



Optimal Design in Hierarchical Random Effect Models for Individual Prediction with Application in Precision Medicine

Maryna Prus¹ · Norbert Benda^{2,3} · Rainer Schwabe¹

© The Author(s) 2021, corrected publication 2021

Abstract

Hierarchical random effect models are used for different purposes in clinical research and other areas. In general, the main focus is on population parameters related to the expected treatment effects or group differences among all units of an upper level (e.g. subjects in many settings). Optimal design for estimation of population parameters are well established for many models. However, optimal designs for the prediction for the individual units may be different. Several settings are identified in which individual prediction may be of interest. In this paper, we determine optimal designs for the individual predictions, e.g. in multi-cluster trials or in trials that investigate a new treatment in a number of different subpopulations, and compare them to a conventional balanced design with respect to treatment allocation. Our investigations show that in the case of uncorrelated cluster intercepts and cluster treatments the optimal allocations are far from being balanced if the treatment effects vary strongly as compared to the residual error and more subjects should be recruited to the active (new) treatment. Nevertheless, efficiency loss may be limited resulting in a moderate sample size increase when individual predictions are foreseen with a balanced allocation.

Keywords Optimal design · Random coefficient regression · Multi-cluster trials · Prediction

This research has been partially supported by Grant SCHW 531/16 of the German Research Foundation (DFG).

✉ Maryna Prus
maryna.prus@ovgu.de

¹ Institute for Mathematical Stochastics, Otto von Guericke University, Magdeburg, Germany

² Research Division, Biostatistics and Special Pharmacokinetics Unit, Federal Institute for Drugs and Medical Devices, Bonn, Germany

³ Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

1 Introduction

Hierarchical random effect models are used for different purposes. Common applications, e.g. in clinical research, are random effect meta analyses of several clinical trials using individual patient data (IPD), multi-centre trials assuming random centre effects or mixed or random effect models for repeated measurements (MMRM) for longitudinal data within subjects. Clinical trials or meta-analyses of clinical trials may involve several populations or subgroups of patients (e.g. defined by genetic biomarkers), where in some settings between-population variability may be modelled using a hierarchical approach.

Precision medicine aims at investigating new treatments in different subpopulations of patients, e.g. with specific tumour subtypes, in order to detect potentially differential treatment effects indicating the application in specific subpopulations. Basket trials investigate a new treatment in a number of different groups of patients, where the treatment is compared to a control in each of the different subpopulations. If, for example, two binary or dichotomized genetic biomarkers identify biomarker positive or negative patients, the whole patient population could be split into four disjoint subsets of patients. However, far more subpopulations of interest may be identified, especially in oncological settings, where one treatment may be investigated in a large number of different patient populations. Furthermore, in many oncological indications, biomarker defined subgroups might be rather small and early phase clinical trials aim at exploring different subpopulations that may be highly heterogeneous with respect to the response to treatment, since depending on the individual biomarker status, different treatment effects are often expected. In these difficult settings with a limited number of trial participants available, high design efficiency is an important prerequisite for a successful development of new targeted drugs in small populations.

Further applications refer to observational studies data on quality parameters of drugs arise from hierarchical settings with multiple layers, i.e. with multiple sources of variability according to the underlying manufacturing process, e.g. given by manufacturing sites, production batches and samples within batches. In general, the main focus is on population parameters related to the expected treatment effects or group differences among all units of an upper level (e.g. trials in IPD meta-analyses, centres in multi-centre trials, patients in longitudinal trials, batches in quality control, etc.). Several authors considered optimal design for estimation of population parameters (expected values of random effects) in similar models; see e.g. [2–6, 8]. In general, prediction of the outcome in the individual units may also be of interest in several settings, as for treatment effects in single clusters to assess qualification of individual clinics, manufacturing sites in manufacturing control, treatment effects in different subpopulations of patients. In these cases, the question arises, whether optimal designs for population parameters can also be used for individual predictions of random effects, whether optimal designs for individual predictions differ from those for population parameters and to which extent and which efficiency loss (or sample size increase) can be anticipated if another conventional design is chosen.

In this paper, we investigate optimal designs for random effects to be applied, e.g. in basket trials (or multi-cluster trials), investigating a number of disjoint patient populations and compare them to a conventional balanced design with respect to treatment allocation. Nevertheless, the results are applicable to very different settings as outlined above.

