design, we set $c_{SB} = c_{sp}$ to obtain as an upper lower bound for the dropout rate: $q \geq 1 - ((1 + \rho)c_\lambda) / ((1 - \rho)(0.5c_\lambda + c_{sp}))$. So, no upper lower bound for the dropout rate can be specified that only depends on the intraclass correlation, such that a parallel trial is more efficient than Balaam's design. It is, however, evident that for complete dropout after the first measurement (i.e., $q = 1$), a parallel design is at least as efficient as Balaam's design.

For the relative efficiency of a parallel design versus its extended version (see Table II)

$$RE_{p\text{ vs }ep}(\hat{\beta}_1) = \frac{c_{sep} + c_\lambda}{c_{sp} + c_\lambda/2} \times \frac{1 + \rho}{2 - q(1 - \rho)}, \quad (17)$$

we can show that also here there is no lower or upper bound for q , only depending on the intraclass correlation, such that one design is more efficient than the other. Table III summarizes these results.

5.2. No placebo carry-over

Assuming equal budgets, $c_{ts} = 0$ and identical subject specific costs for the two-period designs, the efficiency ordering of these designs equals the ordering in section 4.2. The extended parallel design and Balaam's design are more efficient than the AB/BA cross-over trial. If $\rho \leq 2 - \sqrt{3}$, then the extended parallel design is more efficient than Balaam's design, and if $\rho > 2 - \sqrt{3}$, then Balaam's design is most efficient.

For no placebo carry-over, $RE_{p\text{ vs }c}(\hat{\beta}_1)$ is also given by Equation (14), implying that a parallel design is more efficient than a cross-over design. The expression $RE_{p\text{ vs }ep}(\hat{\beta}_1)$ is given by Equation (17), implying that no boundaries on q can be derived, depending only on the intraclass correlation, such that one design is more efficient than the other. We also obtain (see Equation (5))

$$RE_{p\text{ vs }B}(\hat{\beta}_1) = \frac{c_{SB} + c_\lambda}{c_{sp} + c_\lambda/2} \times \frac{3(1 - \rho^2)}{(5 - 2q) - 2(1 - q)\rho - (2q + 1)\rho^2}, \quad (18)$$

implying that Balaam's design is more efficient than a parallel design whenever

$$q \leq \frac{(-\rho^2 - 2\rho + 5) - 3(1 - \rho^2)((c_{SB} + c_\lambda)/(c_{sp} + c_\lambda/2))}{2(\rho^2 - \rho + 1)}. \quad (19)$$

The upper bound for the dropout rate is decreasing in c_{SB} , and taking $c_{SB} = 2c_{sp}$ yields as the smallest upper bound, $q \leq (5\rho^2 - 2\rho - 1) / (2(\rho^2 - \rho + 1))$, which is positive when $\rho > 1/5(1 + \sqrt{6})$ ($= 0.690$). Hence, when $\rho > 1/5(1 + \sqrt{6})$ Balaam's design is more efficient than a parallel design provided $q \leq (5\rho^2 - 2\rho - 1) / (2(\rho^2 - \rho + 1))$. To examine when a parallel trial is more efficient than Balaam's design, we take $c_{SB} = c_{sp}$ to obtain an upper lower bound for the dropout rate. This upper lower bound still depends on the costs but is 1 at most. So if $q = 1$, a parallel design is at least as efficient as Balaam's design. The results for no placebo carry-over are summarized in Table III.

5.3. No differential carry-over

The sample sizes for the two-period designs are identical in case of equal budgets. Hence, the results of section 4.3 also apply: a cross-over trial is more efficient than Balaam's design, and Balaam's design is more efficient than an AA/BB design. The relative efficiency of a parallel design versus an extended parallel design is given by (see Equation (7))

$$RE_{p\text{ vs }ep}(\hat{\beta}_1) = \frac{c_{sep} + c_\lambda}{c_{sp} + c_\lambda/2} \times \frac{(1 + \rho)}{2 - q(1 - \rho)}, \quad (20)$$

implying that a parallel design is more efficient than an extended parallel design whenever

$$q \geq \frac{2 - (1 + \rho)(c_{sep} + c_\lambda)/(c_{sp} + c_\lambda/2)}{(1 - \rho)}. \quad (21)$$

Because this function decreases in c_{sep} , an upper lower bound for q is obtained if $c_{sep} = c_{sp}$. This upper lower bound still depends on the costs, but is 1 at most; for $q = 1$, a parallel design is at least as efficient as an extended parallel design. A lower upper bound for q is obtained if $c_{sep} = 2c_{sp}$, yielding $q \leq -2\rho/(1 - \rho)$, which is never true. Hence, no boundary on q can be given such that an extended parallel design is more efficient than a parallel design.

