

The impact of baseline covariates on the efficiency of statistical analyses of crossover designs

Zhiwu Yan^{*†}

We investigate the impact of baseline covariates on the efficiency of statistical analyses of crossover designs. For practical considerations, we contemplate two different baseline methods: study baselines and period-dependent baselines. For each baseline method, we establish analytical upper bounds for the relative efficiency of a large class of crossover designs, the totally balanced designs, under a model with the baseline covariates as compared with the model without the baseline covariates. We present numerical details based on these bounds for assorted scenarios and reveal implications of these results. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: baseline covariate; crossover design; Fisher information matrix; relative efficiency; totally balanced design

1. Introduction

Crossover designs have applications in a wide range of sciences and research areas, such as clinical trials, pharmaceutical studies, psychological experiments, agriculture field trials, and animal feeding experiments. The advantage of such designs is that the study subjects become their own controls, thereby reducing the error variance.

In many experimental situations, notably in clinical trials and pharmaceutical studies, the treatment effects are often adjusted for the baseline values of the response (outcome) variable. It is well known that inclusion of baseline covariates in a parallel study can in general (i.e., when the baselines and responses are moderately to strongly correlated) significantly increase the efficiency of statistical analyses. In the context of crossover designs, however, the impact of baseline covariates on the efficiency of statistical analyses has not been fully appreciated, and quite often, the analysis of covariance has been abused in practice. With the increasing popularity of crossover designs in the past three decades, various baseline approaches for crossover designs have been discussed in the literature, such as Hills and Armitage [1], Wallenstein [2], Fleiss *et al.* [3], Willan and Pater [4], Kenward and Jones [5], Senn [6], Jones and Kenward [7], and the recent work by Liang and Carrière [8] and Kenward and Roger [9]. Some recommended that the change from baseline be modeled, whereas others argued that baselines should be used as additional responses. Although most statisticians would agree that an appropriate way to incorporate the baseline information is the analysis of covariance, baseline-related issues for crossover designs have rarely been well addressed from such a perspective. The mathematical hurdles to obtain an exact inference based on weighted least squares analyses for crossover designs are perhaps the primary reason for which the analysis of covariance approach has been more or less 'bypassed' in the literature. Not surprisingly, most research carried out along this line has relied heavily on computer simulations.

This article is not intended to justify or promote the analysis of covariance approach for crossover designs. Instead, we will argue, through the analysis of covariance, that baseline information in general makes limited additions to the efficiency of statistical analyses, especially for highly efficient crossover designs. We hope that the message we deliver in this article can shed fresh light on other baseline

Department of Biostatistics, Allergan, Inc., 2525 Dupont Drive, Irvine, CA 92612, U.S.A.

^{*}Correspondence to: Zhiwu Yan, Department of Biostatistics, Allergan, Inc., 2525 Dupont Drive, Irvine, CA 92612, U.S.A.

[†]E-mail: yan_zhiwu@allergan.com

approaches for crossover designs as well. To this end, we compare the efficiency of crossover designs between models with and without the baseline covariates. Depending upon whether study baselines or period-dependent baselines are fitted, analytical upper bounds of the relative efficiency for treatment effects are obtained for totally balanced designs in which the number of periods (p) does not exceed the number of treatments (t). The impact of the baseline covariates on the efficiency of statistical analyses is then examined.

In Section 2, we introduce the response models with and without the baseline covariates. Section 3 provides the upper bounds of the relative efficiency for totally balanced designs. For the arguably most important case where $p = t$, we also present upper bounds that are independent of the within-subject correlation. With only a few exceptions, we tabulate numerical results for each of the scenarios where $2 \leq p \leq t \leq 9$ and discuss the implications of these results. We conclude in Section 4 with a discussion and summary of our findings and provide the technical details for our main results in the Appendix.