The structure of the paper is as follows: In Sect. 2, a hierarchical model is specified and the best linear unbiased predictions of the individual parameters (intercepts and treatment effects for the individual units or clusters) are derived. In Sect. 3, analytical results are presented to characterize optimal designs for prediction of the cluster specific effects. The results are illustrated by some numerical examples. The paper is concluded by a short discussion.

2 Model Specification

We consider a randomized comparative trial with K different clusters (e.g. subpopulations). In all of these clusters, individuals are allocated to two treatment groups. In the first group (denoted by $x = 1$), the individuals receive an active treatment while in the second group (denoted by $x = 0$) a placebo or a control treatment is applied.

We consider K clusters $i = 1, \dots, K$. Denote by μ_i and α_i the intercept (mean response at placebo or control) and the effect of the active treatment (compared to placebo or control), respectively, in cluster i , which both may vary across the clusters. The response Y_{ij} of an individual $j = 1, \dots, N_i$ in cluster i can be described as

$$Y_{ij} = \mu_i + \alpha_i x_{ij} + \varepsilon_{ij}, \quad (1)$$

where x_{ij} is equal to 1, if the individual belongs to the treatment group, and x_{ij} is equal to 0 for the control (or placebo) group and ε_{ij} denotes the random variation in the response of the individuals. The individual variations ε_{ij} are assumed to have zero mean and to be homoscedastic with common variance σ^2 .

The cluster specific intercepts μ_i and treatment effects α_i can be assumed as random with (unknown) expected values $E(\mu_i) = \mu$ and $E(\alpha_i) = \alpha$ characterizing the mean intercept and mean treatment effect across the clusters and covariance structure $\text{Cov}((\mu_i, \alpha_i)^\top) = \sigma^2 \mathbf{D}$ for some 2×2 positive definite dispersion matrix \mathbf{D} . All random effects and all individual variations are assumed to be uncorrelated.

For the sake of simplicity, we further assume that the total number N of individuals per cluster is the same for all clusters ($N_i = N$) and that the allocation rate is constant across the clusters, i. e. the number n of individuals in the treatment group is the same for all clusters. The design problem can then be formulated in terms of finding the optimal allocation rate $w = n/N$ for the treatment group.

Because of exchangeability of the individuals within each cluster, we may sort them in the analysis regardless of randomization in such a way that the first n individuals $j = 1, \dots, n$ to be analyzed receive the active treatment and the remaining $N - n$ individuals $j = n + 1, \dots, N$ are in the control group. Then the experimental settings x_{ij} in (1) can be specified by

$$x_{ij} = x_j = \begin{cases} 1, & j = 1, \dots, n \\ 0, & j = n + 1, \dots, N \end{cases}$$

and are independent of the cluster i .

Hence, the multi-cluster model (1) can be identified as a particular case of the random coefficient regression model

$$Y_{ij} = \mathbf{f}(x_j)^\top \boldsymbol{\beta}_i + \varepsilon_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, N \tag{2}$$

investigated by [7] when the regression functions and the cluster parameters are specified by $\mathbf{f}(x) = (1, x)^\top$ and $\boldsymbol{\beta}_i = (\mu_i, \alpha_i)^\top$, respectively. In general, this model can be written in vector notation as

$$\mathbf{Y}_i = \mathbf{F}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \tag{3}$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iN})^\top$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iN})^\top$ are the N -dimensional vectors of observations and individual variations at cluster i , respectively, and $\mathbf{F} = (\mathbf{f}(x_1), \dots, \mathbf{f}(x_N))^\top$ is the within cluster design matrix which is equal across all clusters.

For the present multi-cluster model, the design matrix \mathbf{F} simplifies to

$$\mathbf{F} = \begin{pmatrix} \mathbb{1}_n & \mathbb{1}_n \\ \mathbb{1}_{N-n} & \mathbf{0}_{N-n} \end{pmatrix},$$

where $\mathbb{1}_\ell$ and $\mathbf{0}_\ell$ denote the ℓ -dimensional vectors with all entries equal to 1 and 0, respectively.

