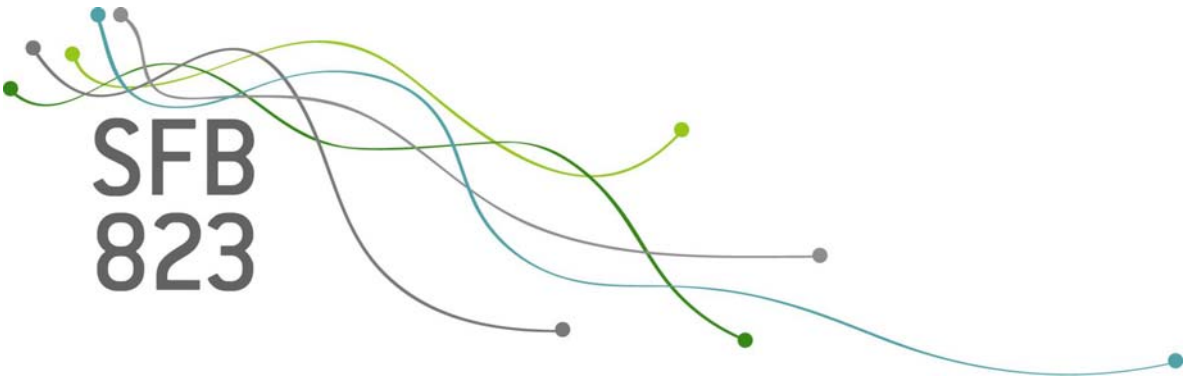
SFB 823

Universally optimal crossover designs for the estimation of mixed-carryover effects with an application to biosimilar development

Discussion Paper

Johanna Mielke, Joachim Kunert

Nr. 3/2018



Universally optimal crossover designs for the estimation of mixed-carryover effects with an application to biosimilar development

Johanna Mielke · **Joachim Kunert**

Received: date / Accepted: date

Abstract Biosimilars are medical products that are developed as copies of already established, large molecule drugs (biologics). For gaining approval, sponsors have to confirm that the proposed biosimilar has the same efficacy and safety as the originator product. This comparability exercise includes also, in most cases, that large clinical trials are conducted in patients. However, even with the evidence gained during the clinical studies, there is still some uncertainty if patients who were already treated with the originator can be switched to the biosimilar or if even multiple switches between the biosimilar and the originator are acceptable. A simple way to address the question of switchability is the estimation of so-called mixed and self-carryover effects, which are carryover effects that not only depend on the treatment in the current period, but also on the treatment in the previous period. In this paper, we determine universally optimal designs for the estimation of mixed-carryover effects in a linear model with treatment, period, subject and self-carryover as nuisance parameters.

Keywords biosimilars · self carryover effects · mixed-carryover effects · crossover design · switchability

The authors gratefully acknowledge the funding from the Deutsche Forschungsgemeinschaft (SFB 823), from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567 and from the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 999754557. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Swiss Government.

Johanna Mielke
Novartis Pharma AG, Basel, Switzerland
Tel.: +41613245702
Fax: +41613243039
E-mail: johanna.mielke@udo.edu

Joachim Kunert
Department of Statistics, University of Dortmund, Germany

1 Introduction

A biosimilar (the test product) is developed as a copy of an already approved large molecule drug, a so-called biologic (the reference product). Market authorization is granted after the sponsor of the biosimilar has shown the comparability of the test and the reference products at the analytical, non-clinical and clinical level (CHMP, 2014). If a biosimilar gains approval, it has been confirmed that patients who are taking the biosimilar can expect the same treatment effect and the same safety profile as patients who are taking the reference product.

Biosimilars are still a fairly new concept with the first biosimilar approved in Europe in 2006 (Omnitrope, Sandoz) and the first biosimilar approved in the USA in 2015 (Zarxio, Sandoz). Even though there are by now already 36 approved biosimilars in Europe (Generics and Biosimilars Initiative, 2017a) and 7 approved biosimilars in the USA (Generics and Biosimilars Initiative, 2017b) and many more are expected to gain market authorization in the next years, there is still some uncertainty if patients who were already treated with the reference product can be switched to the biosimilar or even if multiple switches are allowed. This would mean that the automatic substitution of the reference product by the biosimilar at the pharmacy level could be possible. The practice of switching at the pharmacy level is mostly accepted for generics. For biosimilars, the positions of regulatory agencies are diverse: for example, regulators in Finland state that ”switches between biological products are common and usually not problematic” (Finnish Medicines Agency, 2015) while the regulatory agency in Ireland ”does not recommend that patients are switched back and forth between a biosimilar and the reference medicinal product” (Health Products Regulatory Authority, 2015).

Additional evidence might be required to support the decision as to whether single or multiple switches between a biosimilar and its reference product should be permitted. This evidence could be provided by an additional clinical study that is conducted specifically for analysing the impact of single or multiple switches on the treatment success. A simple statistical methodology for showing the impact of switching between a biosimilar and its reference product could be based on the direct estimation of the impact of switching by calculating the so-called mixed and self-carryover effects. This idea was first proposed by Afsarinejad and Hedayat (2002). They introduced a model in which the usual carryover effect that only depends on the treatment in the previous period is replaced by two different carryover effects per treatment. These carryover effects do not only depend on the treatment in the previous period, but also on the treatment in the current period: if both treatments are different, a mixed-carryover effects is introduced, if the treatment are the same, a self-carryover effect is used. In the case of two treatments, the proposed model is identical to the inclusion of interactions between the direct effects and carryover effects. From previous parallel groups trials, it will generally be known prior to the assessment of switchability that the direct effects of the two treatments can be considered similar. From previous long term experiments (where subjects receive either reference or test over several periods) it will also be known that the self-carryover effects can be considered to

be identical. However, for claiming switchability, we might want to confirm that the mixed-carryover effects have equal values. Hence, it is necessary to perform a study which uses a design that allows the estimation of the mixed-carryover effects.

It is generally desirable to use a study design which is as efficient as possible. While for the showing of biosimilarity, mostly parallel groups designs were conducted in the past (Mielke et al, 2016), such designs are clearly not appropriate for the assessment of switchability because one or multiple switches have to be included in the study design. So far, not many switching studies have been published and there is currently no well-established standard for the choice of the study design. In this paper, we determine universally optimal designs for the estimation of mixed-carryover effects in the case of two treatments (the test and the reference product) in a linear model with treatment, period, subject and self-carryover effects as nuisance parameters. This work builds on results by Kunert and Stufken (2002) and Kunert and Stufken (2008) who studied optimal designs for the estimation of the treatment effect when period, subject and both self and mixed-carryover effects are nuisance parameters. We study efficient designs for the joint estimation of self and mixed-carryover effects in a separate paper.

The rest of the paper is structured as follows: in Section 2, we introduce the necessary notation that is used in Section 3 for identifying the optimal designs. In Section 4, we compare the derived optimal designs with designs used in practice. In Section 5, we investigate if the designs can be improved by including periods without any treatment. We summarize our results in Section 6.