2. The response models

Throughout this article, we shall assume linear mixed-effects models for the responses of interest. When baseline covariates are not fitted, our model is precisely the same as that studied by Hedayat *et al.* [10]. When baseline covariates are accounted for, two different baseline methods are to be considered. The first method uses the baselines obtained before the first treatment period (the so-called study baselines) for each subsequent treatment period, whereas the second uses the baselines obtained for their own treatment period (the so-called period-dependent baselines). In addition, we will assume a joint multivariate normal distribution of the baselines and responses, which is critical in deriving the covariance matrix of the responses conditioning on the baseline covariates.

2.1. Without baseline covariates

Without baseline covariates, we assume that the response y_{ks} obtained for subject s in period k , where $s \in \{1, \dots, n\}$ and $k \in \{1, \dots, p\}$, can be expressed as

$$y_{ks} = \mu + \alpha_k + \xi_s + \tau_{d(k,s)} + \gamma_{d(k-1,s)} + \varepsilon_{ks}, \quad (1)$$

where μ is the general mean, α_k is the k th period effect, ξ_s is the s th subject effect, $d(k, s) \in \{1, \dots, t\}$ denotes the treatment assigned to subject s in period k , $\tau_{d(k,s)}$ is the direct effect due to treatment $d(k, s)$, $\gamma_{d(k-1,s)}$ is the (first-order) carryover or residual effect due to treatment $d(k-1, s)$ assigned to subject s in the preceding period (by convention $\gamma_{d(0,s)} = 0$), and ε_{ks} is the measurement error. We also assume that ξ_s 's are independent and identically distributed random variables with mean zero and a common variance σ_s^2 and that ε_{ks} 's are independent and identically distributed random variables with mean zero and a common variance σ_e^2 . In addition, we assume the ξ_s 's and ε_{ks} 's to be independent. The assumptions on the random effects in model (1) imply that the responses have a block-diagonal, compound-symmetry covariance matrix, with each block being the covariance matrix for an individual subject.

2.2. With baseline covariates

When baseline covariates are accounted for, we assume the response models (conditional models) based on the two aforementioned baseline methods to be

$$y_{ks}|z_{1s} = \mu + \alpha_k + \xi_s + \tau_{d(k,s)} + \gamma_{d(k-1,s)} + \beta_k z_{1s} + \varepsilon_{ks} \quad (2)$$

and

$$y_{ks}|z_{ks} = \mu + \alpha_k + \xi_s + \tau_{d(k,s)} + \gamma_{d(k-1,s)} + \beta_k z_{ks} + \varepsilon_{ks}, \quad (3)$$

where z_{ks} is the baseline value obtained for subject s in period k ($k = 1$ corresponds to the study baseline), β_k is the corresponding regression coefficient, and all other effects and assumptions are analogous to those presented in model (1). However, owing to potential variability accounted for by the baseline covariates, the variances (conditional) of the random effects ξ_s and ε_{ks} in the conditional models (2) and (3) are expected to be no greater (and quite often smaller) than those (σ_s^2 and σ_e^2) in model (1). To address this difference and to be consistent with model (1), we assume that at the design stage (i.e., before data collection), the combined data of baselines and responses follow a multivariate normal

distribution with the same covariance structure as that for model (1) if ignoring the matrix dimensions. Such a distributional assumption will enable us to obtain the covariance matrices for the conditional models (2) and (3), to be shown in the Appendix. Following the convention in analysis of covariance, the assumption of baseline comparability (i.e., the expected values of baselines for all treatment groups are the same for each period) is also imposed in this article. In conjunction with the assumption of a joint normal distribution, this requirement ensures that both models (2) and (3) provide unbiased predictors of the responses and thus unbiased estimators of treatment effects.

3. The relative efficiency

When unbiased inferences are made based on two statistical procedures (a procedure can be an estimator, a model, a design, etc.), relative efficiency is a measure used to evaluate how 'good' (efficient) one procedure is with respect to the other one. The actual definition of such a measure varies depending on the context, but in all cases, it reflects the sample size requirement for obtaining the same precision (and thus power) of inferences under the two procedures. When a single parameter of interest is involved, for example, the relative efficiency of an unbiased estimator T_1 compared with another unbiased estimator T_2 is defined to be the reciprocal of the ratio of their variances. An efficiency of 2 for T_1 relative to T_2 would suggest that the sample size for T_2 should be twice (or approximately twice) as much as that for T_1 in order to achieve the same precision for the two estimators. To avoid confusion, the term 'relative efficiency' used throughout this article should be interpreted as relative efficiency for treatment comparisons.