According to [7], Corollary 1 (see also [2]) in model (2), the best linear unbiased predictors (BLUPs)

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1})^{-1} (\mathbf{F}^\top \mathbf{F} \hat{\boldsymbol{\beta}}_{i;\text{within}} + \mathbf{D}^{-1} \hat{\boldsymbol{\beta}}_0) \tag{4}$$

of the random parameters $\boldsymbol{\beta}_i$ are weighted combinations of the estimates $\hat{\boldsymbol{\beta}}_{i;\text{within}} = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \mathbf{Y}_i$ based only on the observations within cluster i and the best linear unbiased estimator (BLUE) $\hat{\boldsymbol{\beta}}_0 = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \bar{\mathbf{Y}}$ of the population parameter $\boldsymbol{\beta}_0 = E(\boldsymbol{\beta}_i) = (\mu, \alpha)^\top$, where $\bar{\mathbf{Y}} = \frac{1}{K} \sum_{i=1}^K \mathbf{Y}_i$ is the mean observational vector averaged across the clusters.

We additionally assume that the cluster intercepts μ_i and the cluster treatment effects α_i are uncorrelated for all clusters, i. e. $\mathbf{D} = \text{diag}(u, v)$, where $u = \sigma_\mu^2 / \sigma^2 > 0$ and $v = \sigma_\alpha^2 / \sigma^2 > 0$ are the variance ratios of the intercepts and the treatment effects in relation to the observational variance of the individuals.

With the common notations $\bar{Y}_{i\cdot}^{(T)} = \frac{1}{n} \sum_{j=1}^n Y_{ij}$ and $\bar{Y}_{i\cdot}^{(C)} = \frac{1}{N-n} \sum_{j=n+1}^N Y_{ij}$ for the mean response in the treatment (“ T ”) and the control (“ C ”) group in cluster i , $\bar{Y}_{\cdot\cdot}^{(T)} = \frac{1}{K} \sum_{i=1}^K \bar{Y}_{i\cdot}^{(T)}$ and $\bar{Y}_{\cdot\cdot}^{(C)} = \frac{1}{K} \sum_{i=1}^K \bar{Y}_{i\cdot}^{(C)}$ for the overall mean of the treatment and the control group, respectively, the BLUPs for the cluster parameters $\hat{\boldsymbol{\beta}}_i = (\hat{\mu}_i, \hat{\alpha}_i)^\top$ of the random intercepts and the random treatment effects $\boldsymbol{\beta}_i = (\mu_i, \alpha_i)^\top$ in model (1) can be written as weighted averages

$$\hat{\mu}_i = c_0 \left(\bar{Y}_{i\cdot}^{(T)} - \bar{Y}_{\cdot\cdot}^{(T)} \right) + c \bar{Y}_{i\cdot}^{(C)} + (1 - c) \bar{Y}_{\cdot\cdot}^{(C)}, \tag{5}$$

where $c_0 = \frac{nu}{(Nu+1)(nv+1)-n^2uv}$ and $c = \frac{u(N-n)(nv+1)}{(Nu+1)(nv+1)-n^2uv}$, and

$$\hat{\alpha}_i = c_1 \bar{Y}_{i\cdot}^{(T)} + (1 - c_1) \bar{Y}_{\cdot\cdot}^{(T)} - c_2 \bar{Y}_{i\cdot}^{(C)} - (1 - c_2) \bar{Y}_{\cdot\cdot}^{(C)} \tag{6}$$

with weights $c_1 = \frac{(Nu+1)nv-n^2uv}{(Nu+1)(nv+1)-n^2uv}$ and $c_2 = \frac{Nu+nv+1}{(Nu+1)(nv+1)-n^2uv}$.

The derivation of formulas (5) and (6) is deferred to “Appendix”.

