Open Journal of Statistics, 2017, 7, 305-322
http://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

# A Comparison of Within-Subjects and Between-Subjects Designs in Studies with Discrete-Time Survival Outcomes

**Maryam Safarkhani, Mirjam Moerbeek**

Department of Methodology and Statistics, Utrecht University, Utrecht, the Netherlands
Email: M.Safarkhani@uu.nl, M.Moerbeek@uu.nl

**How to cite this paper:** Safarkhani, M. and Moerbeek, M. (2017) A Comparison of Within-Subjects and Between-Subjects Designs in Studies with Discrete-Time Survival Outcomes. *Open Journal of Statistics*, **7**, 305-322.
https://doi.org/10.4236/ojs.2017.72023

**Received:** March 6, 2017
**Accepted:** April 23, 2017
**Published:** April 26, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

## Abstract

Crossover designs are well-known to have major advantages when comparing the effects of various non-curative treatments. We compare efficiencies of several crossover designs along with the Balaam's design with that of a parallel group design pertaining to longitudinal studies where event time can only be measured in discrete time intervals. With equally sized sequences, the parallel group design results in the greater efficiency if the number of time periods is small. However, the crossover and Balaam's designs tend to be more efficient as the study duration increases. The degree to which these designs add efficiency depends on the baseline hazard function and effect size. Additionally, we incorporate different cost considerations at the subject level when comparing the designs to determine the most cost-efficient design. Researchers might consider the crossover or Balaam's design more efficient if the duration of the study is long enough, especially if the costs of applying the baseline treatment are higher.

## Keywords

Balaam's Design, Crossover Design, Cost-Efficiency, Discrete-Time Survival Endpoint, Efficiency Measure, Parallel Group Design

## 1. Introduction

A well-known type of outcome in longitudinal studies is the survival endpoint with the research interest focused on the occurrence and timing of some events. The timing of events can be measured continuously using thin precise time units (e.g. minutes or days). Under some circumstances, however, it is far less feasible to measure the timing of events this precisely. Instead, it is measured discretely using a set of discrete intervals like years, months or weeks. Here, an event might

occur at any point in time during an interval but is only measured once at the end of each interval until event occurrence, drop-out or the end of the trial. Since the time of event occurrence is rounded upward to the nearest measurement time point, a loss of information will occur if the exact time is unknown. Data with this type of survival endpoints are called discrete-time survival data as opposed to data recorded on a continuous scale, *i.e.* continuous-time survival data.

It is useful to measure time discretely in retrospective studies where subjects can only supply event times in ranges or round numbers due to memory failure. In a suicide ideation study, for instance, subjects may not remember the exact day of their first suicidal thought, but may remember how old they were at the time. Discrete-time survival data are also encountered in prospective studies where it may not be feasible or practical to follow subjects continuously. In a smoking initiation study, for example, researchers are not able to contact subjects every day to record the onset of smoking, but may do so on a regular basis, say once a month. Another reason for measuring event occurrence in discrete time is if events can only occur at a few points in time, e.g. a student can graduate from college on a few occasions in the academic year.

Optimal designs and statistical power analysis have been important tools in designing longitudinal studies with a variety of types of outcomes. For trials with discrete-time survival outcomes, [1] [2] have recently studied the optimal combination of the number of subjects and measurements per subject to achieve a sufficient power at a minimal cost or to maximize the power level for a fixed budget. These papers solely focus on randomized controlled trials with a parallel group design with subjects receiving only a single treatment in the course of the trial. However, in studies evaluating a new and promising treatment, it may be considered unethical to not offer the treatment to some of the subjects as is done in parallel group trials. In addition, if subjects do not receive the treatment during the study, they are more likely to withdraw from the study [3]. Here, crossover designs are a more efficient option than parallel group designs for comparing the effect of treatments.

Crossover designs are powerful designs in bioequivalence, clinical and pharmaceutical trials if the disease is chronic and treatments have a reversible or non-curative effect. The major advantage of crossover over parallel group designs is that crossover designs eliminate part of the inter-subject variability from the treatment comparisons, and thus might require fewer subjects to provide the same level of power. For a discussion on the analysis of crossover designs with continuous and binary outcomes, see [4] [5] [6]. Moreover, the work of [7] discusses sample size determinations in crossover designs with binary outcomes in subjects measured until the end of the study even if they experience the event in an earlier period.

Crossover designs have rarely been applied with right-censored survival data [8]. Nevertheless, survival analysis can be used in crossover trials with survival outcomes. An example is a crossover study to compare disease-free survival among postmenopausal women with receptor-positive early breast cancer [9].

Another example is a two-period crossover design comparing atenolol with a combination of atenolol and nifedipine to treat angina pectoris [10]. In these studies, the time-to-event endpoint was measured continuously and a continuous-time survival model was used to analyze the data [11]. On rare occasions, parallel group designs have been compared with crossover designs in studies with continuous time-to-event outcomes [12]. However, to the authors' knowledge, no longitudinal studies at all have been conducted on this subject with discrete-time survival data.

The aim here is to determine whether and if so, to what extent, a crossover design is more efficient than a parallel group design with discrete-time survival outcomes if subjects are not further observed after experiencing the event. An example of a crossover design with discrete-time survival endpoints is a fertility study was conducted to study the effect of ovarian stimulation on increasing the chance of conception [13]. The outcome was timing of pregnancy that was recorded discretely after each treatment cycle. Here, a proportional odds model was used for analysis repeated crossover outcomes. We compare the designs' efficiencies for different numbers of time periods, allocation proportions to treatment sequences, baseline hazard probabilities and treatment effect sizes. We assume the main objective of the trials is to compare the treatments, and the best design is the one that provides an efficient estimate of the treatment differences.

We consider the most common $AB/BA$ crossover design where subjects switch to the other treatment after one time period. For practical purposes, other variations of this design are also considered such as the $AABB/BBAA$ design where subjects alternate the treatments after multiple time periods or applications of a given treatment. The efficiency of these designs are also compared with that of the Balaam's design [14], which is a combination of the crossover and parallel group designs. It is logical that these studies are affected by dropout if subjects leave the study permanently due to unforeseen reasons rather than event occurrence. We thus compare the designs with and without attrition.

The organization of this paper is as follows. In the next section, an overview of the logistic regression model for analysing discrete-time survival data is presented; see [15] for an extensive discussion of this model. This section is followed by an introduction of the various designs and the optimality criterion. Section 4 reports on the results. The comparison between different designs is illustrated with an example in Section 5. The final section presents the conclusions and discussion and gives suggestions for future work.

## 2. The Statistical Model

We consider designs with two treatments $A$ and $B$, and $s$ sequences of treatments. Let $N=\sum_{k=1}^{s} n_k$ be the total number of subjects in the design, with $n_k$ the number of subjects randomly assigned to sequence $k$ at baseline. The underlying continuous event times are recorded in discrete time intervals indexed by $j=1,2,\cdots,p$. These intervals represent a series of consecutive

periods in continuous time with equidistant cut points $t_0 = 0,\ t_1 = 1, \cdots, t_p = p$. The baseline measure is taken at time $t_0 = 0$, just before randomization, and the total duration of the follow-up of a study with $p$ periods ends at time $t_p = p$. Note that $t_0 = 0$ is the "beginning of time" when no one has experienced the event yet but everyone is eligible to do so. The first measurement of event occurrence is taken at time point $t_1$, and any event occurring after $t_0$ and before $t_1$ is classified as happening during the first time interval $[t_0, t_1)$. The $j$ th time interval $[t_{j-1}, t_j)(j = 2, 3, \cdots, p)$ begins immediately at time point $t_{j-1}$ and ends just before time $t_j$.