2 Notations and definitions

We consider the model that was described in Kunert and Stufken (2002) and Kunert and Stufken (2008). We assume that the response $y_{u,r}$ of subject u ($u = 1,...,n$) in period r ($r = 1,...,p$) can be written by

$$y_{u,r} = \begin{cases} \alpha_u + \beta_r + \tau_{d(u,r)} + \rho_{d(u,r-1)} + e_{u,r} & \text{if } d(u,r) \neq d(u,r-1) \\ \alpha_u + \beta_r + \tau_{d(u,r)} + \chi_{d(u,r-1)} + e_{u,r} & \text{if } d(u,r) = d(u,r-1) \end{cases},$$

where $d(u,r)$ gives the treatment of subject u in period r, α_u is the subject effect of subject u, β_r is the period effect in period r, τ_i is the treatment effect of treatment i, ρ_i is the mixed-carryover effect of treatment i and χ_i is the self-carryover effect of treatment i. The residual error $e_{u,r}$ is assumed to be independent and identically distributed with expectation 0 and variance σ^2. To simplify the notation, we set $\sigma^2 = 1$.

The set of all designs with t treatments, n subjects and p periods is denoted as $\Omega_{t,n,p}$. Here, we focus on the case of two treatments (Test - T, Reference - R; $t = 2$) and $p > 2$. Using the notation of Kunert and Stufken (2002), we define the matrices $\mathbf{U} = \mathbf{I}_n \otimes \mathbf{1}_p$ (subject effect), $\mathbf{P} = \mathbf{1}_n \otimes \mathbf{I}_p$ (period effect), $\mathbf{T}_d$ (treatment effect), $\mathbf{M}_d$ (mixed-carryover effects) and $\mathbf{S}_d$ (self-carryover effect), where $\otimes$ denotes the Kronecker product, $\mathbf{I}_m$ is the identity matrix of dimension m and $\mathbf{1}_m$ is a vector with

length m that only contains the entries 1.

The goal of this analysis is to derive the characteristics of a design d that is universally optimal for the estimation of the mixed-carryover effects in the class of design $\Omega_{2,n,p}$. Universally optimal is a term introduced by Kiefer (1975). If all information matrices $\mathbf{C}_d$ have row sums and column sums 0, then a design d^* is universally optimal if its information matrix $\mathbf{C}_{d^*}$ is completely symmetric and the design maximises the trace of $\mathbf{C}_d$ over all $d \in \Omega_{2,n,p}$. A matrix $\mathbf{A}$ is called completely symmetric if it can be written in the form

$$\mathbf{A} = a\mathbf{I} + b\mathbf{1}\mathbf{1}^T,$$

where a and b are real numbers. The information matrix for the estimation of the mixed-carryover effects is given by

$$\tilde{\mathbf{C}}_d = \mathbf{M}_d^T \omega^{\perp}([\mathbf{P}, \mathbf{U}, \mathbf{T}_d, \mathbf{S}_d])\mathbf{M}_d,$$

where $\omega^{\perp}(\mathbf{A}) = \mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-}\mathbf{A}^T$ is the projection on the space of all vectors that are orthogonal to the columns of $\mathbf{A}$ and where $\mathbf{A}^T$ gives the transpose and $\mathbf{A}^{-}$ the g-inverse of $\mathbf{A}$. It can be easily confirmed that all information matrices $\tilde{\mathbf{C}}_d$ have column and row sums 0: for that, define $\mathbf{q} = \begin{pmatrix} 0 \\ \mathbf{1}_{p-1} \end{pmatrix}$. Then, we observe that

$$\mathbf{M}_d\mathbf{1}_2 + \mathbf{S}_d\mathbf{1}_2 = \mathbf{P}\mathbf{q} \Rightarrow \mathbf{M}_d\mathbf{1}_2 = \mathbf{P}\mathbf{q} - \mathbf{S}_d\mathbf{1}_2 \in im([\mathbf{S}_d, \mathbf{P}]) \subset im([\mathbf{T}_d, \mathbf{S}_d, \mathbf{U}, \mathbf{P}]),$$

where $im(\mathbf{A})$ gives the image of matrix $\mathbf{A}$ and the first equality holds because each subject experiences a mixed or a self-carryover effect in all periods except in the first period. With that, we know that

$$\omega^{\perp}([\mathbf{T}_d, \mathbf{S}_d, \mathbf{U}, \mathbf{P}])\mathbf{M}_d\mathbf{1}_2 = 0,$$

which directly leads to the conclusion that the column sums and row sums of $\tilde{\mathbf{C}}_d = \mathbf{M}_d^T \omega^{\perp}([\mathbf{P}, \mathbf{U}, \mathbf{T}_d, \mathbf{S}_d])\mathbf{M}_d$ are 0. Therefore, the concept of universally optimal is applicable.

3 Identification of universally optimal designs

The strategy presented in this paper for identifying universally optimal designs follows the ideas of Kunert and Stufken (2002) and Kunert and Stufken (2008) and consists of two main steps: first, a matrix $\mathbf{C}_d$ that is larger in the Loewner sense than the information matrix $\tilde{\mathbf{C}}_d$ is derived (Section 3.1). Then, an upper bound for the trace of $\mathbf{C}_d$ is determined and a class of designs is identified that reaches this bound (Section 3.2).

3.1 An upper bound for the information matrix

In this section, we derive an upper bound for the information matrix $\tilde{\mathbf{C}}_d$ in the Loewner sense. Since $\tilde{\mathbf{C}}_d$ has row and column sums 0 and is completely symmetric, we can multiply the information matrix from both sides with a matrix $\mathbf{B}_2 \in \mathbb{R}^{2x2}$ which has diagonal elements $1/2$ and off-diagonal elements $-1/2$ without changing the matrix and we have the equality

$$\tilde{\mathbf{C}}_d = \mathbf{B}_2^T \tilde{\mathbf{C}}_d \mathbf{B}_2 = (\mathbf{M}_d \mathbf{B}_2)^T \omega^{\perp}([\mathbf{P}, \mathbf{U}, \mathbf{T}_d, \mathbf{S}_d])(\mathbf{M}_d \mathbf{B}_2).$$

The upper bound of the information matrix is obtained by removing the period effect from the information matrix. More concrete, we use Proposition 2.3 in Kunert (1983) which states that

$$\tilde{\mathbf{C}}_d \leq (\mathbf{M}_d \mathbf{B}_2)^T \omega^{\perp}([\mathbf{U}, \mathbf{T}_d, \mathbf{S}_d])\mathbf{M}_d \mathbf{B}_2 = \mathbf{C}_d, \text{ say,} \tag{1}$$

which equality if and only if

$$(\mathbf{M}_d \mathbf{B}_2)^T \omega^{\perp}([\mathbf{U}, \mathbf{T}_d, \mathbf{S}_d])\mathbf{P} = 0. \tag{2}$$