For the purpose of this article, we shall be particularly interested in a large class of crossover designs that are extensively used in applications, namely the totally balanced designs defined by Kunert and Stufken [11]. Such designs have strong practical appeal and have been revealed to be optimal or highly efficient in most situations under model (1) [10]. Many useful crossover designs (such as balanced uniform designs [12], which in turn include the well-known Williams designs as a special case, and designs constructed from mutually orthogonal Latin squares) belong to such a class. For those who are not familiar with totally balanced designs obtained from mutually orthogonal Latin squares, an example of such a design is given as follows, where the rows correspond to the periods, the columns correspond to the subjects/sequences, and the entries 1, 2, 3, 4 stand for the treatments.

$$\begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 4 \\ 2 & 3 & 4 & 1 & 3 & 4 & 1 & 2 & 4 & 1 & 2 & 3 \\ 4 & 2 & 3 & 3 & 4 & 1 & 4 & 1 & 2 & 2 & 3 & 1 \end{bmatrix}$$

We shall now present our main results for comparing the efficiency of totally balanced designs between the model (1) and each of the models (2) and (3). We evaluate the relative efficiency based on the universal optimality criteria defined by Kiefer [13], which is particularly relevant when all treatment effects are of equal interest. The definition of the universal optimality is mathematically involved and thus omitted here. The interested reader is referred to Kiefer [13] for more details.

Theorem 1

Let RE^s and RE^{pd} be the relative efficiency of a totally balanced design with $p \leq t$ under models (2) and (3), respectively, as compared with model (1); then

$$RE^s \leq g_{(t,p)}(\theta) = \frac{f_{(t,p)}(\theta(1+\theta)^{-1})}{f_{(t,p)}(\theta)} \quad (4)$$

and

$$RE^{pd} \leq h_{(t,p)}(\theta) = \frac{f_{(t,p)}(\theta(1+p\theta)^{-1})}{f_{(t,p)}(\theta)}, \quad (5)$$

where $\theta = \sigma_s^2 / \sigma_\varepsilon^2$ and

$$f_{(t,p)}(x) = (p-1) + \frac{(t-p)}{t(1+px)} - \frac{(p-1)(tx+1)^2}{t^2(1+px)^2 \left[1 - \frac{1}{tp} - \frac{tpx+p-1}{tp(1+px)} \right]}.$$

Moreover, if $p = t$, then

$$RE^s \leq 1 + \frac{1}{(t^2 - t - 2)(t^3 - 2t)} \quad (6)$$

and

$$RE^{pd} \leq 1 + \frac{1}{(t^2 - t - 2)(2t^2 - 2t - 1)}. \quad (7)$$

The proof of Theorem 1 is given in the Appendix.

The general upper bounds given by (4) and (5) have to be evaluated numerically for a given θ . The results given by (6) and (7) for the special case $p = t$ are quite informative because they do not depend on θ . For example, if $p = t = 3$, then $RE^s \leq 1 + 1/84 \approx 1.012$ and $RE^{pd} \leq 1 + 1/44 \approx 1.023$, and if $p = t = 4$, then $RE^s \leq 1 + 1/560 \approx 1.002$ and $RE^{pd} \leq 1 + 1/230 \approx 1.004$. It is clear that these bounds monotonically and quickly tend to 1 as t increases.