The relative efficiency of a parallel versus Balaam's design is given by (see Table II)

$$RE_{p\text{ vs }B}(\hat{\beta}_1) = \left(\frac{c_\lambda + c_{SB}}{0.5c_\lambda + c_{sp}} \right) \times \frac{(1 - \rho^2)}{2 - q(1 + \rho^2)}. \quad (22)$$

From Equation (22), Balaam's design is more efficient than a parallel design whenever

$$q \leq \frac{2 - (1 - \rho^2)(c_\lambda + c_{SB})/(c_\lambda/2 + c_{sp})}{(1 + \rho^2)}. \quad (23)$$

Because this function decreases in c_{SB} , the upper bound for q is lowest for $c_{SB} = 2c_{sp}$ yielding $q \leq 2\rho^2/(1 + \rho^2)$. When the dropout rate is smaller than this boundary, Balaam's design is more efficient than a parallel trial. Taking the lowest value for c_{SB} in Equation (23), $c_{SB} = c_{sp}$, yields the largest lower bound for the dropout rate such that a parallel design is more efficient than Balaam's design. This largest lower bound still depends on the research costs, but is 1 at most.

For a parallel design versus a cross-over trial, we have (see Table II)

$$RE_{p\text{ vs }c}(\hat{\beta}_1) = \left(\frac{c_\lambda + c_{sc}}{1/2c_\lambda + c_{sp}} \right) \times \frac{(1 - \rho)}{2 - q(1 + \rho)}. \quad (24)$$

This extends a result in Brown [19, p.73], where it is assumed that $q = 0$ and $c_{sc} = c_{sp}$. It can be shown that $RE_{p\text{ vs }c}(\hat{\beta}_1) \leq 1$ whenever

$$q \leq \frac{2 - (1 - \rho)(c_\lambda + c_{sc})/(c_\lambda/2 + c_{sp})}{(1 + \rho)}. \quad (25)$$

The boundary for the dropout rate q decreases as c_{sc} increases. Assuming that $c_{sc} \leq 2c_{sp}$, a lower upper bound for the dropout rate would then be $q \leq 2\rho/(1 + \rho)$. No simple guideline can be given as to when a parallel trial is more efficient than a cross-over trial. From Equation (25) and $c_{sc} = c_{sp}$, the dropout rate has to be $q \geq (c_\lambda\rho/(1 + \rho) + c_{sp}) / (0.5c_\lambda + c_{sp})$. So if $q = 1$, a parallel trial is more efficient than a cross-over trial.

5.4. Saturated carry-over

For saturated carry-over, also self carry-over is involved, implying that we are interested in the total effect of *A* versus *B* in the second period. The single relative efficiency that we need to consider is that of an extended parallel design versus Balaam's design (see Equation (9)). For equal budgets, equal subject specific costs ($c_{sep} = c_{sB}$) and $c_{ts} = 0$, both designs have the same sample size, implying that an extended parallel design is at least as efficient as Balaam's design (see section 4.4).

5.5. No placebo carry-over with self carry-over

Because self carry-over is involved, interest is in the total treatment effect and only the relative efficiency of an *AA/BB* design versus Balaam's design needs to be considered (see Equation (10)). Assuming identical subject specific costs and $c_{ts} = 0$, the sample sizes are equal and the results of section 4.5 apply. As long as $\rho \leq 0.865$, the *AA/BB* design is most efficient, but if $\rho > 0.865$ this only is true as long as the dropout rate q exceeds $(2\rho^3 + 3\rho^2 - 2\rho - 1 - (1 - \rho^2)\sqrt{8\rho^2 + 4\rho + 1}) / 2\rho^2$, otherwise Balaam's design is most efficient.

6. MAXIMIN DESIGNS

Choosing the most efficient design requires knowledge on the intraclass correlation ρ and the dropout rate q . In many cases, this a priori knowledge may be rather vague. Taking a Bayesian approach, one might choose the design that maximizes the power of the treatment test (cf. [20]) by integrating over prior distributions for the intraclass correlation and dropout rate. In this approach, the expected power of the treatment test is maximized.