To measure the quality of a design, we will use the mean squared error (MSE) matrix of the BLUP $\hat{\beta} = \left(\hat{\beta}_1^\top, \dots, \hat{\beta}_K^\top \right)^\top$ of the complete vector $\beta = \left(\beta_1^\top, \dots, \beta_K^\top \right)^\top$ of all random parameters. The mean squared error (MSE) matrix for the BLUP $\hat{\beta}$ can be computed by means of the following formula :

$$\begin{aligned} \text{Cov}(\hat{\beta} - \beta) &= \sigma^2 \left(\frac{1}{K} \left(\mathbb{1}_K \mathbb{1}_K^\top \right) \otimes \left(\mathbf{F}^\top \mathbf{F} \right)^{-1} \right. \\ &\quad \left. + \left(\mathbb{1}_K - \frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top \right) \otimes \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1} \right)^{-1} \right) \end{aligned} \tag{7}$$

(see [7], Corollary 2), where $\mathbb{1}_\ell$ denotes the $\ell \times \ell$ identity matrix and \otimes is the symbol for the Kronecker product of matrices or vectors.

Further denote by $\alpha = \left(\alpha_1, \dots, \alpha_K \right)^\top$ the vector of treatment effects for all clusters. Then $\alpha = \left(\mathbb{1}_K \otimes (0, 1) \right) \beta$ and, hence,

$$\text{Cov}(\hat{\alpha} - \alpha) = \left(\mathbb{1}_K \otimes (0, 1) \right) \text{Cov}(\hat{\beta} - \beta) \left(\mathbb{1}_K \otimes (0, 1) \right)^\top \tag{8}$$

for the MSE matrix of the BLUP $\hat{\alpha} = \left(\hat{\alpha}_1, \dots, \hat{\alpha}_K \right)^\top$ of α . Using this and formula (7), we obtain the MSE matrix

$$\begin{aligned} \text{Cov}(\hat{\alpha} - \alpha) &= \sigma^2 \left(\frac{N}{n(N-n)} \frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top \right. \\ &\quad \left. + \frac{v(Nu+1)}{(Nu+1)(nv+1)-n^2uv} \left(\mathbb{1}_K - \frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top \right) \right). \end{aligned} \tag{9}$$

Note that the MSE matrix for $\hat{\alpha}$ is completely symmetric and that, hence, the MSE of $\hat{\alpha}_i$ attains the same value for all cluster treatment effects.

3 Optimal Design

As individuals are interchangeable within treatment groups, we may define an exact within cluster design

$$\begin{pmatrix} T & C \\ n & N - n \end{pmatrix} \tag{10}$$

by the allocation numbers n and $N - n$ to the treatment and control groups T and C , respectively.

For analytical purposes, we generalize this to the definition of an approximate design:

$$\xi = \begin{pmatrix} T & C \\ w & 1 - w \end{pmatrix}, \tag{11}$$

where $w = \frac{n}{N}$ is the allocation rate to the treatment group and $1 - w = \frac{N-n}{N}$ is the allocation rate to the control group. For finding an optimal design, only the optimal allocation rate w^* to the treatment group has to be determined.

For an approximate design, the definition of the MSE matrix (9) of the BLUP $\hat{\alpha}$ is extended in a straightforward manner and can be rewritten (neglecting σ^2) as :

$$\begin{aligned} \text{MSE}(w) &= \frac{1}{Nw(1-w)} \frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top \\ &+ \frac{v(Nu+1)}{(Nu+1)(Nwv+1) - N^2w^2uv} \left(\mathbb{1}_K - \frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top \right) \end{aligned} \tag{12}$$

in terms of the allocation rate w .

For the assessment of the MSE matrix, we focus on the A -optimality criterion for the cluster treatment effects α , which averages the mean squared errors of the predicted cluster treatment effects $\hat{\alpha}_i$. More specifically, the A -criterion $\Phi_{A,\alpha}$ is the trace of the MSE matrix of the prediction $\hat{\alpha}$ of the cluster treatment effects. For an approximate design, we obtain

$$\Phi_{A,\alpha}(w) = \frac{1}{Nw(1-w)} + (K-1) \frac{v(Nu+1)}{(Nu+1)(Nwv+1) - N^2w^2uv} \tag{13}$$

for the criterion function $\Phi_{A,\alpha}$ in terms of the allocation rate w .

Alternatively, one may consider the MV-optimality criterion for the cluster treatment effects α , which measures the maximal MSE over all predicted cluster treatment effects $\hat{\alpha}_i$. More specifically, the MV-criterion $\Phi_{MV,\alpha}$ is the maximal diagonal entry of the MSE matrix of the prediction $\hat{\alpha}$ of the cluster treatment effects. Because the MSE is equal for all $\hat{\alpha}_i$, the MV-criterion is related to the A -criterion by $\Phi_{MV,\alpha} = \frac{1}{K} \Phi_{A,\alpha}$, and hence, the A -optimal designs are also MV-optimal.