The binary response $Y_{ijk}$ for subject $i(i = 1, 2, \cdots, n_k)$ in period $j(j = 1, 2, \cdots, p)$ receiving sequence $k(k = 1, 2, \cdots, s)$ is measured once at the end of each time interval and defined according to whether the subject experiences the event of interest $(Y_{ijk} = 1)$ or not $(Y_{ijk} = 0)$. The expected value and variance of $Y_{ijk}$ are $E(Y_{ijk}) = h_{d[j,k]}(t_j)$ and $\operatorname{var}(Y_{ijk}) = h_{d[j,k]}(t_j)\left[1 - h_{d[j,k]}(t_j)\right]$ where $h_{d[j,k]}(t_j)$ is the discrete-time hazard probability that the subject experiences the event in period $j$ under treatment $d[j,k]$. It is given by $h_{d[j,k]}(t_j) = \Pr\left(Y_{ijk} = 1 \middle| Y_{ij'k} = 0 \text{ for } j' < j\right)$. It is the conditional probability that an event occurs in interval $j$ under sequence $k$ for subject $i$ given that the event has not yet occurred before period $j$. Note that $d[j,k] = A$ or $B$ denotes the treatment used for subject $i$ in period $j$ under sequence $k$. Note that repeated measurements per subject are considered to be conditionally independent.

The discrete-time hazard probability $h_{d[j,k]}(t_j)$ for subject $i$ assigned to sequence $k$ in period $j$ is modeled as:

$$\operatorname{logit} h_{d[j,k]}(t_j) = \log \frac{h_{d[j,k]}(t_j)}{1 - h_{d[j,k]}(t_j)} = \sum_{j=1}^{p} \alpha_j D_{ijk} + \beta Z_{ijk}, \tag{1}$$

with the time-dependent explanatory variable $Z_{ijk}$ denoting the treatment condition and $Z_{ijk} = 1$ if the subject receives treatment $B$, and $= 0$ otherwise. For a given time period, the parameter $\beta$ denotes the effect of treatment $B$ relative to treatment $A$ on the probability of event occurrence on the logit scale, so $\beta = \log \dfrac{h_B(t_j) / \left[1 - h_B(t_j)\right]}{h_A(t_j) / \left[1 - h_A(t_j)\right]}$. As can be seen, the parameter $\beta$ is constant across time. We assume that model (1) is a proportional odds model. The dummy variable $D_{ijk}$ is set to 1 in time interval $j$ and 0 elsewhere. The corresponding intercept parameter $\alpha_j$ is the value of the logit hazard probability corresponding to treatment $A$ in that particular time period so $\alpha_j = \log \dfrac{h_A(t_j)}{\left[1 - h_A(t_j)\right]}$ for $j = 1, 2, \cdots, p$.

Model (1) can be formulated in matrix form as:

$$\operatorname{logit} \boldsymbol{h}(\boldsymbol{t}) = \boldsymbol{X\theta},$$

where the vector $\boldsymbol{h}(\boldsymbol{t})$ contains discrete-time hazard probabilities of event

occurrence for all $p$ time periods and all $N$ subjects until they experience the event or leave the study before event occurrence or the study concludes (*i.e.*, if $j = p$). The parameter vector $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \cdots, \alpha_p, \beta)'$ is a column vector of $(p+1)$ unknown parameters. The design matrix $\boldsymbol{X}$ is of order $\sum_{k=1}^{s}\sum_{j=1}^{p} n_{jk} \times (p+1)$, with $n_{jk}$ representing the number of subjects in the $k$ th sequence entering the $j$ th period and leave the study neither due to event occurrence nor unforeseen reasons prior to time period $j$. The total number of subjects at the beginning of the study in sequence $k$ is $n_{1k} = n_k$, and the total number of subjects entering period $j \geq 2$ is

$$n_{jk} = n_{(j-1)k}\left[1-\hat{h}_{d[j-1,k]}(t_{j-1})\right]\left[1-r_k(t_{j-1})\right] = n_k \hat{S}_{d[j-1,k]}(t_{j-1})\prod_{h=1}^{j-1}\left[1-r_k(t_h)\right].$$

Here, $\hat{h}_{d[j,k]}(t_j)$ is the estimate of the discrete-time hazard probability, and $\hat{S}_{d[j,k]}(t_j)$ is the estimate of the probability of the subject will experience the event after time $t_j$. The notation $d[j-1,k]$ refers to the treatment in the preceding period, *i.e.* the $(j-1)$th time period. It can be concluded that the risk of event occurrence in period $j$ depends on the survival probability then and in the previous period using $h_{d[j,k]}(t_j) = \left[S_{d[j,k]}(t_{j-1}) - S_{d[j,k]}(t_j)\right] / S_{d[j,k]}(t_{j-1})$. What is more, the attrition rate $r_k(t_h)$ denotes the proportion of subjects in sequence $k$ who leaves the study during time period $h$ due to reasons other than event occurrence. In this study, we assume a constant attrition rate across all the time periods and treatment sequences, *i.e.* $r_k(t_h) = r$ for any $h \in \{1, 2, \cdots, p\}$ and $k \in \{1, 2, \cdots, s\}$. We assume non-informative attrition (*i.e.* missing at random), that is the non-censored subjects do not differ systematically from the censored subjects. This means those who remain in the study are representative of everyone who would have remained in the study had there been no censoring.

The common method for estimating the vector of unknown parameters $\boldsymbol{\theta}$ is iteratively re-weighted least squares [16]. The asymptotic variance-covariance matrix of the estimator $\hat{\boldsymbol{\theta}}$ has the form:

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}}) = \left[\sum_{k=1}^{s}\sum_{j=1}^{p} \boldsymbol{X}'_{jk}\hat{w}_{d[j,k]}(t_j)\boldsymbol{X}_{jk} n_{jk}\right]^{-1}. \tag{2}$$

The vector $\boldsymbol{X}_{jk}$ corresponds to subjects in the $j$ th time interval in the $k$ th sequence, and has $(p+1)$ elements with value 1 on the $j$ th element, value 0 or 1 on the $(p+1)$th element, and 0 elsewhere. So the first $p$ elements represent the values on the dummies $D_1, D_2, \cdots, D_p$, and the $(p+1)$th element represents the value of $Z_{jk}$. The scalar $\hat{w}_{d[j,k]}(t_j)$ is the least squares weight for subjects in period $j$ under sequence $k$. For a logit link function, it is given as $\hat{w}_{d[j,k]}(t_j) = \hat{h}_{d[j,k]}(t_j)\left[1-\hat{h}_{d[j,k]}(t_j)\right]$. It should be noted that the $(p+1, p+1)$th entry of $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}})$ is proportional to the variance of the estimator of the treatment difference (*i.e.*, $\beta$) and it will be used for the definition of the optimal designs.

## 3. Crossover Designs and Efficiencies

We consider trials with a maximum duration of $p_{\max} = 12$ time periods where

subjects may be observed over $p = 1, 2, \cdots,$ or $p_{\max}$ time periods. So $p_{\max}$ is the maximum number of time periods a trial can be conducted in and $p$ is the number of time periods at hand. Note that the larger $p$ is, the longer the duration of the follow-up. For easy comparison of the hazard probabilities in each period, we assume the time points are equally spaced and the distance between any pair of adjacent time points is fixed in advance. Under this assumption, the duration of a trial with $p_{\max} = 12$ time periods is twice as long as the duration of a trial with $p = 6$ time periods. Table 1 presents designs for studies with $p_{\max} = 12$ time periods.