Tables I and II show the numerical results (rounded to the third decimal position) of the upper bound functions $g_{(t,p)}(\theta)$ and $h_{(t,p)}(\theta)$ for totally balanced designs with $2 \leq p \leq t \leq 9$. We exclude the cases where $t = 6$ with $p \neq 6$ because it is not known if a totally balanced design exists for such

Table I. The upper bound ($g_{(t,p)}(\theta)$) of the relative efficiency RE^s .							
p	t	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.99$
2	2	1.333	1.563	1.961	2.778	5.263	50.251
	3	1.229	1.347	1.510	1.743	2.093	2.593
	4	1.220	1.331	1.480	1.688	1.989	2.400
	5	1.218	1.325	1.471	1.671	1.959	2.346
	7	1.216	1.322	1.465	1.660	1.939	2.310
	8	1.215	1.321	1.463	1.658	1.935	2.303
	9	1.215	1.321	1.462	1.656	1.932	2.298
3	3	1.007	1.009	1.009	1.010	1.011	1.012
	4	1.042	1.051	1.059	1.067	1.074	1.080
	5	1.058	1.072	1.085	1.097	1.108	1.118
	7	1.074	1.092	1.110	1.127	1.144	1.158
	8	1.079	1.098	1.118	1.136	1.154	1.170
	9	1.082	1.103	1.123	1.143	1.162	1.179
4	4	1.001	1.001	1.001	1.002	1.002	1.002
	5	1.016	1.018	1.020	1.022	1.023	1.025
	7	1.030	1.035	1.040	1.044	1.047	1.050
	8	1.035	1.040	1.046	1.050	1.054	1.057
	9	1.038	1.044	1.050	1.055	1.060	1.063
5	5	1.000	1.000	1.000	1.000	1.000	1.000
	7	1.013	1.014	1.016	1.017	1.018	1.018
	8	1.016	1.018	1.020	1.022	1.023	1.024
	9	1.019	1.022	1.024	1.025	1.027	1.028
6	6	1.000	1.000	1.000	1.000	1.000	1.000
	7	1.004	1.005	1.005	1.005	1.006	1.006
	8	1.008	1.008	1.009	1.009	1.010	1.010
	9	1.010	1.011	1.012	1.012	1.013	1.014
7	7	1.000	1.000	1.000	1.000	1.000	1.000
	8	1.003	1.003	1.003	1.003	1.003	1.004
	9	1.005	1.005	1.006	1.006	1.006	1.006
8	8	1.000	1.000	1.000	1.000	1.000	1.000
	9	1.002	1.002	1.002	1.002	1.002	1.002
9	9	1.000	1.000	1.000	1.000	1.000	1.000

Table II. The upper bound ($h_{(t,p)}(\theta)$) of the relative efficiency RE^{pd} .

p	t	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$	$\rho = 0.99$
2	2	1.500	1.818	2.361	3.462	6.786	66.779
	3	1.350	1.515	1.739	2.053	2.519	3.181
	4	1.337	1.492	1.697	1.978	2.380	2.923
	5	1.333	1.485	1.684	1.955	2.339	2.851
	7	1.331	1.480	1.676	1.940	2.312	2.803
	8	1.330	1.479	1.674	1.937	2.307	2.794
	9	1.330	1.478	1.673	1.935	2.303	2.787
3	3	1.014	1.016	1.018	1.020	1.021	1.023
	4	1.087	1.103	1.118	1.131	1.144	1.154
	5	1.121	1.145	1.167	1.188	1.208	1.224
	7	1.153	1.186	1.217	1.246	1.273	1.297
	8	1.162	1.197	1.231	1.262	1.293	1.318
	9	1.168	1.205	1.241	1.275	1.307	1.335
4	4	1.003	1.003	1.004	1.004	1.004	1.004
	5	1.041	1.046	1.051	1.054	1.058	1.060
	7	1.079	1.090	1.099	1.107	1.114	1.120
	8	1.089	1.102	1.113	1.123	1.131	1.138
	9	1.097	1.111	1.124	1.134	1.144	1.151
5	5	1.001	1.001	1.001	1.001	1.001	1.001
	7	1.040	1.044	1.047	1.050	1.052	1.054
	8	1.051	1.056	1.061	1.064	1.067	1.070
	9	1.059	1.066	1.071	1.075	1.079	1.082
6	6	1.000	1.001	1.001	1.001	1.001	1.001
	7	1.016	1.017	1.018	1.019	1.020	1.021
	8	1.027	1.030	1.031	1.033	1.034	1.035
	9	1.036	1.039	1.041	1.043	1.045	1.046
7	7	1.000	1.000	1.000	1.000	1.000	1.000
	8	1.011	1.012	1.013	1.013	1.014	1.014
	9	1.020	1.021	1.022	1.023	1.024	1.025
8	8	1.000	1.000	1.000	1.000	1.000	1.000
	9	1.009	1.009	1.010	1.010	1.010	1.010
9	9	1.000	1.000	1.000	1.000	1.000	1.000