Then, as shown in Kunert and Martin (2000), it is possible to use the following decomposition of the information matrix $\mathbf{C}_d$:

$$\begin{aligned}\mathbf{C}_d =& \mathbf{C}_{d11} - \mathbf{C}_{d12}\mathbf{C}_{d22}^{-}\mathbf{C}_{d12}^T - (\mathbf{C}_{d13} - \mathbf{C}_{d12}\mathbf{C}_{d22}^{-}\mathbf{C}_{d23}) \\ & (\mathbf{C}_{d33} - \mathbf{C}_{d23}^T\mathbf{C}_{d22}^{-}\mathbf{C}_{d23})^{-}(\mathbf{C}_{d13} - \mathbf{C}_{d12}\mathbf{C}_{d22}^{-}\mathbf{C}_{d23})^T,\end{aligned}$$

where, in the situation of this paper,

$$\begin{aligned}\mathbf{C}_{d11} &= \mathbf{B}_2\mathbf{M}_d^T\mathbf{M}_d\mathbf{B}_2 - \frac{1}{p}\mathbf{B}_2\mathbf{M}_d^T\mathbf{U}\mathbf{U}^T\mathbf{M}_d\mathbf{B}_2, \\ \mathbf{C}_{d12} &= \mathbf{B}_2\mathbf{M}_d^T\mathbf{T}_d - \frac{1}{p}\mathbf{B}_2\mathbf{M}_d^T\mathbf{U}\mathbf{U}_d^T\mathbf{T}, \\ \mathbf{C}_{d13} &= \mathbf{B}_2\mathbf{M}_d^T\mathbf{S}_d - \frac{1}{p}\mathbf{B}_2\mathbf{M}_d^T\mathbf{U}\mathbf{U}^T\mathbf{S}_d, \\ \mathbf{C}_{d22} &= \mathbf{T}_d^T\mathbf{T}_d - \frac{1}{p}\mathbf{T}_d^T\mathbf{U}\mathbf{U}^T\mathbf{T}_d^T, \\ \mathbf{C}_{d23} &= \mathbf{T}_d^T\mathbf{S}_d - \frac{1}{p}\mathbf{T}_d^T\mathbf{U}\mathbf{U}^T\mathbf{S}_d^T, \\ \mathbf{C}_{d33} &= \mathbf{S}_d^T\mathbf{S}_d - \frac{1}{p}\mathbf{S}_d^T\mathbf{U}\mathbf{U}^T\mathbf{S}_d^T.\end{aligned}$$

In order to use the reduced information matrix stated in Equation (1), it is necessary to show that the condition stated in Equation (2) is fulfilled. We first note that there is always a dual-balanced design among the optimal designs. A sequence s is called dual to a sequence s' if sequence s can be changed to sequence s' by interchanging the two treatments (e.g., TRTR, dual sequence: RTRT). A design d is dual-balanced if it uses sequence s exactly as often as sequence s'. It is noteworthy that dual-balanced designs fulfill all conditions stated in Kunert and Stufken (2002)

(in each period, all treatments appear equally often, the mixed carryover effects of all treatments appear equally often, the self carryover effects of all treatments appear equally often) and we can therefore directly use the results obtained in Kunert and Stufken (2002). We refer to the conditions stated in Kunert and Stufken (2002) as conditions $(*)$. For showing that Equation (2) holds, we first rewrite the equation:

$$\begin{aligned}(\mathbf{M}_d\mathbf{B}_2)^T\omega^\perp([\mathbf{U},\mathbf{T}_d,\mathbf{S}_d])\mathbf{P} =& \mathbf{B}_2\mathbf{M}_d^T\omega^\perp(\mathbf{U})\mathbf{P} - \mathbf{C}_{d12}\mathbf{C}_{d22}^-\mathbf{T}_d^T\omega^\perp(\mathbf{U})\mathbf{P}\\ &-(\mathbf{C}_{d13}-\mathbf{C}_{d12}\mathbf{C}_{d22}^-\mathbf{C}_{d23})(\mathbf{C}_{d33}-\mathbf{C}_{d23}^T\mathbf{C}_{d22}^-\mathbf{C}_{d23})^-\\ &(\mathbf{T}_d^T\omega^\perp(\mathbf{U})\mathbf{P}-\mathbf{C}_{d23}\mathbf{C}_{d22}^-\mathbf{S}_d^T\omega^\perp(\mathbf{U})\mathbf{P}).\end{aligned}$$

From Kunert and Stufken (2002), we use that $\mathbf{T}_d^T\omega^\perp(\mathbf{U})\mathbf{P}=0$ and

$$\mathbf{B}_2\mathbf{M}^T\omega^\perp(\mathbf{U})\mathbf{P}=0$$

for the designs which fulfill conditions $(*)$, i.e., for all dual-balanced designs, which simplifies the term to

$$\begin{aligned}(\mathbf{M}_d\mathbf{B}_2)^T\omega^\perp([\mathbf{U},\mathbf{T}_d,\mathbf{S}_d])\mathbf{P} =& -(\mathbf{C}_{d13}-\mathbf{C}_{d12}\mathbf{C}_{d22}^-\mathbf{C}_{d23})(\mathbf{C}_{d33}-\mathbf{C}_{d23}^T\mathbf{C}_{d22}^-\mathbf{C}_{d23})^-\\ &(-\mathbf{C}_{d23}\mathbf{C}_{d22}^-\mathbf{S}_d^T\omega^\perp(\mathbf{U})\mathbf{P}).\end{aligned}$$

The matrices $\mathbf{C}_{d22}$ and $\mathbf{C}_{d23}$ have column sums 0 (Kunert and Stufken, 2002) and as these matrices are in addition completely symmetric, they are a multiple of $\mathbf{B}_2$. Therefore, we can replace

$$\begin{aligned}\mathbf{C}_{d22} &= k_{22}\mathbf{B}_2,\\ \mathbf{C}_{d23} &= k_{23}\mathbf{B}_2.\end{aligned}$$

Furthermore, as $\mathbf{B}_2$ is idempotent,

$$\mathbf{B}_2^- = \mathbf{B}_2.$$

This leads to

$$\begin{aligned}(\mathbf{M}_d\mathbf{B}_2)^T\omega^\perp([\mathbf{U},\mathbf{T}_d,\mathbf{S}_d])\mathbf{P} =& -(\mathbf{C}_{d13}-\mathbf{C}_{d12}(k_{22}\mathbf{B}_2)^-k_{23}\mathbf{B}_2)\\ &(\mathbf{C}_{d33}-(k_{23}\mathbf{B}_2)^T(k_{22}\mathbf{B}_2)^-k_{23}\mathbf{B}_2)^-\\ &(-k_{23}\mathbf{B}_2(k_{22}\mathbf{B}_2)^-\mathbf{S}_d^T\omega^\perp(\mathbf{U})\mathbf{P})\\ =& -(\mathbf{C}_{d13}-\mathbf{C}_{d12}\frac{k_{23}}{k_{22}}\mathbf{B}_2)\left(\mathbf{C}_{d33}-\frac{k_{23}}{k_{22}}\mathbf{B}_2\right)^-\\ &\left(-\frac{k_{23}}{k_{22}}\mathbf{B}_2\mathbf{S}_d^T\omega^\perp(\mathbf{U})\mathbf{P}\right)\\ =& 0,\end{aligned}$$

where the last equality uses that $\mathbf{B}_2\mathbf{S}_d^T\omega^\perp(\mathbf{U})\mathbf{P}=0$ (Kunert and Stufken, 2002). Therefore, we only need to determine sequences which maximise the trace of the matrix

$$\mathbf{C}_d = (\mathbf{M}_d\mathbf{B}_2)^T\omega^\perp([\mathbf{U},\mathbf{T}_d,\mathbf{S}_d])\mathbf{M}_d\mathbf{B}_2.$$