The first design is the parallel group (PG) design where subjects are randomly assigned to sequences with a single treatment *A* or *B*. The other three designs are crossover (CO) designs where subjects receive treatments *A* and *B* according to a pre-established order during the study, but switch to a different treatment after one or more multiple periods of using a given treatment. In the CO1 design, subjects use the two treatments sequentially for fixed periods of time, and switch to treatment *A* or *B* after one time period. In the CO3 design, subjects alternate the use of the two treatments after three applications of a given treatment so the switching time point is three. The PG and CO designs are based on two treatment sequences in which some part of the subjects are randomly assigned to the first sequence and the remaining subjects to the second sequence. The last design is the Balaam's (BM) design, a four-treatment-sequences design that assigns some parts of the subjects to the (*AB*/*BA*) sequence and the remainder to the (*AA*/*BB*) sequence. This design may be considered a combination of the PG and CO1 designs. It should be noted that we compare the designs of equal total duration or follow-up time $(p)$ and sample size at baseline $(N)$.

We study the efficiency of the PG design compared with that of an alternative design to determine which design estimates the parameter $\beta$ more efficiently.

**Table 1.** 2-treatment and 12-period designs.

| Design | Sequence | Period | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| PG | 1 | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* |
| | 2 | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* |
| CO1 | 1 | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* |
| | 2 | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* |
| CO3 | 1 | *A* | *A* | *A* | *B* | *B* | *B* | *A* | *A* | *A* | *B* | *B* | *B* |
| | 2 | *B* | *B* | *B* | *A* | *A* | *A* | *B* | *B* | *B* | *A* | *A* | *A* |
| CO6 | 1 | *A* | *A* | *A* | *A* | *A* | *A* | *B* | *B* | *B* | *B* | *B* | *B* |
| | 2 | *B* | *B* | *B* | *B* | *B* | *B* | *A* | *A* | *A* | *A* | *A* | *A* |
| BM | 1 | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* |
| | 2 | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* |
| | 3 | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* | *A* |
| | 4 | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* | *B* |

To do so, we consider the PG design as the reference design and compare the performance of the other designs using the relative efficiency (RE):

$$\mathrm{RE}_{\mathrm{CO|PG}} = \frac{\mathrm{var}_{\mathrm{PG}}\left(\hat{\beta}\right)}{\mathrm{var}_{\mathrm{CO}}\left(\hat{\beta}\right)}, \quad \mathrm{RE}_{\mathrm{BM|PG}} = \frac{\mathrm{var}_{\mathrm{PG}}\left(\hat{\beta}\right)}{\mathrm{var}_{\mathrm{BM}}\left(\hat{\beta}\right)}.$$

If $\mathrm{RE}_{\mathrm{CO|PG}} = 1$, the CO design is as efficient as the PG design. If $\mathrm{RE}_{\mathrm{CO|PG}} < 1$, the PG design is more efficient, and if $\mathrm{RE}_{\mathrm{CO|PG}} > 1$, the PG design is less efficient than the CO design. In addition, the $\mathrm{RE}_{\mathrm{CO|PG}}^{-1}$ indicates how many subjects should be taken under the CO design to be as efficient as the PG design [6] [17]. For example, if $\mathrm{RE}_{\mathrm{CO|PG}} = 0.8$, then $\left(0.8^{-1} - 1\right) \times 100\% = 25\%$ more subjects are required under the CO design to have the same efficiency as under the PG design. The interpretation of $\mathrm{RE}_{\mathrm{BM|PG}}$ is similar.

## 4. Results

We assume the probability of event occurrence for treatment $A$ does not vary across the time intervals, so $h_A\left(t_j\right) = h_A$ for $j = 1, 2, \cdots, p$. Since finding a closed-form formula for the variance-covariance matrix in (2) is complicated, the results are presented for selected choices of $h_A$ and the difference between the probabilities of treatments $A$ and $B$ $\left(\delta = h_B - h_A\right)$. We study the efficiency and cost efficiency of a PG design in comparison to three alternative CO designs, namely CO1, CO3, and CO6, along with the BM design.

Figure 1 presents the REs on the vertical axis as a function of the number of time periods $p$ on the horizontal axis for various values of $h_A$ (rows in matrix of graphs) and $\delta$ (columns in matrix of graphs). These selected values of $h_A$ and $\delta$ result in a maximum difference of 50% in the survival probabilities between two treatment sequences by the end of a study with the maximum duration if a PG design is conducted. Here, the total number of subjects in each design $(N)$ is equally divided over the treatment sequence groups. Each group in the PG and CO designs contains $\frac{N}{2}$ subjects and in the BM design, each group contains $\frac{N}{4}$ subjects.

As can be seen in Figure 1, all the designs are equally efficient if $p = 1$ since all the designs are the same in this case (see Table 1). In addition, the CO3 design is as efficient as the PG design if $p \le 3$ and the CO6 design is as efficient as the PG design if $p \le 6$. We also observe that the REs of the CO and the BM designs generally decrease from unity as $p$ increases from 1 and the size of the decrease depends to some extent on the $h_A$ and $\delta$ values. The decrease is smaller if $h_A$ is larger for a given $\delta$, and for a given $h_A$, it is larger with a larger $\delta$. However, at some value of $p$, the REs start to increase and approach unity if $p$ increases further. They may exceed unity if $p$ becomes even larger. The CO and BM designs are thus less efficient than the PG design if $p$ is small, though the designs may become more efficient than the PG design if the duration of the trial is large enough. Of all the designs, we observe that the BM