situations. Here $\rho = \theta(1 + \theta)^{-1}$ is the within-subject correlation varying from 0.5 to 0.99, which covers a reasonable range for most practical applications. We observe that both tables show similar patterns of changes in the functions $g_{(t,p)}(\theta)$ and $h_{(t,p)}(\theta)$, and because $g_{(t,p)}(\theta) \leq h_{(t,p)}(\theta)$ (because $f_{(t,p)}(x)$ is decreasing in x), all the entries in Table I are consistently smaller than those in Table II. This difference, however, is almost negligible if $p (\neq 2)$ and t are close or if p is relatively large.

For $p > 2$, we note that if p and t are close enough, then the potential increase in efficiency due to inclusion of baseline covariates is typically extremely limited, which is more so when p is relatively large. Even otherwise, the increase in efficiency is at most moderate ($RE^s \leq 1.179$ and $RE^{pd} \leq 1.335$), and this may occur only when t is much bigger than p while p is small (e.g., $p = 3$). It is conceivable that if p and t are very close, then most (if not all) of the information about the treatment effects for a totally balanced design under model (1) is provided by within-subject comparisons (in other words, the design is highly efficient), and thus, very little or no additional information can be obtained by inclusion of baseline covariates. On the other hand, if t is relatively large compared with p , then more between-subject information is expected to be used for treatment comparisons, therefore leaving room for using baseline information.

The results for $p = 2$ appear to be quite different; however, the interpretations of these results are straightforward. When $p = t = 2$, the totally balanced design is the standard 2×2 crossover design. It is well known that if carryover effects are accounted for, then any unbiased estimator of the treatment effects under such a design has to rely on the data from the first period only. In such a situation, the crossover design does not provide any within-subject information for the treatment effects and therefore performs just like a parallel design. The results for $p = t = 2$ simply reveal the fact that in theory the efficiency increase for a parallel design can be arbitrarily large (as ρ approaches to 1) once baseline covariates are fitted. When $p = 2$ and $t > 2$, a totally balanced design does use the data from both periods; however, it is also well known that such designs are in general not efficient owing to the presence of carryover effects. It is therefore not a surprise that baseline covariates could make a big impact on the efficiency of statistical analyses.

4. Discussion

We have delivered an important message in this short article: when a crossover design is highly efficient, that is, when most of the information about the treatment effects is provided by within-subject comparisons, the baseline covariates in general have very limited impact on the efficiency of statistical analyses. For a less efficient crossover design, however, baseline covariates could turn out to be important. Even though these findings are based on models assuming carryover effects, we are quite confident that such a statement remains valid for other different models as well. To support this optimism, we shall give two illustrating examples while considering models (1) and (3) without the carryover effects. The first example is the standard 2×2 crossover design, which is highly efficient without the carryover effects (indeed the treatment effects are entirely based on the within-subject comparisons). It can be shown that $RE^{pd} \leq 1$ holds for any ρ . In such a situation, ironically, inclusion of baseline covariates can only do harm to the efficiency owing to the loss of certain degrees of freedom when estimating the error variance in the data analysis. The second example is a totally balanced design with $p = 2$ and $t = 9$, which is slightly less efficient because the treatment effects are more or less (depending on ρ) based on between-subject information. Numerical computation shows that the upper bound for RE^{pd} ranges between 1.200 and 1.494 when ρ varies from 0.5 to 0.99.