3.2 Maximising the trace of $\mathbf{C}_d$

Since the column sums of $\mathbf{C}_{d11}, \mathbf{C}_{d12}$ and $\mathbf{C}_{d13}$ are 0 because of the multiplication by $\mathbf{B}_2$, we are in the same setting as in Proposition 2 in Kunert and Martin (2000) and can therefore use the same strategy to find the optimal design. Let $x, y \in \mathbb{R}$ and l be an equivalence class of sequences which consists of a specific sequence s and its dual-balanced sequence s'. Since we are in the two-treatment case and we consider only dual-balanced designs, $\Omega_{2,n,p}$ consists of 2^{p-1} different equivalence classes. Let π_{dl} be a vector of length 2^{p-1} which gives the proportion of sequences of the design d which belongs to the lth equivalence class. We use from Kunert and Stufken (2008) that for any design $d \in \Omega_{2,n,p}$,

$$tr(\mathbf{C}_d) \leq n \cdot \min_{x,y} \sum_{l=1}^{2^{p-1}} \pi_{dl} h_l(x,y) =: q_d^*,$$

with

$$h_l(x,y) = c_{11}(l) + 2xc_{12}(l) + x^2 c_{22}(l) + 2yc_{13}(l) + y^2 c_{33}(l) + 2xyc_{23}(l),$$

where, in our situation,

$$\begin{aligned}
c_{11}(l) &:= tr(\mathbf{B}_2 \mathbf{M}_u^T w^{\perp}(\mathbf{U}_u) \mathbf{M}_u \mathbf{B}_2) = tr(\mathbf{B}_2 (\mathbf{M}_u^T (\mathbf{I} - \mathbf{U}_u (\mathbf{U}_u^T \mathbf{U}_u)^{-} \mathbf{U}_u^T) \mathbf{M}_u) \mathbf{B}_2) \\
&= tr\left(\mathbf{B}_2 \left(\mathbf{M}_u^T \mathbf{M}_u - \frac{1}{p} \mathbf{M}_u^T \mathbf{U}_u \mathbf{U}_u^T \mathbf{M}_u \right) \mathbf{B}_2 \right) \\
&= tr\left(\mathbf{B}_2 \left(\mathbf{M}_u^T \mathbf{M}_u - \frac{1}{p} \mathbf{M}_u^T \mathbf{U}_u \mathbf{U}_u^T \mathbf{M}_u \right) \right), \\
c_{12}(l) &:= tr(\mathbf{B}_2 \mathbf{M}_u^T w^{\perp}(\mathbf{U}_u) \mathbf{T}_u) = tr\left(\mathbf{B}_2 \left(\mathbf{M}_u^T \mathbf{T}_u - \frac{1}{p} \mathbf{M}_u^T \mathbf{U}_u \mathbf{U}_u^T \mathbf{T}_u \right) \right), \\
c_{13}(l) &:= tr(\mathbf{B}_2 \mathbf{M}_u^T w^{\perp}(\mathbf{U}_u) \mathbf{S}_u) = tr\left(\mathbf{B}_2 \left(\mathbf{M}_u^T \mathbf{S}_u - \frac{1}{p} \mathbf{M}_u^T \mathbf{U}_u \mathbf{U}_u^T \mathbf{S}_u \right) \right), \\
c_{22}(l) &:= tr(\mathbf{B}_2 \mathbf{T}_u^T w^{\perp}(\mathbf{U}_u) \mathbf{T}_u) = tr\left(\mathbf{B}_2 \left(\mathbf{T}_u^T \mathbf{T}_u - \frac{1}{p} \mathbf{T}_u^T \mathbf{U}_u \mathbf{U}_u^T \mathbf{T}_u \right) \right), \\
c_{23}(l) &:= tr(\mathbf{B}_2 \mathbf{T}_u^T w^{\perp}(\mathbf{U}_u) \mathbf{S}_u) = tr\left(\mathbf{B}_2 \left(\mathbf{T}_u^T \mathbf{S}_u - \frac{1}{p} \mathbf{T}_u^T \mathbf{U}_u \mathbf{U}_u^T \mathbf{S}_u \right) \right), \\
c_{33}(l) &:= tr(\mathbf{B}_2 \mathbf{S}_u^T w^{\perp}(\mathbf{U}_u) \mathbf{S}_u) = tr\left(\mathbf{B}_2 \left(\mathbf{S}_u^T \mathbf{S}_u - \frac{1}{p} \mathbf{S}_u^T \mathbf{U}_u \mathbf{U}_u^T \mathbf{S}_u \right) \right).
\end{aligned}$$

In these equations, $\mathbf{M}_u, \mathbf{U}_u, \mathbf{T}_u, \mathbf{S}_u$ are the design matrices for the uth subject of the mixed-carryover effects, the subject effects, the treatment effects and the self-carryover effects, respectively. Due to the properties of π_{dl} (non-negative, $\sum_{l=1}^{2^{p-1}} \pi_{dl} = 1$), it is clear that

$$\sum_{l=1}^{2^{p-1}} \pi_{dl} h_l(x,y) \leq \max_l h_l(x,y)$$

which leads to

$$tr(\mathbf{C}_d) \leq \min_{x,y} \max_l h_l(x,y).$$

Let x^*, y^* be values such that

$$\max_l h_l(x^*, y^*) \leq \max_l h_l(x, y).$$

Following Kunert and Stufken (2008), we define L^* as the set of equivalence classes with

$$l \in L^* \text{ if and only if } h_l(x^*, y^*) = \max_\mu h_\mu(x^*, y^*).$$

As in previous publications, the main technical difficulty is to identify numbers x^*, y^* and optimal classes of sequences $l \in L^*$. In the derivation of optimal designs for the estimation of the direct treatment effects in the presence of mixed and self-carryover effects as nuisance parameters (two treatments, Kunert and Stufken, 2008), the task was massively simplified because it was possible to show that the function h_l is, at the optimal values x^*, y^*, independent of the choice of the equivalence class l^*. Unfortunately, this trick is not applicable in our case. Therefore, we optimize x^*, y^* and the optimal classes of sequences $l \in L^*$ simultaneously. It is important to note that since the terms $c_{ij}(l)$ are invariant for a sequence and its dual sequence and therefore are invariant for all sequences within one equivalence class, we represent an equivalence class l, without loss of generality, with a sequence s that ends with treatment T. For example, if TRT and RTR are the dual sequences of equivalence class l, we focus on the sequence $s = TRT$. In the following we introduce additional notation which is required for the identification of optimal designs. For that, let $n_j(s)$ be the number of appearances of treatment T ($j = 1$) or R ($j = 2$) in the sequence s, $\tilde{n}_j(s)$ is the number of appearances of mixed-carryover effects and $t_{pj}(s)$ is 1 if treatment j is in the last period and 0 otherwise. With that notations, it follows that for every s, there is exactly one j such that $t_{pj}(s) = 1$ and it is 0 in all other cases. The number of appearances of self-carryover effects is given by