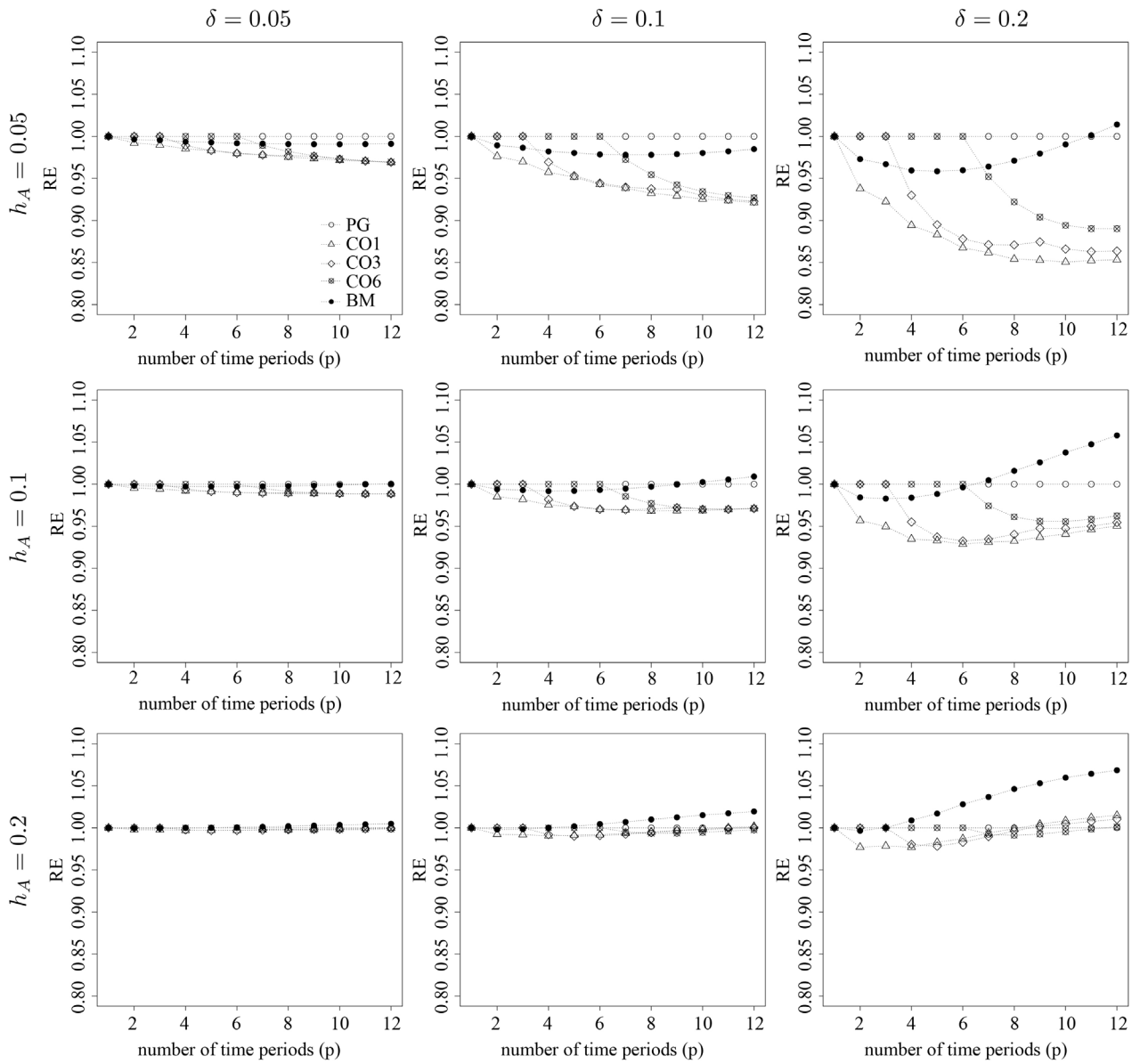


**Figure 1.** Efficiency of selected designs with two treatments $A$ and $B$ in comparison with the PG design as a function of the number of time periods $p$ for various $h_A$ and $\delta$ given equal sample sizes when treatment sequences are equally sized.

design more often tends to become more efficient than the PG design for a larger $p$ than the CO designs. Figure 1 also shows that a more extreme result is given if $\delta$ becomes larger for a given $h_A$ or $h_A$ becomes smaller for a given $\delta$.

We observe almost similar results if a constant attrition rate $r$ is taken into account. The only difference is that the REs approach unity more gradually as $p$ increases if $5\%$ or $10\%$ of the subjects are lost to follow-up in each period within each sequence, implying that the CO and BM designs require more time periods to be equal or more efficient than the PG design (results not shown).

We now look for the time point when the CO1 and BM designs are as efficient as the PG design. We have $\tilde{p}$ denote the smallest number of time points when $\mathrm{RE}_{\mathrm{CO1|PG}} > 1$ or $\mathrm{RE}_{\mathrm{BM|PG}} > 1$. For computational and practical reasons, we limit

our search to sixty periods. Table 2 presents the value of $\tilde{p}$ for various combinations of $h_A$, $\delta$ and $r$. For the BM design, we observe that for a given $\delta$, $\tilde{p}$ decreases as $h_A$ increases, and similarly it decreases as $\delta$ increases for a given $h_A$. In addition, the decrease in $\tilde{p}$ accompanying an increase in $h_A$ is larger with a smaller $\delta$. Likewise, the decrease in $\tilde{p}$ with an increase in $\delta$ is larger if $h_A$ is smaller. We note that the same effect of $\delta$ on $\tilde{p}$ is not observed for the CO1 design if $h_A = 0.1$. For each combination of $h_A$ and $\delta$, the CO1 design requires a longer study duration than the BM design to become more efficient than the PG design. The table also shows that if 5% of the subjects drop out of the study in each period, only a few if any more periods are required for the CO1 and BM designs in comparison with the case of no attrition, and if $r$ increases further, the designs need to be expanded to include even more periods (results not shown). Lastly, we emphasize that the value of $\tilde{p}$ for the CO3 design is almost similar to that of the CO1 design, but the study duration for the CO6 design needs to be extended for one to three more time periods (results not shown).

## A Cost-Efficiency Comparison between Designs

Previous section shows a pair-wise comparison of efficiency of the designs for studies with 1 to 12 time periods. In such a comparison, we did not make a distinction between the costs of sampling subjects and the costs of treating and measuring them. However, if recruiting a subject costs differently than taking measurements from that subject and the cost of treating this subject with treatment *A* is different than that with treatment *B*, we should account for the cost differential when we compare the five types of the designs. To this end, we take two cost functions as a function of the number of time periods for each type of the designs into account. Let $c_0$ represent the initial cost for setting up a study. If $c_1$ denotes the cost to include a subject in the study, let the cost of taking one measurement be denoted by $c_2$. If $c_A$ denotes the cost to treat a subject with treatment *A*, $c_B$ denotes the cost to treat a subject with treatment *B*. Cost function I is then computed for a study with $p$ time periods, $s$

Table 2. The number of time periods $\tilde{p} \in \{2,3,\cdots,59,60\}$ at which a BM or CO1 design is equally efficient as a PG design with several $h_A$, $\delta$ and $r$ values when treatment sequences are equally sized.

| Study designs | | $r = 0.00$ | | | $r = 0.05$ | | |
|---|---|---|---|---|---|---|---|
| | | $\delta = 0.05$ | $\delta = 0.1$ | $\delta = 0.2$ | $\delta = 0.05$ | $\delta = 0.1$ | $\delta = 0.2$ |
| BM | $h_A = 0.05$ | 22 | 17 | 11 | 28 | 20 | 12 |
| BM | $h_A = 0.1$ | 12 | 10 | 7 | 13 | 10 | 7 |
| BM | $h_A = 0.2$ | 5 | 4 | 3 | 5 | 5 | 4 |
| CO1 | $h_A = 0.05$ | >60 | >60 | >60 | >60 | >60 | >60 |
| CO1 | $h_A = 0.1$ | 37 | 48 | >60 | >60 | >60 | >60 |
| CO1 | $h_A = 0.2$ | 13 | 11 | 9 | 17 | 14 | 10 |

treatment sequences, and $N$ subjects at baseline as follows:

$$C - c_0 = Nc_1 + Nc_2 \sum_{k=1}^{s}\sum_{j=0}^{p} S_{d[j,k]}\left(t_j\right) + \frac{N}{2}\sum_{k=1}^{s}\sum_{j=0}^{p-1} c_{d[j,k]} S_{d[j,k]}\left(t_j\right). \quad (3)$$

With this cost function, we assume that subjects leave the study once they have experienced the event and measurements are not taken after event occurrence. $S_{d[j,k]}\left(t_j\right)$ is the survival function for the treatment $d\left[j,k\right]$ under sequence $k$ by the end of time period $j$ and $t_0 = 0$ is the baseline. Therefore, the number of measurements (including one baseline measurement) for each subject in sequence $k$ is given by $1 + \sum_{j=1}^{p} S_{d[j,k]}\left(t_j\right)$. Note that subjects are treated by treatment $d\left[j,k\right] = A$ or $B$ at the beginning of each time period, and so the number of treatment applications for a subject in sequence $k$ is $1 + \sum_{j=1}^{p-1} S_{d[j,k]}\left(t_j\right)$.