We have considered two different baseline methods, although the period-dependent baselines seem to be more often used in practice. However, if potential confounding between ‘late-period’ (those other than the first period) baselines and other model effects is suspected, which seems to be a legitimate concern for crossover designs, then the study baseline method is more relevant and thus preferred. In addition, we have also assumed different regression coefficients for baseline covariates fitted for different periods (i.e., baseline-by-period interaction effects). If a common regression coefficient is considered instead, then all the upper bounds obtained in Theorem 1 are still valid (see the proof in the Appendix).

We have adopted a simple compound (complete) symmetry covariance structure for all the measurements (both baselines and responses) made on the same subject. This assumption not only makes our problem mathematically more tractable but is also reasonable and quite sufficient in many applications. For example, the measurement errors (ε_{ks} ’s) made on the same subject could be assumed to be autocorrelated (e.g., AR(1)), but in practice, the addition to the overall covariance due to such an error correlation is often negligible as compared with the covariance due to the within-subject correlation. For a totally balanced design with $p > 2$, we believe that the baseline covariates will have more impact on the efficiency if the within-subject correlation between the baseline and the response is much higher than that between the responses themselves (i.e., the baselines are more informative than the responses). But we also feel that such an assumption is unlikely to be supported in the data analysis.

Finally, it is interesting to note that our conclusions are consistent with those reported by Liang and Carrière [8]. Although their findings were based on a relatively small class of crossover designs ($t = 2$ and $p = 2, 3$, and 4) with different assumptions on the carryover effects and a different baseline approach (baselines were modeled as additional responses).

Appendix A

Proof of Theorem 1

The proofs for RE^s and RE^{pd} are similar, so here we only provide necessary details for RE^{pd} .

Let $\mathbf{1}_a$, $\mathbf{I}_a$, and $\mathbf{J}_a$ denote the column vector with all a entries equal to 1, the $a \times a$ identity matrix, and the $a \times a$ matrix with all entries equal to 1, respectively. For any $a \times b$ matrix $\mathbf{X}$, we define $\text{pr}^\perp(\mathbf{X}) = \mathbf{I}_a - \mathbf{X}(\mathbf{X}^T \mathbf{X})^- \mathbf{X}^T$, where $(\mathbf{X}^T \mathbf{X})^-$ is a generalized inverse of $\mathbf{X}^T \mathbf{X}$.

In matrix form, the mean response under models (1) and (3), respectively, can be written as

$$E(\mathbf{Y}) = \mathbf{1}_{pn} \mu + \mathbf{P} \alpha + \mathbf{T}_d \tau + \mathbf{F}_d \gamma \quad (8)$$

and

$$E(\mathbf{Y}|\mathbf{Z}) = \mathbf{1}_{pn} \mu + \mathbf{P} \alpha + \mathbf{T}_d \tau + \mathbf{F}_d \gamma + \sum_{k=1}^p \mathbf{Z}_k \beta_k, \quad (9)$$

where $\mathbf{Y}$ is the column vector of the responses that are ordered period by period for each subject in turn; α , τ , and γ are the column vectors of the period, treatment, and carryover effects with $\mathbf{P}$, $\mathbf{T}_d$, and $\mathbf{F}_d$ being the corresponding design matrices, respectively; $\mathbf{Z}$ is the column vector of the period-dependent baseline covariates; and $\mathbf{Z}_k$ ($k = 1, \dots, p$) is the vector obtained from $\mathbf{Z}$ by keeping the entries $z_{k1}, z_{k2}, \dots, z_{kn}$ (i.e., the baseline covariates for the k th period) while replacing all other ones by 0.