$$\bar{n}_j(s) = n_j(l) - \tilde{n}_j(s) - t_{pj}(s).$$

It is necessary to distinguish between the sequences that start with treatment T and the sequences that start with treatment R. In the first case, if $\tilde{n}_1(s)$ is the number of mixed-carryovers for treatment T, the number of mixed carryovers for treatment R is $\tilde{n}_2(s) = \tilde{n}_1(s)$. If the treatment sequence starts with an R, the number of mixed carryovers for R is $\tilde{n}_2(s) = \tilde{n}_1(s) + 1$. Using this notation, it is possible to derive simpler terms for c_{ij} (see Appendix) and this allows the simplification of the function h_l. In our situation, for the special case of two treatments, we can write the function h_l as

$$\begin{aligned} h_{l,1}(x,y) =& \tilde{n}_1(s)\left(1 - 2x - y^2 - 2xy\right) + n_1^2(s)\left(\frac{-2x^2}{p} - \frac{2y^2}{p} - \frac{4xy}{p}\right) \\ &+ n_1(s)\left(2x^2 + 2y^2 + \frac{2y^2}{p} + 4xy + \frac{2xy}{p}\right) \\ &+ \left(-\frac{y^2}{2} - \frac{y^2}{2p} - y^2 - 2xy\right) \end{aligned}$$

in case the sequence s starts with T and as

$$\begin{aligned} h_{l,2}(x,y) =& \tilde{n}_1(s)\left(1-2x-y^2-2xy\right)+n_1(s)^2\left(\frac{-2x^2}{p}-\frac{2y^2}{p}-\frac{4xy}{p}\right) \\ &+n_1(s)\left(\frac{2x}{p}+\frac{2y}{p}+4xy+2x^2+2y^2\right) \\ &+\left(\frac{1}{2}-\frac{1}{2p}-2x-y-y^2-2xy\right) \end{aligned}$$

in case the sequence s starts with R. It is noteworthy that the trace of the information matrix is completely determined by $n_1(s)$ and $\tilde{n}_1(s)$. Therefore, these are the parameters that need to be identified for the determination of the optimal design. The following new proposition is the main tool for the identification of optimal sequences in the case of designs with an odd number of periods.

Proposition 1 *Let p be an odd number. Then, for sequences starting and ending with the same treatment, the upper boundary of the trace of $\mathbf{C}_d$, q_d^*, is attained by*

$$x^* = \frac{p}{p+1},\ y^* = -\frac{p}{p+1}$$

and

$$n_1^*(s) = \frac{p+1}{2},\ \tilde{n}_1^*(s) = \frac{p-1}{2}.$$

Proof The proof consists of three main steps: first, we start with a pair of characteristics of the sequence s, $n_1(s)$ and $\tilde{n}_1(s)$, that we consider to be optimal. In the next step, we determine optimal values x^*, y^* for this choice and confirm afterwards that our choice of $n_1(s)$ and $\tilde{n}_1(s)$ was indeed optimal.

For that, we consider sequences s with the characteristics

$$n_1^*(s) = \frac{p+1}{2} \text{ and } \tilde{n}_1^*(s) = \frac{p-1}{2}.$$

Next, the corresponding optimal values of x, y should be determined. The partial derivatives of $h_{l,1}(x,y)$ with respect to x, y are given by

$$\begin{aligned} \frac{dh_{l,1}(x,y)}{dx} =& -2\tilde{n}_1(s)-2y\tilde{n}_1(s)-\frac{4x}{p}n_1(s)^2-\frac{4y}{p}n_1(s)^2+4xn_1(s) \\ &+4yn_1(s)+\frac{2yn_1(s)}{p}-2y, \\ \frac{dh_{l,1}(x,y)}{dy} =& -2y\tilde{n}_1(s)-2x\tilde{n}_1(s)-\frac{4y}{p}n_1(s)^2-\frac{4x}{p}n_1(s)^2 \\ &+4yn_1(s)+\frac{4yn_1(s)}{p}+4xn_1(s)+\frac{2xn_1(s)}{p}-y-\frac{y}{p}-2y-2x. \end{aligned}$$

Settings these equations equal to 0, we note that all values of y and

$$x^* = \frac{p}{p+1}$$

are optimal. In the following, we set

$$y^* := -\frac{p}{p+1} = -x^*.$$

In the direction of x, the second derivative is positive, i.e., there exists only one minimum in the direction of x. The function is constant in the direction of y. Therefore, we have identified a global minimum. Including our knowledge in the function $h_{l,1}$ leads to

$$\tilde{h}_{l,1} := h_{l,1}(x^*, y^*) = \frac{\tilde{n}_1(s)}{(p+1)^2} + \frac{p^2-p}{2(p+1)^2}.$$

This function appears to be independent of the appearance of treatment T ($n_1(s)$ is not present in the formula), but this is not the case: the appearance of the treatments influences the range of possible mixed-carryover effects and therefore indirectly still influences the choice of optimal designs.

In the third step, we need to confirm that our choice of $n_1(s), \tilde{n}_1(s)$ was indeed optimal. For this, it is sufficient to confirm that the derived function $\tilde{h}_{l,1}$ is maximal for

$$n_1(s) = \frac{p+1}{2} \text{ and } \tilde{n}_1(s) = \frac{p-1}{2}.$$

We only need to confirm optimality at the point of the identified values x^*, y^*, since this is already the minimum of the function $h_l^*(x,y)$ with respect to x, y. The derived function $\tilde{h}_{l,1}$ clearly shows that choosing $\tilde{n}_1(s)$ as large as possible is optimal. The largest value for a sequence starting and ending with T is $\tilde{n}_1(s) = \frac{p-1}{2}$. This choice directly determines that $n_1(s) = \frac{p+1}{2}$ which completes the proof. □

For the chosen values of mixed-carryover effects $\tilde{n}_1(l^*)$, the function $\tilde{h}_{l,1}$ simplifies to

$$\tilde{h}_{l^*,1} = \frac{p^2-1}{2(p+1)^2}. \tag{3}$$

Proposition 1 only refers to sequences which uses the same treatment in the first and in the last period. Therefore, it is necessary to confirm that starting with a different treatment than the one used in the last period cannot improve the design in the case of an odd number of periods.