If it is assumed that subjects leave the study after experiencing the event or due to reasons other than event occurrence, the number of measures for each subject is given by $1 + \sum_{j=1}^{p} S_{d[j,k]}\left(t_j\right)\left(1-r\right)^j$. Cost function II can then be represented as follows:

$$C - c_0 = Nc_1 + Nc_2 \sum_{k=1}^{s}\sum_{j=0}^{p} S_{d[j,k]}\left(t_j\right)\left(1-r\right)^j + \frac{N}{2}\sum_{k=1}^{s}\sum_{j=0}^{p-1} c_{d[j,k]} S_{d[j,k]}\left(t_j\right)\left(1-r\right)^j.$$

To determine the most cost-efficient design for a given number of time periods, we normalize the optimality criterion (*i.e.* $\mathrm{var}\left(\hat{\beta}\right)$) by multiplying it by the cost $C - c_0$. In other words, we compare the designs based on:

$$\mathrm{RE}_{\mathrm{CO|PG}} = \frac{\mathrm{var}_{\mathrm{PG}}\left(\hat{\beta}\right)}{\mathrm{var}_{\mathrm{CO}}\left(\hat{\beta}\right)} \times \frac{C_{\mathrm{PG}} - c_0}{C_{\mathrm{CO}} - c_0}, \qquad \mathrm{RE}_{\mathrm{BM|PG}} = \frac{\mathrm{var}_{\mathrm{PG}}\left(\hat{\beta}\right)}{\mathrm{var}_{\mathrm{BM}}\left(\hat{\beta}\right)} \times \frac{C_{\mathrm{PG}} - c_0}{C_{\mathrm{BM}} - c_0}.$$

In this case, $\mathrm{var}\left(\hat{\beta}\right)$ under each design is penalised by the amount of costs of that design which accounts for the number of time periods and for different costs of treatment *A* and *B*. We compare the designs for different combinations of the costs at the subject-level (*i.e.* $c_1$, $c_A$ and $c_B$). The costs at the measurement level is fixed to $c_2 = 1$.

Figure 2 presents the REs plots as a function of $p$ for deigns with an equal allocation proportion for three combinations of $h_A$ and $\delta$ (columns in matrix of graphs). We consider three different combinations for the costs $c_1$, $c_A$, and $c_B$ (rows in matrix of graphs). It should be mentioned that in all three cases the costs at the subject-level are higher than the cost at the measurement-level (*i.e.* $c_2 < c_1$). The first combination $c_1 < c_A < c_B$ corresponds to studies where treating subjects is more expensive than sampling subjects and the costs to treat a subject with treatment *B* are high in relation to the costs to treat the subject with treatment *A*. The second combination $c_1 = c_A = c_B$ represents studies where all the three costs at the subject-level are equal. Finally, the last combination $c_1 < c_B < c_A$ represents a reverse scenario compared to the first combination where application of treatment *B* is less expensive than treatment *A*.

Figure 2 shows that when adjusting for design cost, the PG design is often a

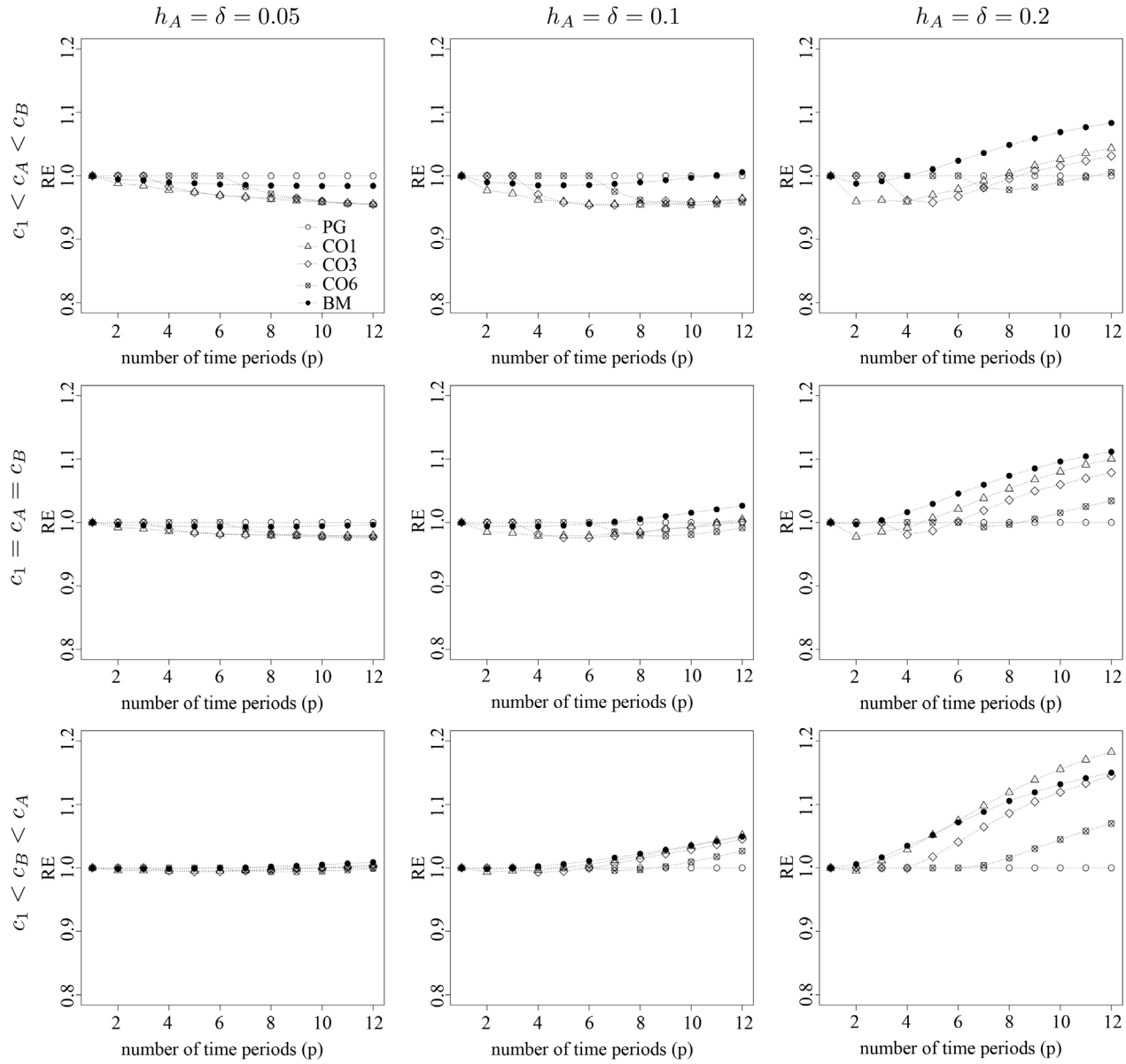


**Figure 2.** Efficiency of selected designs with two treatments $A$ and $B$ in comparison with the PG design as a function of the design allocation proportion $\pi$ for various $h_A$ and $\delta$ and the cost ratios $c_1$, $c_A$ and $c_B$ using cost function I in Equation (3) $(r = 0)$ when treatment sequences are equally sized.