Our assumptions in Sections 2.1 and 2.2 imply that $\text{cov}(\mathbf{Y}) = \sigma_\varepsilon^2 \mathbf{V}_0$ and $\text{cov}([\mathbf{Y}^T, \mathbf{Z}^T]^T) = \sigma_\varepsilon^2 [\mathbf{I}_2 \otimes \mathbf{V}_0 + (\mathbf{J}_2 - \mathbf{I}_2) \otimes \mathbf{V}_{YZ}]$, where $\mathbf{V}_0 = \mathbf{I}_n \otimes (\mathbf{I}_p + \theta \mathbf{J}_p)$ with $\theta = \sigma_s^2 / \sigma_\varepsilon^2$ and $\otimes$ being the Kronecker product and $\sigma_\varepsilon^2 \mathbf{V}_{YZ} = \sigma_s^2 (\mathbf{I}_n \otimes \mathbf{J}_p)$ is the covariance matrix between $\mathbf{Y}$ and $\mathbf{Z}$. The covariance matrix of $\mathbf{Y}$ conditioning on $\mathbf{Z}$ is then given by $\text{cov}(\mathbf{Y}|\mathbf{Z}) = \sigma_\varepsilon^2 \mathbf{V}_1$, where $\mathbf{V}_1 = \mathbf{V}_0 - \mathbf{V}_{YZ} \mathbf{V}_0^{-1} \mathbf{V}_{YZ}^T = \mathbf{I}_n \otimes [\mathbf{I}_p + \theta(1 + p\theta)^{-1} \mathbf{J}_p]$.

The Fisher information matrices, denoted by $\mathbf{C}_\tau(\theta)$ and $\mathbf{C}_\tau(\theta|\mathbf{Z})$, for the direct treatment effects τ under models (8) and (9), respectively, can be expressed as

$$\mathbf{C}_\tau(\theta) = \mathbf{T}_d^T \mathbf{V}_0^{-\frac{1}{2}} \text{pr}^\perp \left(\mathbf{V}_0^{-\frac{1}{2}} [\mathbf{1}_{pn}, \mathbf{P}, \mathbf{F}_d] \right) \mathbf{V}_0^{-\frac{1}{2}} \mathbf{T}_d$$

and

$$\mathbf{C}_\tau(\theta|\mathbf{Z}) = \mathbf{T}_d^T \mathbf{V}_1^{-\frac{1}{2}} \text{pr}^\perp \left(\mathbf{V}_1^{-\frac{1}{2}} [\mathbf{1}_{pn}, \mathbf{P}, \mathbf{F}_d, \mathbf{Z}_1, \dots, \mathbf{Z}_p] \right) \mathbf{V}_1^{-\frac{1}{2}} \mathbf{T}_d, \quad (10)$$

with $\mathbf{V}_0^{-\frac{1}{2}} \mathbf{V}_0^{-\frac{1}{2}} = \mathbf{V}_0^{-1}$ and $\mathbf{V}_1^{-\frac{1}{2}} \mathbf{V}_1^{-\frac{1}{2}} = \mathbf{V}_1^{-1}$.

An upper bound for $\mathbf{C}_\tau(\theta|\mathbf{Z})$, in the Loewner sense (i.e., $\mathbf{A} \leq \mathbf{B}$ if and only if $\mathbf{A} - \mathbf{B}$ is non-negative definite), can be obtained by ignoring all $\mathbf{Z}_1, \dots, \mathbf{Z}_p$ on the right-hand side of (10). That is,

$$\mathbf{C}_\tau(\theta|\mathbf{Z}) \leq \mathbf{T}_d^T \mathbf{V}_1^{-\frac{1}{2}} \text{pr}^\perp \left(\mathbf{V}_1^{-\frac{1}{2}} [\mathbf{1}_{pn}, \mathbf{P}, \mathbf{F}_d] \right) \mathbf{V}_1^{-\frac{1}{2}} \mathbf{T}_d = \mathbf{C}_\tau(\theta(1 + p\theta)^{-1}),$$

with equality if the columns of $\mathbf{Z}$ belong to the column span of the matrix $[\mathbf{1}_{pn}, \mathbf{P}, \mathbf{F}_d]$ (note that $\mathbf{Z} = \sum_{k=1}^p \mathbf{Z}_k$).