Another solution is the maximin strategy [5], consisting of 2 steps: (1) for each design determine the minimum efficiency of the treatment effect estimator across the range of values for the intraclass correlation and the dropout rate, 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 that minimizes the maximum variance of the estimator of the treatment effect. In determining sample sizes, choosing values for the intraclass correlation and the dropout rate for which the variance is maximum, will guarantee the desired power level also for all other values of the intraclass correlation and the dropout rate. Choosing the maximin design implies the desired power level being guaranteed at the lowest costs.

From the asymptotic variances in Table II, one can derive that the maximum variance for each type of carry-over occurs at the highest dropout rates p and q , but not always at the highest intraclass correlation ρ . From an empirical overview of cross-over trials, it appears that $\rho > 0.12$, and in the majority of cases ρ easily exceeds 0.30 (e.g., [1, p.17-20, p.69], [2, p. 268], [3, p.275, p.277], [15], [16, p.5]). To be on the safe side, we assume ρ is between 0.10 and 1.0. Furthermore, the dropout rate q can vary to a large extent, but is at most 0.5 in a scenario with modest dropout and at most 0.9 in a scenario with large dropout. No set of values needs to be specified for the dropout rate p , because the maximum variance for each design is attained at the same (maximum) value for p and the relative efficiencies of designs do not depend on this common value for p (see Table II). Like before, we assume that

$c_{sp} \leq c_{sB} = c_{sc} = c_{sep}$, but we do not restrict $c_{ts} = 0$ and c_{sB} , c_{sc} , $c_{sep} \leq 2c_{sp}$.

Table IV shows which design minimizes the maximum (asymptotic) variance of the ML estimator of the treatment effect. The proofs are given in Appendix B (available online as supporting information). In case of equal sample sizes and no carry-over, a cross-over trial is maximin, whereas in case of steady-state carry-over, each of the four designs are maximin. For equal sample sizes and no placebo carry-over, only Balaam's design is maximin. For both saturated carry-over and no placebo carry-over with self carry-over, where interest is in the total treatment effect in the second period, the extended parallel design is the maximin choice.

In case of equal budgets and no carry-over, a cross-over trial is maximin if the sample ratio $n_p/n_c \leq 1.61$ (for $q \leq 0.5$) or $n_p/n_c \leq 1.12$ (for $q \leq 0.9$). Otherwise, a parallel design is maximin. For equal budgets and steady-state carry-over a parallel design is the maximin choice. For no placebo carry-over, a parallel design is maximin if the sample ratio $n_p/n_B \geq 1.29$ (for $q \leq 0.5$) or $n_p/n_B \geq 1.06$ (for $q \leq 0.9$), otherwise Balaam's design is maximin. Finally, for both carry-over types that include self carry-over the *AA/BB* design turns out to be maximin.

7. APPLICATION IN PLANNING A TRIAL

Suppose one would like to perform a randomized trial on the effectiveness of two drugs, say omeprazole and lansoprazole, among subjects suffering from heartburn, similar to Miner *et al.* [21]. After 4 days of treatment, one plans to evaluate the effect on 24-h gastric acid suppression (e.g., the percentage of time that the intragastric pH level exceeds 4). Similar to Miner *et al.* [21], one is not sure whether carry-over can be prevented, but is rather confident that steady-state will be reached within four days of treatment. Because this excludes self-carry over, this would mean that steady-state carry-over could occur. For such a scenario, Table III shows that, depending on the exact research costs, either a parallel or an extended parallel design is most efficient. The costs of the study are unknown, but suppose that $c_{sep} = 1.5c_{sp}$ and $1.5c_{\lambda} = c_{sep}$, then it can be derived from Equation (17) that a parallel trial is most efficient if $q > 1/3(1 - 5\rho)/(1 - \rho)$, which is always true if $\rho \geq 0.2$. Regardless of the research costs, for complete dropout at the second measurement, a parallel design is most efficient. If there is not enough information on ρ or q , then the maximin design could be chosen. For both a maximum dropout q of 0.5 and 0.9, the parallel design would be maximin (see Table IV).

If one would not be convinced that 4 days of treatment is enough to reach steady-state, self carry-over may occur. In such a scenario of saturated carry-over, interest is in the total treatment effect of omeprazole versus lansoprazole at the second measurement, implying that the extended parallel design would be the most efficient choice (see Table III).

8. CONCLUSION AND DISCUSSION

We examined the asymptotic efficiency of the ML estimator of the treatment effect for four two-treatment designs—a parallel, an extended parallel, a cross-over and Balaam's design. Assuming equal allocation ratios and equal dropout rates for the treatment sequences, and employing a flexible cost function, for different types of carry-over the most efficient designs were derived. If

Table IV. Maximin designs for the maximum likelihood estimator of the treatment effect^{1,2}.