more efficient choice when treating subjects costs more than recruiting the subjects and also when treatment $A$ is less costly than treatment $B$ (*i.e.* $c_1 < c_A < c_B$). The efficiency of the other alternative designs, however, tend to increase as $p$ becomes larger so that the BM design becomes more efficient than the PG for $p \in \{5, 6, \cdots, 12\}$ when $h_A = \delta = 0.2$. The CO designs, on the other hand, need to be conducted for a longer time to become more efficient than the PG design; the CO1 and CO3 designs become more efficient for $p \geq 9$ when $h_A = \delta = 0.2$. When all the subject-level costs are equal (*i.e.* $c_1 = c_A = c_B$), the efficiencies of the CO designs and the BM design is very close to that of the PG design, and the efficiency of these designs tend more often to exceed unity as

$h_A$ and $\delta$ become larger. In the last scenario, when sampling subjects is less costly than treating subjects and the cost of treatment $B$ is lower relative to the cost of treatment $A$, the CO and BM designs are most often more efficient than the PG design. In this case, the CO and BM designs are always preferable. Overall, we observed very similar results to those of **Figure 2** when comparing the design efficiencies using cost function II $(r > 0)$. However, the increase in the efficiencies with increasing $p$ becomes smaller in cases where $r \neq 0$ (results not shown).

Up until now we have focused on equal allocation proportions for each of the treatment sequences of the designs. From a clinical or ethical point of view, there might be reasons for an unequal assignment of subjects. For the BM design, for example, it might be considered unethical to give subjects the same treatment multiple times if its efficacy is unknown [18]. In this paper, we define $\pi \in [0,1]$ as the design allocation proportion and assume for a given $\pi$ that the CO and PG designs randomly allocate $\pi N$ subjects to the first treatment sequence and the BM design randomly allocates $\pi \frac{N}{2}$ to the ($AB/BA$) sequence. We now compare the designs efficiency as a function of $\pi$ for a given $p$. We limit our search to $\pi \in [0.25, 0.75]$ and assume that the treatment sequences contain at least a quarter of the subjects. **Figure 3** depicts efficiency comparisons across the designs as a function of $\pi$ under the same condition as **Figure 2** if $p = 12$.

We first focus on the results for $c_1 < c_A < c_B$ which implies treatment $B$ is more costly than treatment $A$. When $h_A = \delta = 0.05$, the REs of the CO designs are larger than unity if $\pi$ is small; the designs are almost equally efficient if $\pi \in [0.40, 0.45]$, and the PG design is more efficient than the CO designs as $\pi$ increases further. If $h_A$ and $\delta$ increase to 0.2, the REs of the CO designs become closer to that of the PG design which implies the effect of $\pi$ becomes smaller and thus a negligible gain in efficiency from any CO or PG design is obtained. As can be seen, the RE line of the BM design has almost a similar $\cup$-shape. This makes sense since the BM design is a compromised design between the CO1 and the PG designs when $\pi = 0.5$. For a smaller $\pi$, the BM design allocates more subjects to the sequences of the PG design and therefore its efficiency is in favor of the efficiency of the PG design for a small $\pi$. Similarly, the BM design allocates more subjects to the sequences of the CO1 design for a large $\pi$ and therefore its efficiency is in favor of the efficiency of the CO1 in this case. So the BM design tends to be more efficient than the PG design as the sequence sizes become more unequal.

As treatment $B$ becomes as expensive as or less expensive than treatment $A$ (*i.e.* $c_1 = c_A = c_B$ or $c_1 < c_B < c_A$), a higher efficiency is most often maintained by the CO designs compared to the PG design as $\pi$ becomes larger, especially if $h_A = \delta = 0.2$. In this case, it is seen that the BM design becomes even more efficient among all the designs as $\pi$ increases.

## 5. Example

In the introduction, an example of a CO design with right-censored survival

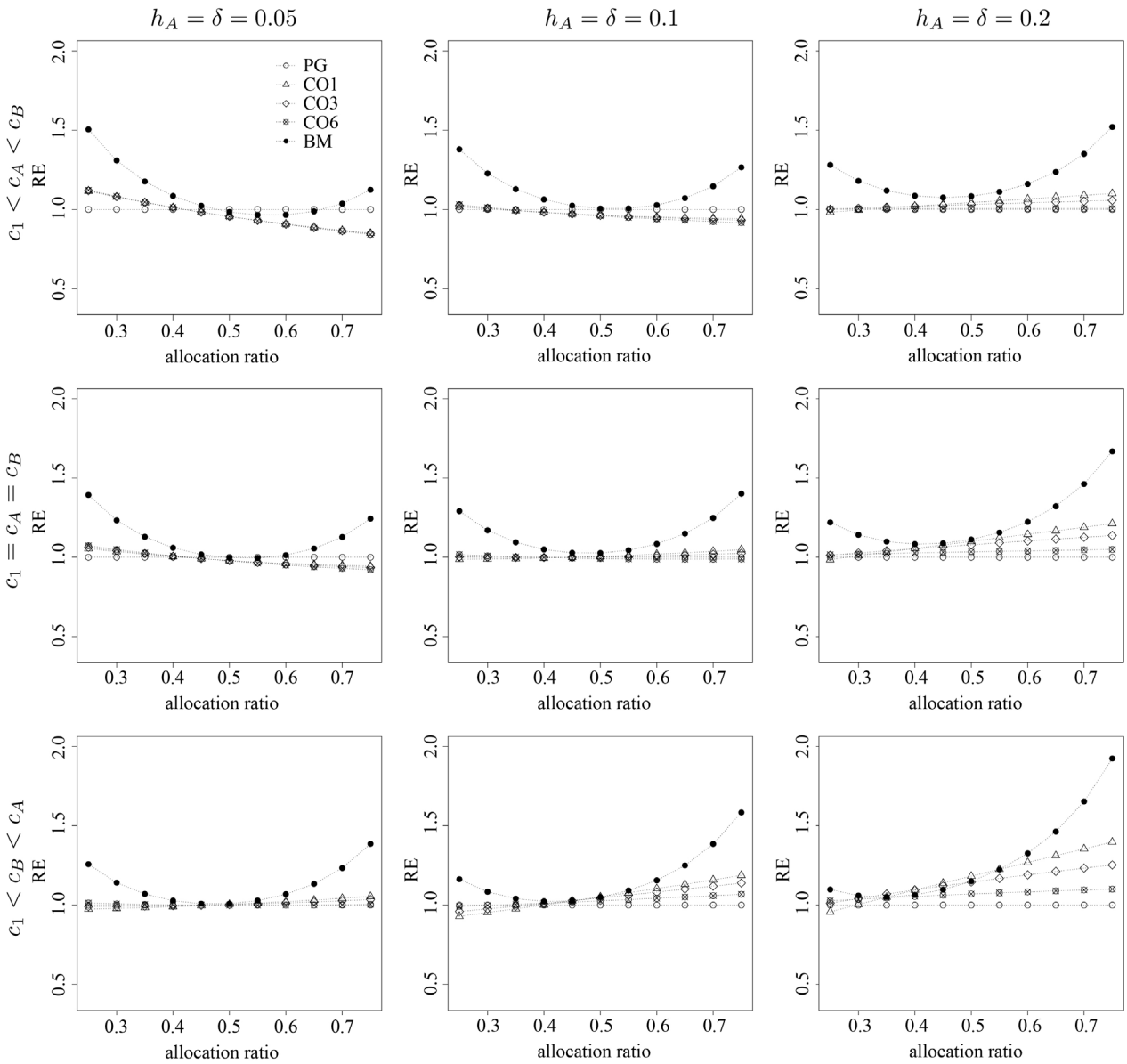


**Figure 3.** Efficiency of selected designs with two treatments $A$ and $B$ in comparison with the PG design as a function of the design allocation proportion $\pi$ for various $h_A$ and $\delta$ and the cost ratios $c_1$, $c_A$ and $c_B$ using cost function I in Equation (3) $(r = 0)$ when $p = 12$.

outcomes is given by a study investigating the effect of using controlled ovarian hyper-stimulation on the probability of conception via intrauterine insemination (IUI) [13]. A total of 74 couples with male sub-fertility are randomized to IUI in a natural cycle or to IUI in a cycle with ovarian stimulation. Each couple is given a total of six treatment cycles, three with IUI in natural cycles and three with IUI in cycles with ovarian stimulation. The couples alternate the treatments according to a CO1 design. The primary outcome measure is the pregnancy rate over the cycles. The study reports the pregnancy rates per completed cycle after IUI in either treatment.