In light of Hedayat *et al.* [10], for any totally balanced design with $p \leq t$, $\mathbf{C}_\tau(\theta)$ is completely symmetric (compound symmetric) with $\text{tr}[\mathbf{C}_\tau(\theta)] = n \cdot f_{(t,p)}(\theta)$, where $f_{(t,p)}(\theta)$ is as defined in Theorem 1. On the basis of Kiefer's universal optimality criteria, we would have

$$RE^{pd} \leq \frac{\text{tr}[\mathbf{C}_\tau(\theta(1 + p\theta)^{-1})]}{\text{tr}[\mathbf{C}_\tau(\theta)]} = \frac{f_{(t,p)}(\theta(1 + p\theta)^{-1})}{f_{(t,p)}(\theta)} = h_{(t,p)}(\theta). \quad (11)$$

For the special case $p = t$, it can be shown that $h_{(t,t)}(\theta)$ is non-decreasing in θ for any t . Indeed, with the help of the mathematical software MAPLE, we can obtain

$$\frac{\partial h_{(t,t)}(\theta)}{\partial \theta} = \frac{\theta(t-1)^2 [3t^4 - 6t^3 - 2t^2 + 5t + 1] + 2t^3 - 4t^2 + 2}{t[\theta(2t^2 - 2t - 1) + t - 1]^2 [\theta(t^2 - t - 1) + t - 1]^2 f_{(t,t)}^2(\theta)},$$

which is non-negative for any t and θ . Now we take $\theta = \infty$ in (11); then simple algebra yields

$$RE^{pd} \leq 1 + \frac{1}{(t^2 - t - 2)(2t^2 - 2t - 1)}.$$

□

Acknowledgements

The author thanks the editor, an associate editor, and anonymous referees for their valuable comments and suggestions that have greatly enhanced the manuscript. The author is indebted to Dr. Vince Shu at Allergan for many helpful discussions related to this paper.

References

1. Hills M, Armitage P. The two-period crossover clinical trial. *British Journal of Clinical Pharmacology* 1979; **8**:7–20.
2. Wallenstein S. Inclusion of baseline values in the analysis of crossover designs (abstract). *Biometrics* 1979; **35**:894.
3. Fleiss JL, Wallenstein S, Rosenfeld R. Adjusting for baseline measurements in the two-period crossover study: a cautionary note. *Controlled Clinical Trials* 1985; **6**:192–197.
4. Willan AR, Pater JL. Using baseline measurements in the two period cross-over trial. *Controlled Clinical Trials* 1986; **7**:282–289.
5. Kenward MG, Jones B. The analysis of data from 2×2 cross-over trials with baseline measurements. *Statistics in Medicine* 1987; **6**:911–926.
6. Senn S. *Cross-over Trials in Clinical Research (Statistics in Practice)*, 2nd edn. Wiley: Chichester, 2002.
7. Johns B, Kenward MG. *Design and Analysis of Cross-over Trials*, 2nd edn. Chapman & Hall/CRC Press: Boca Raton, 2003.
8. Liang Y, Carrière KC. On the role of baseline measurements for crossover designs under the self and mixed carryover effects model. *Biometrics* 2010; **66**(1):140–148.
9. Kenward MG, Roger JH. The use of baseline covariates in crossover studies. *Biostatistics* 2010; **11**(1):1–17.
10. Hedayat AS, Stufken J, Yang M. Optimal and efficient crossover designs when subject effects are random. *Journal of the American Statistical Association* 2006; **101**:1031–1038.
11. Kunert J, Stufken J. Optimal crossover designs in a model with self and mixed carryover effects. *Journal of the American Statistical Association* 2002; **97**:898–906.
12. Cheng CS, Wu CFJ. Balanced repeated measurement designs. *The Annals of Statistics* 1980; **8**:1272–1283.
13. Kiefer J. Construction and optimality of generalized Youden designs. *A Survey of Statistical Design and Linear Models*, North-Holland, Amsterdam, 1975; 333–353.