Proposition 2 *Let p be an odd number. Then,*

$$h_{l,2}(x^*, y^*) < h_{l^*,1}(x^*, y^*),$$

where l^* is the equivalence class with a sequence s with n_1^* and $\tilde{n}_1^*$ as given in Proposition 1. The values x^* and y^* are the real numbers which were also identified in Proposition 1.

Proof Evaluating the function $h_{l,2}$ at x^*, y^* leads to

$$\tilde{h}_{l,2} = \frac{2p\tilde{n}_1(l) + p^3 - p^2 - p - 1}{2p(p+1)^2}.$$

It is necessary to compare the value of $\tilde{h}_{l,2}$ to the value of the function $\tilde{h}_{l^*,1}$ which is given in Equation (3) at x^*, y^* and therefore to confirm that

$$\begin{aligned}\frac{p^2-1}{2(p+1)^2} > \frac{2p\tilde{n}_1(l) + p^3 - p^2 - p - 1}{2p(p+1)^2} &\Leftrightarrow 0 > 2p\tilde{n}_1(l) - p^2 - 1 \\ &\Leftrightarrow \tilde{n} < \frac{p}{2} + \frac{1}{2p}\end{aligned}$$

The number of mixed-carryovers, $\tilde{n}_1(l)$, cannot be larger than $\frac{p}{2}$. Therefore, this condition always holds true. □

Combining Proposition 1 and Proposition 2, we have now identified optimal sequences in designs with an odd number of periods. Next, we focus on designs with an even number of periods.

Proposition 3 *Let p be an even number and $p > 2$. Then, for sequences with different treatments in the first and last period (start with R, end with T), the upper boundary of the trace of $\mathbf{C}_d$ is reached for*

$$x^* = \frac{p-1}{p},\ y^* = -\frac{1-p}{p}$$

and

$$n_1^*(s) = \frac{p}{2},\ \tilde{n}_1^*(s) = \frac{p-2}{2}.$$

Starting with the same treatment that is used in the last period cannot increase this value, i.e.,

$$h_{l,1}(x^*, y^*) < h_{l^*,2}(x^*, y^*).$$

Proof Very similar to the proof of Proposition 1 and 2. See the Appendix for details.

3.3 Example

The analysis in the previous section showed that for designs with an odd number of periods, only sequences should be used in which the subjects switch after each period and the starting and ending period should be the same. For example for 5 periods, this restricts our attention to sequences $s = TRTRT$, $s' = RTRTR$. If the number of periods is even, the subjects should again switch between T and R after each period, but the treatment in the first and in the last period should be different. For 6 periods, this leads to the sequences $s = RTRTRT$, $s' = TRTRTR$. As optimal designs have to be dual-balanced, the sequences s and s' have to be appear equally often.

4 Study designs in practice

It is interesting to see how these optimized study designs relate to study designs which are applied for switchability assessment of biosimilars in practice. As an example, we discuss the study design of the EGALITY study (Griffiths et al, 2017). It is important to note that the aim of this section is not to criticize the study design of the EGALITY study: a study design is always optimized for a specific analysis method and the analysis of mixed-carryover effects, for which the designs we identified in this paper are optimal, was not one of the analyses conducted for the EGALITY study. We also acknowledge constraints in the choice of designs in practice due to, for example, operational constraints and expectations of health authorities (e.g., the recommended design to assess interchangeability (the term used by the health authority in the US for switchability) as stated in the FDA's draft guidance (FDA, 2017)). However, we find it nonetheless interesting to compare the theoretical optimal designs with the designs applied in practice.

The EGALITY study was conducted in patients with moderate to severe chronic plaque-type psoriasis. 531 patients were randomised 1:1 to Erelzi® (the biosimilar) or Enbrel® (the reference product). The study design is given in Table 1. In the following, we ignore that the time intervals between the follow-ups are not equidistant and consider each follow-up assessment as one period.

Table 1 Study design of the EGALITY study (Griffiths et al, 2017). T is the test treatment, R is the reference treatment.

Sequence/Follow-up	Week 12	Week 18	Week 24	Week 30	Week 52
1	T	T	T	T	T
2	R	R	R	R	R
3	R	T	R	T	T
4	T	R	T	R	R

The trace of the information matrix for the study design used in the EGALITY study is 0.8636 (assuming one subject per sequence). The optimal design identified in this paper had a much larger trace of the information matrix (1.3333) and is clearly superior for the estimation of mixed-carryover effects. The much lower value of the criterion for the EGALITY study is due to the fact that half of the subjects (the subjects in the non-switching sequences) do not contribute much to the precision of the estimation of the mixed-carryover effects. Although the EGALITY study was clearly not tailored for the estimation of mixed-carryover effects, this shows the potential of using optimized designs.

5 The inclusion of dummy treatments

For the designs derived in Kunert and Stufken (2008), it was shown that adding additional periods does not lead to a relevant improvement of the design, but the inclusion of dummy treatments (for example: a placebo treatment or no treatment) increased the precision of the estimator. In this section, we investigate if periods with dummy treatments can also improve the design in our setting. The dummy treatment is denoted by N. For that, we focus on a design with five periods. In this case, $s = RTRTR$, $s' = TRTRT$ are the optimal sequences that were derived in Section 3.3.

We focus on the linear model that was already introduced in Section 2, i.e., the response $y_{u,r}$ of subject u in period r can be written by

$$y_{u,r} = \begin{cases} \alpha_u + \beta_r + \tau_{d(u,r)} + \rho_{d(u,r-1)} + e_{u,r} & \text{if } d(u,r) \neq d(u,r-1) \\ \alpha_u + \beta_r + \tau_{d(u,r)} + \chi_{d(u,r-1)} + e_{u,r} & \text{if } d(u,r) = d(u,r-1) \end{cases},$$

where $d(u,r)$ gives the treatment of subject u in period r ($1 \leq u \leq n; 1 \leq r \leq p$), α_u is the subject effect of subject u and β_r is the period effect in period r. τ_i is the direct effect of treatment i with the levels τ_T (treatment T), level τ_R (treatment R) and level τ_N (treatment with the dummy treatment N). The mixed-carryover is ρ_i which is present in the model for the switches from T to R, from T to N (level ρ_T) and from R to N or from R to T (level ρ_R). If a subject receives the dummy treatment (no treatment) in period $k-1$ and T or R in period k, no mixed-carryover effect is assumed because the dummy treatment corresponds to "no treatment" or "placebo" and the situation is therefore comparable to the situation in the first period. The self-carryover effect is denoted as χ_i. More concretely, the level χ_T (treatment in period k and period $k-1$ with T) and χ_R (treatment in period k and period $k-1$ with R) are used. No self-carryover effect is introduced for the dummy treatment. In summary, the mixed and self-carryover effects are defined as:

$$\rho_i = \begin{cases} \rho_T & \text{if } d(u,r-1) = T \text{ and } d(u,r) \neq T, \\ \rho_R & \text{if } d(u,r-1) = R \text{ and } d(u,r) \neq R, \\ 0 & \text{if } d(u,r-1) = d(u,r) \text{ or } d(u,r-1) = N \end{cases}$$

$$\chi_i = \begin{cases} \chi_T & \text{if } d(u,r-1) = d(u,r) = T, \\ \chi_R & \text{if } d(u,r-1) = d(u,r) = R, \\ 0 & \text{if } d(u,r-1) \neq d(u,r) \text{ or } d(u,r-1) = d(u,r) = N. \end{cases}$$

As this investigation is only for illustration, we will use the optimal design with five periods without the dummy treatment that was stated above and include systematically the dummy treatment. Since we start with a fixed sequence, it is important to note that this approach might not lead to the optimal design for a study with a dummy treatment.