	Saturated carry-over ³	No placebo & self carry-over ³	No placebo carry-over	Steady-state carry-over	No carry-over
$q \leq 0.5$					
Equal sample sizes	EP	EP	BA	PA/CO EP/BA	CO
Equal budgets: ⁴	EP	EP		PA	
$\left(\frac{n_p}{n_x}\right) \leq 1.29$			BA		CO
$1.29 < \left(\frac{n_p}{n_x}\right) \leq 1.61$			PA		CO
$1.61 < \left(\frac{n_p}{n_x}\right)$			PA		PA
$q \leq 0.9$					
Equal sample sizes	EP	EP	BA	PA/CO EP/BA	CO
Equal budgets: ⁴	EP	EP		PA	
$\left(\frac{n_p}{n_x}\right) \leq 1.06$			BA		CO
$1.06 < \left(\frac{n_p}{n_x}\right) \leq 1.12$			PA		CO
$1.12 < \left(\frac{n_p}{n_x}\right)$			PA		PA

¹ PA, Parallel design; EP, Extended parallel design; BA, Balaam's design; CO, AB/BA cross-over design.

² It is assumed that the research costs in Equations (11)–(13) satisfy $c_{SB} = c_{SC} = c_{sep} \geq c_{sp}$.

³ For these types of carry-over, the estimator of the total treatment effect (direct + carry-over effect onto treatment itself) is examined.

⁴ n_x represents n_c for no carry-over, and n_B for no placebo carry-over.

carry-over can be excluded or is negligible, a cross-over design is most efficient as long as the dropout rate at the second measurement does not exceed $2\rho/(1 + \rho)$. For larger dropout rates, either a cross-over trial or a parallel design is most efficient, this depending on the research costs involved. For steady-state carry-over either a parallel or an extended parallel design is most efficient, this depending on the dropout rate, intraclass correlation and research costs. For no placebo carry-over, the pattern is more complicated, in that either a parallel design, an extended parallel design or Balaam's design can be most efficient (see Table III for details). For all these types of carry-over, a parallel design is most efficient if there is complete dropout at the second measurement.

For saturated carry-over, all forms of carry-over are occurring, including self carry-over, which shifts the interest to the total treatment effect of A versus B (direct effect plus carry-over effect onto a treatment itself). This can only be tested in Balaam's design and an extended parallel design, the latter being most efficient. In case of no placebo carry-over combined with self carry-over, the extended parallel design is most efficient if the intraclass correlation is smaller than 0.865. For higher intraclass correlations, this depends on the dropout rate (see Table III).

If knowledge concerning the dropout rate or the intraclass correlation is too vague, a maximin design can be chosen. Maximin designs start from a worst case scenario, the maximum variance of the treatment estimator, and minimize this maximum variance. Assuming two different maxima for the dropout rates and a realistic range of intraclass correlations, it was shown that for steady-state carry-over a parallel design is maximin, whereas for no placebo carry-over, either Balaam's design or a parallel design is maximin, this depending on the research costs involved. If there is no carry-over either a cross-over design or a parallel design

is maximin, this, again, depending on the research costs. For saturated carry-over and no placebo carry-over with self carry-over, the extended parallel design is the maximin choice.

The variance of the maximum likelihood estimator of the treatment effect was derived, assuming the variance components are known. Asymptotically, these are also the variances in case of unknown variance components. For realistic sample sizes the variances of the fixed effect estimators may have a downward bias [3, p.49]. There are no firm rules, but simulation studies indicate that the bias can exceed 5% if the number of random effect categories relating to the random intercept variance is less than 10 and the intraclass correlation is less than 0.5 [3, p.75]. Hence, one should be cautious in applying these results to small sample studies.

The analysis models studied have no auto-correlation across time between the residual error terms. Adding such a parameter would make the analysis model unidentifiable for the two-period designs considered. However, to model the dependencies across time, instead of including a random intercept as in Equations (2) and (3), one could also let the errors be (auto)correlated. The resulting expressions for the variances of the treatment effect estimators are similar to those in Table II, the total variance being replaced by the residual error variance and the intraclass correlation being replaced by the autocorrelation. Provided that the autocorrelation is positive, the results of the present study immediately translate also to these auto-correlated error models.