We use the rates in the ovarian stimulation cycles as the probability of con-

ception in each cycle for the current treatment. We presume there is a newly developed treatment expected to further improve the efficacy of IUI in increasing the probability of conception, and the difference between the two treatments on the logit scale is $\beta = 0.5$. Figure 4 presents the survival probabilities and hazard probabilities on the logit scale when a PG design is applied. A value of $\beta = 0.5$ results in a decrease of about 16% in the survival after six cycles. A number of study designs might be suggested for this study including a PG design, a CO1 design (ABABAB/BABABA), a CO2 design (AABBAA/ BBAABB), a CO3 design (AAABBB/BBBAAA), and a BM design (ABABAB/BABABA/AAAAAA/ BBBBBB).

Table 3 reports the REs of the selected designs in comparison with the PG

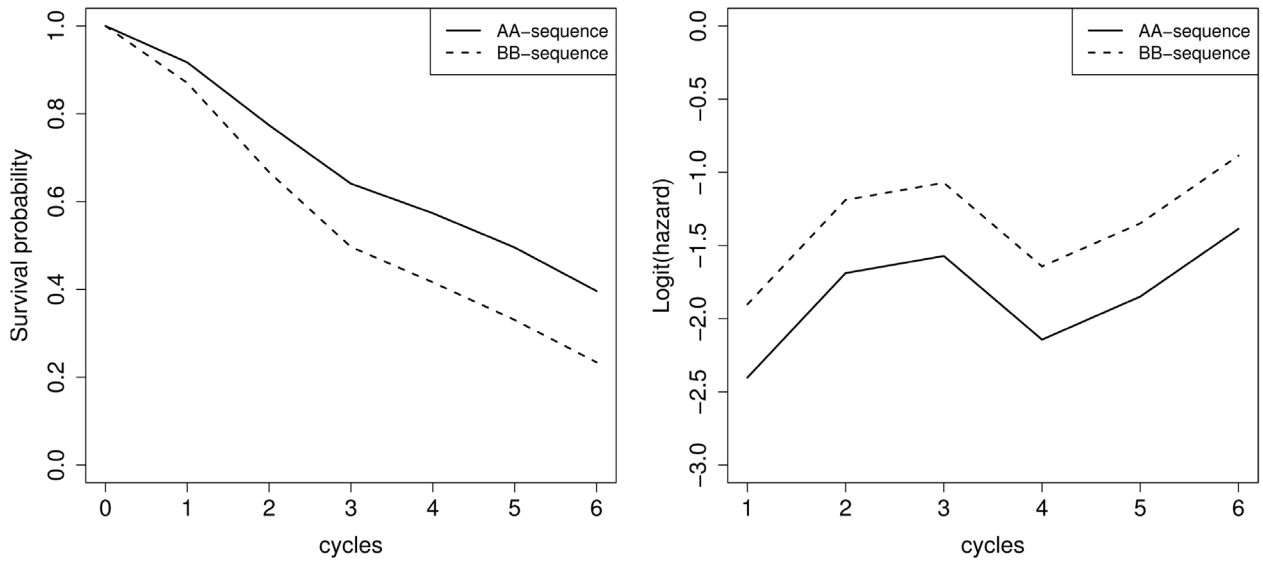


Figure 4. Fitted survivor function (left side) and logit (hazard) function (right side) for the first pregnancy example based on a PG design when $\beta = 0.5$, $\pi = 0.5$, and $r = 0.00$.

Table 3. Cost efficiency of the CO and BM designs relative to the PG design in the first pregnancy example for different $\pi$ and $c_1$, $c_A$ and $c_B$ values when $\beta = 0.5$ and $p = 6$.

| Allocation proportion | Subject-level costs | Study designs | | | |
|---|---|---|---|---|---|
| | | CO1 | CO2 | CO3 | BM |
| $\pi = 0.2$ | $c_1 = 10, c_A = 10, c_B = 10$ | 1.02 | 1.01 | 1.04 | 1.58 |
| | $c_1 = 10, c_A = 10, c_B = 20$ | 1.15 | 1.09 | 1.14 | 1.84 |
| | $c_1 = 10, c_A = 20, c_B = 10$ | 0.89 | 0.92 | 0.94 | 1.33 |
| $\pi = 0.5$ | $c_1 = 10, c_A = 10, c_B = 10$ | 0.99 | 0.99 | 0.99 | 1.00 |
| | $c_1 = 10, c_A = 10, c_B = 20$ | 0.97 | 0.97 | 0.97 | 0.99 |
| | $c_1 = 10, c_A = 20, c_B = 10$ | 1.02 | 1.02 | 1.02 | 1.01 |
| $\pi = 0.8$ | $c_1 = 10, c_A = 10, c_B = 10$ | 0.98 | 0.98 | 0.96 | 1.53 |
| | $c_1 = 10, c_A = 10, c_B = 20$ | 0.82 | 0.87 | 0.82 | 1.26 |
| | $c_1 = 10, c_A = 20, c_B = 10$ | 1.15 | 1.11 | 1.08 | 1.81 |

design for different allocation proportions and three combinations of costs at the subject-level. In the first combination, we assume sampling couples to be as expensive as treating them with either treatment. For the other combinations, we assume that stimulated cycles cost differently than those in which IUI is applied with the new treatment. We observe that if $\pi = 0.5$, all the designs are almost equally efficient since the REs are either very close or equal to 1; researchers can then choose the design that best counters the practical objections, regardless of design cost. If $\pi = 0.2$, the CO designs are almost as efficient as the PG design when both treatments cost equally (*i.e.* $c_A = c_B$), or they become more efficient when offering IUI in a new treatment cycle is more expensive than in cycles with ovarian stimulation (*i.e.* $c_A < c_B$). The reverse happens if $\pi = 0.8$, meaning that the CO designs are less efficient when $c_B \geq c_A$. With unequal treatment sequences, the BM design is always a more efficient choice among the designs. However, the RE of this design depends on design cost when $\pi \neq 0.5$.

## 6. Discussion

The present study is designed to compare the efficiency and cost efficiency of the crossover (CO) design and Balaam's (BM) design with that of a parallel group (PG) design in trials with discrete-time survival endpoints. We consider designs with two treatments *A* and *B* and focus on how efficient a design is for estimating differences between the treatment conditions. We consider CO designs that differ in the number of time periods after which subjects switch to the other treatment. All the calculations are performed in *R* and our *R* syntax is available upon request from the first author.