Table 2 shows the results for the inclusion of one dummy treatment. All shown combinations led to a completely symmetric information matrix and are therefore in

the class of designs that can be universally optimal. The value of the optimality criterion for the optimal design without any dummy treatment was 0.6667 for one subject per sequence. Table 2 shows that this value can be improved with a dummy treatment: for all sequences apart from the one in which the dummy treatment is included in the first period, the value of the optimality criterion is increased. Interestingly, the criterion for Sequence 2 and 4 is lower than for 3 and 5. While this might be unexpected at first sight, it can easily be explained: in Sequences 2 and 4 we replaced the treatment which was already given more often with the dummy treatment (e.g., in Sequence 2 in Table 2 we replaced an R with an N, therefore subjects received T three times, R once, N once instead of T three times, R twice). This unbalance leads to a lower value of the criterion.

Including two dummy treatments still increases the value of the criterion (2.2857), but the value decreases if more than two dummy treatments are included. For study designs with more periods, the improvement is even higher: for example, for nine periods, the largest optimality criterion is 4.6364 (with inclusion of four dummy treatments) compared to 0.8 without any dummy treatments.

Table 2 Study designs with one dummy treatment per sequence. The criterion is the trace of the information matrix. We only give one sequence, but the design consists of another sequence with is dual-balanced in terms of T and R, but copies the dummy treatment, e.g., for the sequence TRNTR, the second sequence would be RTNRT. The first sequence is the sequence without any dummy treatment that serves as the reference.

No.	Sequence	Criterion
-	TRTRT	0.6667
1	NRTRT	0.55
2	TNTRT	1
3	TRNRT	1.8
4	TRTNT	1
5	TRTRN	1.75

6 Conclusion

In this paper, we showed that for the estimation of mixed-carryover effects it is optimal to use a study design in which the subjects switch between T and R after each period. If an odd number of periods is used in the design, it is optimal to start with the same treatment that is used in the last period. In the situation with an even number of periods, the treatment in the last and in the first period has to be different. The rational for this is that as many mixed-carryover effects as possible should be included in the design and this number is maximised for the described choice.

We also showed that if the aim of a study is to estimate mixed-carryover effects, designs which were already used in practice (e.g., the EGALITY study) could be greatly improved by using our optimized designs. However, it is important to keep in mind that so far, the estimation of mixed-carryover effects was not among the objec-

tives of the studies conducted in practice.

Last, we showed that the estimation could be improved if one is willing to include dummy treatments (i.e., periods without any treatment or placebo). This finding is similar to the one in Kunert and Stufken (2008). It is important to point out that these results might not be relevant for an application in biosimilar development: since biosimilars are used for treating serious and often chronic diseases, it might not be possible to stop the treatment even if this provides an advantage in terms of precision of the estimator. However, if the methodology is applied to other areas (e.g., sensory trials that are mentioned in Kunert and Stufken (2002)) or if the studies are undertaken in healthy volunteers, it might be possible to consider the use of dummy treatments if this provides an advantage from a statistical point of view. It is also important to note that even though our results show that – from a theoretical point of view – the inclusion of dummy treatments improves the performance of the design, it is important to point out that this conclusion is purely based on the considered model and the main assumption, which is that after one period without treatment (the dummy treatment – a wash-out period), no carryover effects are observed. Before using the results as a justification why the inclusion of dummy treatments is necessary, it is important to justify that the model assumptions are met.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Afsarinejad K, Hedayat A (2002) Repeated measurements designs for a model with self and simple mixed carryover effects. J Stat Plan Inference 106(1):449–459

CHMP (2014) Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/01/WC500180219.pdf. Accessed 02 Nov 2015

FDA (2017) Considerations in demonstrating interchangeability with a reference product. Available at https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM537135.pdf. Accessed 02 Nov 2017

Finnish Medicines Agency (2015) Interchangeability of biosimilars – position of Finnish Medicines Agency Fimea. Available at https://www.fimea.fi/documents/542809/838272/29197_Biosimilaarien_vaihtokelpoisuus_EN.pdf. Accessed 24 Mar 2017

Generics and Biosimilars Initiative (2017a) Biosimilars approved in Europe. Available at http://www.gabionline.net/Biosimilars/General/Biosimilars-approved-in-Europe. Accessed 01 Nov 2017

Generics and Biosimilars Initiative (2017b) Biosimilars approved in the US. Available at http://gabionline.net/Biosimilars/General/Biosimilars-approved-in-the-US. Accessed 01 Nov 2017

Griffiths C, Thai D, Gerdes S, Arenberger P, Pulka G, Kingo K, Weglowska J, the EGALITY study group, Hattebuhr N, Poetzl J, Woehling H, Wuerth G, Afonso M (2017) The EGALITY study: a confirmatory, randomized, double-blind study comparing the efficacy, safety and immunogenicity of GP2015, a proposed etanercept biosimilar, vs. the originator product in patients with moderate-to-severe chronic plaque-type psoriasis. Br J Dermatol 176(4):928–938

Health Products Regulatory Authority (2015) Guide to biosimilars for healthcare professionals and patients. Available at https://www.hpra.ie/docs/default-source/publications-forms/guidance-documents/guide-to-biosimilars-for-healthcare-professionals-and-patients-v2.pdf?sfvrsn=18. Accessed 24 Mar 2017

Kiefer J (1975) Construction and optimality of generalized Youden designs. A Survey of Statistical Designs and Linear Models pp 333–353

Kunert J (1983) Optimal design and refinement of the linera model with applications to repeated measurements designs. Ann Stat 11(1):247–257

Kunert J, Martin R (2000) On the determination of optimal designs for an interference model. Ann Stat 28(6):1728–1742

Kunert J, Stufken J (2002) Optimal crossover designs in a model with self and mixed carryover effects. J Am Stat Assoc 97(459):898–906

Kunert J, Stufken J (2008) Optimal crossover designs for two treatments in the presence of mixed and self-carryover effects. J Am Stat Assoc 103(484):1641–1647

Mielke J, Jilma B, Koenig F, Jones B (2016) Clinical trials for authorized biosimilars in the European Union: a systematic review. Br J Clin Pharmacol 82(6):1444–1457