If pre-randomization measurements of the outcome variable are available, these could be included as covariates in the analysis [1, 22]. This will lead to a reduction of the intercept variance and thus of the intraclass correlation [3]. Provided that the costs of pre-randomization measurements are the same for all designs

and there are no missing values on this variable, the results of the present study also apply. If there are missing values on pre-randomization covariates, this will lead to omitting all data of a person from the analysis (cf. [16]). With increasing rates of missing values it would then become more efficient to switch to an analysis, where pre-randomization measurements are considered as one of the outcome measurements (which does not lead to omitting complete cases). Deriving efficient designs then also involves considering the most efficient analysis model.

The present analysis model could be extended by allowing the outcome variances to differ across treatment conditions. This would occur if people react more differently to one treatment as opposed to the other. Further, as shown by Yuan and Zhou [18], given budget constraints Balaam's design is less efficient compared with designs involving three or four periods. Yuan and Zhou [18], however, did not take care of dropout, which will be larger the longer the trial lasts. To the extent that designs involving three or more periods are practically feasible, it would therefore be useful to compare these with the two-period designs of the present study, also allowing for dropout.

Acknowledgements

The author is grateful to two reviewers for their constructive and insightful comments.

REFERENCES

- [1] Jones B, Kenward MG. *Design and analysis of cross-over trials*. Chapman & Hall: London, 1989.
- [2] Senn S. *Cross-over Trials in Clinical Research*. Wiley: Chichester, 2002.
- [3] Brown H, Prescott R. *Applied Mixed Models in Medicine*. Wiley: Chichester, 2006.
- [4] Berger MPF, Wong WK. *An Introduction to Optimal Designs for Social and Biomedical Research*. Wiley: Chichester, 2009.
- [5] Atkinson AC, Donev AN, Tobias RD. *Optimum Experimental Designs, with SAS*. Oxford university press: Oxford, 2007.
- [6] Low JL, Lewis SM, Prescott P. Assessing robustness of crossover designs to subjects dropping out. *Statistics and Computing* 1999; **9**:219–227.
- [7] Grizzle JE. The two-period change over design in clinical trials. *Biometrics* 1965; **21**:467–480.
- [8] Freeman P. The performance of the two-stage analysis of two-treatment, two-period cross-over trials. *Statistics in Medicine* 1989; **8**:1421–1432.
- [9] Liang Y, Carriere KC. On the role of baseline measurements for crossover designs under the self and mixed carryover effects model. *Biometrics* 2010; **66**:140–148.
- [10] Putt ME. Power to detect clinically relevant carry-over in a series of cross-over studies. *Statistics in Medicine* 2006; **25**:2567–2586.
- [11] Jones B, Donev AN. Modelling and design of cross-over trials. *Statistics in Medicine* 1996; **15**:1433–1446.
- [12] Senn S, Lambrou D. Robust and realistic approaches to carry-over. *Statistics in Medicine* 1998; **17**:2849–2864.
- [13] Fleiss JL. A critique of recent research on the two-treatment cross-over design. *Controlled Clinical Trials* 1989; **10**:237–243.
- [14] Senn S. Cross-over trials in drug development: theory and practice. *Journal of Statistical Planning and Inference* 2001; **96**:29–40.
- [15] Garcia R, Benet M, Arnau C, Cobo E. Efficiency of the cross-over design: An empirical evaluation. *Statistics in Medicine* 2004; **23**:3773–3780.
- [16] Kenward MG, Roger JH. The use of baseline covariates in crossover studies. *Biostatistics* 2010; **11**:1–17.
- [17] Lasserre V. Determination of optimal designs using linear models in crossover trials. *Statistics in Medicine* 1991; **10**:909–924.
- [18] Yuan Y, Zhou J. Cost-efficient higher-order crossover designs for two-treatment clinical trials. *Pharmaceutical Statistics* 2005; **4**: 245–252.
- [19] Brown BW, Jr. The crossover experiment for clinical trials. *Biometrics* 1980; **36**:69–79.
- [20] Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. Wiley: Chichester, 2007.
- [21] Miner PB, Jr, McKean LA, Gibb RB, Erasalu GN, Ramsey DL, MCrorie JW. Omeprazole-Mg 20.6 mg is superior to lansoprazole 15 mg for control of gastric acid: a comparison of over-the-counter doses. *Alimentary Pharmacology & Therapeutics* 2010; **31**:846–851.
- [22] Senn S. Cross-over trials in Statistics in Medicine: The first 25 years. *Statistics in Medicine* 2006; **25**:3430–3442.