Using this efficiency comparison, our study shows that the efficiency of estimating treatment differences can be increased by a proper choice of the design. Deciding on whether the CO and BM designs are more efficient than the PG design depends on the size of true treatment differences $(\delta)$, the baseline hazard probability $(h_A)$, and on the study duration $(p)$. This depends also on whether or not the efficiency comparison is penalized by the amount of costs that a design has and whether or not attrition is taken into account. In general, we find that if the treatment sequences are equally sized, the CO and BM designs are less efficient than the PG design if $p$ is small, and a larger gain in efficiency may be obtained using the CO or BM designs instead of the PG design if $p$ is larger. The effect of a prolonged study duration on the efficiency of the CO and BM designs is larger if $\delta$ and $h_A$ are larger. We also observe that the BM design requires fewer time periods than the CO designs to become as efficient as the PG design. The CO and BM designs are either as efficient as or more efficient than the PG design when treatment *B* costs less or the same as treatment *A*. In cases where the baseline treatment is more expensive, the PG design is most often more efficient. In addition, all the designs perform almost equally well if the treatment sequences are of almost equal sizes for a given number of time periods. In studies with unequal allocation proportions, the BM design is preferable.

A similar comparison between a CO design and a PG design can be seen in the work of [12], where the outcome is a continuous-time survival endpoint. In general, they conclude that using the CO design might result in an efficiency gain. They focus on designs with two treatments and only two periods with subjects switching from one treatment to the other halfway through the study. They also study the optimal switch point, which in their case is at one-fifth of the total study length. In our study, we assume the total study length is fixed beforehand and subjects switch from one treatment to the other at a change point of one, three or six periods. Our results seem to show that the total study length and total amount of design costs play an important role in determining when a CO or BM design is more effective. The BM design generally results in a smaller loss in efficiency or provides a greater efficiency if it is used instead of the PG design. The CO designs are more preferable as the study duration becomes longer, or the more effective treatment is as expensive as or less expensive than the baseline treatment.

In the current study, we confine our focus to a model where the subject effects are fixed. However, if a design is efficient under the fixed effect model, it will also perform well under the random subject model [6]. We also limit the use of the designs to situations where assumptions of no sequence, period or carryover effects are valid. Since we study a random assignment of subjects to the sequences, the assumption of no sequence effect is not unrealistic. Moreover, the plausibility of the assumption of no carryover effect can be heightened by including an effective washout period between any two consecutive time periods. Nevertheless, the extent to which our findings are true if these assumptions are in doubt deserves to be explored further. What is more, our findings are based on a constant attrition rate across the time periods and treatment sequences. However, our *R* syntax is also suitable for unequal attrition rates across time periods and sequences. Another subject of future research might be the effect of baseline covariates on the optimal designs of within subject designs, as was studied in [19] with a parallel group design. The degree to which their results apply to trials where subjects receive different treatments over time deserves further study.

## 7. Conclusion

In conclusion, the possible advantages of a CO design compared to those of a PG design have been previously addressed in longitudinal studies with a variety of outcomes including the survival outcome. A similar investigation of the discrete-time survival data where the event time can only be measured on a discrete scale instead of a continuous scale has yet to be conducted. Our study provides additional findings on the usefulness of the CO and BM designs over the PG design if the treatment effect, baseline hazard function and number of time periods are varied.

## Acknowledgements

This research was funded by a VIDI grant from the Netherlands Organization

for Scientific Research (NWO) number 452-08-004.

## References

[1] Jóźwiak, K. and Moerbeek, M. (2012) Cost-Effective Designs for Trials with Discrete-Time Survival Endpoints. *Computational Statistics and Data Analysis*, **56**, 2086-2096. https://doi.org/10.1016/j.csda.2011.12.018

[2] Jóźwiak, K. and Moerbeek, M. (2013) Optimal Treatment Allocation and Study Duration for Trials with Discrete-Time Survival Endpoints. *Journal of Statistical Planning and Inference*, **143**, 971-982. https://doi.org/10.1016/j.jspi.2012.11.006

[3] Lindstrom, D., Sundberg-Petersson, I., Adami, J. and Tonnesen, H. (2010) Disappointment and Dropout Rate after Being Allocated to Control Group in a Smoking Cessation Trial. *Contemporary Clinical Trials*, **31**, 22-26. https://doi.org/10.1016/j.cct.2009.09.003

[4] Jones, B. and Kenward, M.G. (2002) Design and Analysis of Cross-Over Trials. Chapman & Hall/CRC, Boca Raton.

[5] Senn, S. (2002) Cross-Over Trials in Clinical Research. Willey, Chichester. https://doi.org/10.1002/0470854596

[6] Berger, M.P. and Wong, W.K. (2009) An Introduction to Optimal Designs for Social and Biomedical Research. Willey, Chichester. https://doi.org/10.1002/9780470746912

[7] Morel, J.G. and Neerchal, N.K. (2012) Sample Size Determination for Alternative Periods of Use Study Designs with Binary Responses. *Journal of Biopharmaceutical Statistics*, **22**, 351-367. https://doi.org/10.1080/10543406.2010.539082

[8] Senn, S. (2006) Cross-Over Trials in Statistics in Medicine: The First "25" Years. *Statistics in Medicine*, **25**, 3430-3442. https://doi.org/10.1002/sim.2706

[9] Bryant, J. and Wolmark, N. (2003) Letrozole after Tamoxifen for Breast Cancer—What Is the Price of Success? *The New England Journal of Medicine*, **349**, 1855-1857. https://doi.org/10.1056/NEJMe038167

[10] France, L.A., Lewis, J.A. and Kay, R. (1991) The Analysis of Failure Time Data in Crossover Studies. *Statistics in Medicine*, **10**, 1099-1113. https://doi.org/10.1002/sim.4780100710

[11] Cox, D.R. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.

[12] Buyze, J. and Goetghebeur, E. (2013) Crossover Studies with Survival Outcomes. *Statistical Methods in Medical Research*, **22**, 612-629. https://doi.org/10.1177/0962280211402258

[13] Cohlen, B., Velde, E., Kooij, R., Looman, C. and Habbema, J. (1998) Controlled Ovarian Hyperstimulation and Intrauterine Insemination for Treating Male Subfertility: A Controlled Study. *Human Reproduction*, **13**, 1553-1558. https://doi.org/10.1093/humrep/13.6.1553

[14] Balaam, L. (1968) A Two Period Design with $t^2$ Experimental Units. *Biometrics*, **24**, 61-73. https://doi.org/10.2307/2528460

[15] Singer, J.D. and Willett, J.B. (2003) Applied Longitudinal Data Analysis. Modeling Change and Event Occurrence. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

[16] McCullagh, R.M. and Nelder, J.A. (1989) Generalized Linear Models. Chapman and Hall, London.

[17] Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007) Optimum Experimental De-

sign, with SAS. Clarendon, Oxford.

[18] Carriere, K. and Reinsel, G. (1992) Investigation of Dual-Balanced Crossover Designs for Two Treatments. *Biometrics*, **48**, 1157-1164. https://doi.org/10.2307/2532706

[19] Safarkhani, M. and Moerbeek, M. (2014) The Influence of a Covariate on Optimal Designs in Longitudinal Studies with Discrete-Time Survival Endpoints. *Computational Statistics & Data Analysis*, **75**, 217-226. https://doi.org/10.1016/j.csda.2014.02.012

Scientific Research Publishing

**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/
Or contact ojs@scirp.org