Appendix

A. Derivation of a simpler expression for h_l

Using the results obtained by Kunert and Stufken (2002) and the simplifications in the case of two treatments as described by Kunert and Stufken (2008), easier formulas for $c_{ij}(l)$ can be derived:

$$c_{11}(l) := \frac{t-1}{t}\sum_j \tilde{n}_j(l) - \frac{1}{p}\sum_j \tilde{n}_j^2(l) + \frac{1}{pt}\left(\sum_j \tilde{n}_j(l)\right)^2$$
$$= \begin{cases} \tilde{n}_1(l) & \text{if sequence starts with T} \\ \tilde{n}_1(l) + \frac{p-1}{2p} & \text{if sequence starts with R} \end{cases},$$
$$c_{12}(l) := -\frac{1}{p}\sum_j \tilde{n}_j(l) n_j(l)$$
$$= \begin{cases} -\tilde{n}_1(l) & \text{if sequence starts with T} \\ -\tilde{n}_1(l) - \frac{p-n_1(l)}{p} & \text{if sequence starts with R} \end{cases},$$

$$c_{13}(l) := -\frac{1}{p}\sum_j \tilde{n}_j(l)\,(n_j(l)-\tilde{n}_j(l)-t_{pj}(l)) + \frac{1}{pt}\left(\sum_j \tilde{n}_j(l)\right)\left(p-1-\sum_j \tilde{n}_j(l)\right)$$
$$= \begin{cases} 0 & \text{if sequence starts with T} \\ -\frac{p-2n_1(l)}{2p} & \text{if sequence starts with R} \end{cases},$$
$$c_{22}(l) := p - \frac{1}{p}\sum_j n_j^2(l) = \frac{2n_1(l)(p-n_1(l))}{p},$$
$$c_{23}(l) := p-1-\sum_j \tilde{n}_j(l) - \frac{1}{p}\sum_j (n_j(l)(n_j(l)-\tilde{n}_j(l)-t_{pj}(l)))$$
$$= \begin{cases} -\tilde{n}_1(l)-1+\frac{2n_1(l)(p-n_1(l))}{p}+\frac{n_1(l)}{p} & \text{if sequence starts with T} \\ -\tilde{n}_1(l)-1+\frac{2n_1(l)(p-n_1(l))}{p} & \text{if sequence starts with R} \end{cases},$$
$$c_{33}(l) := \frac{t-1}{t}\left(p-1-\sum_j \tilde{n}_j(l)\right) - \frac{1}{p}\sum_j \tilde{n}_j^2(l) + \frac{1}{pt}\left(p-1-\sum_j \tilde{n}_j(l)\right)^2$$
$$= \begin{cases} \frac{p-2\tilde{n}_1(l)-1}{2} - \frac{(p-2n_1(l)+1)^2}{2p} & \text{if sequence starts with T} \\ \frac{p-2\tilde{n}_1(l)-2}{2} - \frac{(p-2n_1(l))^2}{2p} & \text{if sequence starts with R} \end{cases},$$

Plugging these results into the formula for h_l lead to the simpler expression.

B. Proof of Proposition 3

The proof follows the lines of the proofs of Proposition 1 and 2. Again, we first choose values $n_1(s)$ and $\tilde{n}_1(s)$ we consider to be optimal, derive the values x^*, y^* which minimises $h_{l,2}$ and show that the choice of $n_1(s)$ and $\tilde{n}_1(s)$ was indeed optimal. Last, we show that starting with the same treatment that is also used in the last period cannot improve the design. In the first step, we assume that

$$n_1(s) = \frac{p}{2} \text{ and thus } \tilde{n}_1(s) = \frac{p-2}{2}$$

is the optimal choice. Next, the optimal values for x, y are derived by calculating the partial derivations of $h_{l,2}$ with respect to x and y:

$$\frac{dh_{l,2}(x,y)}{dx} = -2\tilde{n}_1(s) - 2y\tilde{n}_1(s) - \frac{4x}{p}n_1(s)^2$$
$$- \frac{4y}{p}n_1(s)^2 + \frac{2n_1(s)}{p} + 4yn_1(s) + 4xn_1(s) - 2 - 2y,$$
$$\frac{dh_{l,2}(x,y)}{dy} = -2y\tilde{n}_1(s) - 2x\tilde{n}_1(s) - \frac{4y}{p}n_1(s)^2$$
$$- \frac{4x}{p}n_1(s)^2 + \frac{2}{p}n_1(s) + 4xn_1(s) + 4yn_1(s) - 1 - 2y - 2x.$$

Setting the equations to 0, we find that

$$x^* = \frac{1+2\tilde{n}}{2\tilde{n}+2} = \frac{p-1}{p}.$$

It is important to note that all values are optimal for y. For the rest of the proof, we set

$$y^* := \frac{2-2p}{2p} = \frac{1-p}{p} = -x^*.$$

We note with the same arguments that we used in the proof of Proposition 1 that we have indeed identified a global minimum. The choice of y again makes the function $\tilde{n}_1(l)$ independent of the number of appearances of the treatment (see below). Plugging x^* and y^* into the function $h_{l,2}$ leads to

$$\tilde{h}_{l,2} := h_{l,2}(x^*, y^*) = \frac{\tilde{n}_1(s)}{p^2} + \frac{p^2-3p+2}{2p^2}.$$

For showing that the choice of $n_1^*(s)$ and $\tilde{n}_1^*(s)$ is optimal, we use exactly the same arguments as in the proof of Proposition 1 and confirm that other choices of $\tilde{n}_1(s)$ and $n_1(s)$ cannot increase the value of $h_{l,2}$ at x^* and y^* because function $h_{l,2}(x^*, y^*)$ is optimal if as many mixed-carryover effects as possible are used. The maximum number of mixed-carryover effects is $\tilde{n}_1(s) = \frac{p-2}{2}$ and this leads directly to $n_1(s) = \frac{p}{2}$. Function $\tilde{h}_{l,2}$ simplifies for this choice of mixed-carryover effects to

$$\tilde{h}_{l^*,2} = \frac{p^2-2p}{2p^2}. \tag{4}$$

In the last step, it is necessary to verify that starting with a different treatment than the one used for the last sequence is optimal. This is confirmed by showing that the value of $h_{l,1}(x^*, y^*)$ is smaller than the obtained optimum. Plugging the values x^* and y^* into the function leads to

$$h_{l,1}(x^*, y^*) = \frac{\tilde{n}_1(s)}{p^2} + \frac{(p-3)^3}{2p^3}.$$

Comparing $h_{l,2}(x^*, y^*)$ to Equation (4) leads to the conclusion that $h_{l,2}(x^*, y^*)$ is larger than $h_{l,1}(x^*, y^*)$ if

$$7p^2 + 27 - 27p - 2\tilde{n}_1(s)p > 0.$$

This condition holds true for $p > 2$ for all possible choices of $\tilde{n}_1(s) = 1, ..., \frac{p}{2}$. Therefore, for a design with an even number of periods, it is optimal to start with R and end with T or vice versa and to use

$$n_1(s) = \frac{p}{2} \text{ and } \tilde{n}_1(s) = \frac{p-2}{2}$$

as the number of appearances of treatment T and the number of mixed-carryovers, respectively. □