Because, in general, there is no explicit solution for the optimal allocation rate, which minimizes (13), we will give an insight in the qualitative behaviour by some numerical examples below.

It is worthwhile mentioning that the criterion (13) is convex, and therefore, an optimal exact design may be obtained by choosing the best of the two exact designs adjacent to an optimal approximate design.

Example 1 For illustrative purposes, we consider a numerical example with $K = 16$ clusters and $N = 4$ individuals in each clusters, corresponding, for example, to a phase II trial investigating the potentially heterogeneous treatment effect in a population that is subdivided by four different dichotomous biomarkers. Obviously, the

resulting sample size per subgroup will be small resulting in a high need for optimal designs.

Figure 1 exhibits the behaviour of the optimal allocation rate w^* to the treatment group in dependence on the variance ratio v of the treatment effects for some fixed values 0.01, 0.1, 0.25, 0.5 and 1.5 of the variance ratio u of the intercept. For reasons of presentation, we plot the optimal allocation rate w^* against the rescaled variance ratio $r_v = v/(1 + v)$ in the spirit of intra-class correlation in order to cover all possible values of the treatment effects variance ratio by a finite interval $((0, 1))$, where $r_v \rightarrow 0$ and $r_v \rightarrow 1$ relate to $v \rightarrow 0$ and $v \rightarrow \infty$, respectively, and r_v is monotonically increasing in v . What can be seen from the picture is that for fixed values of u the optimal allocation rate w^* is equal to 0.5 for $v \rightarrow 0$ and increases with increasing values of the variance ratio v of the treatment effects. The different lines associated with the different values of u appear in descending order which means that the optimal allocation rates decrease when the variance ratio u of the intercepts gets larger.

The next figure (Fig. 2) shows the behaviour of the optimal allocation rate in dependence on the variance ratio u of the intercepts for fixed values 0.01, 0.1, 0.2, 0.5 and 2 of the variance ratio v of the treatment effects, where again the variance ratio is rescaled ($r_u = u/(u + 1)$). Also here it can be seen that the optimal allocation rate decreases with increasing values of u and increases with increasing values of v .

Figures 3 and 4 present the efficiency of the equal allocation rate $w_0 = 0.5$ which is optimal in the fixed effects model ($u = v = 0$). The efficiency for the A-criterion (A-efficiency) has been computed using the common formula

$$\text{eff}_A(w_0) = \frac{\Phi_{A,\alpha}(w^*)}{\Phi_{A,\alpha}(w_0)}. \quad (14)$$

Note that for the MV-criterion the MV-efficiency, which is defined analogously, coincides with the A-efficiency. The efficiencies decrease with increasing values of

Fig. 1 A-optimal allocation rate w^* to the treatment group in dependence on the rescaled variance ratio $r_v = v/(1 + v)$ of the treatment effects for $u = 0.01$ (solid line), 0.1 (dashed line), 0.25 (dotted line), 0.5 (dashed-dotted line) and 1.5 (dashed line, long dashes)

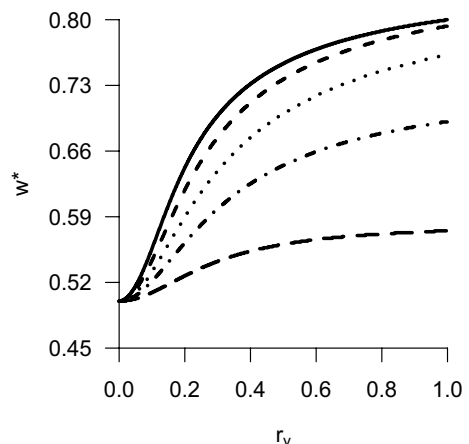


Fig. 2 A-optimal allocation rate w^* to the treatment group in dependence on the rescaled variance ratio $r_u = u/(1 + u)$ of the intercepts for $v = 0.01$, (solid line), 0.1 (dashed line), 0.25 (dotted line), 0.5 (dashed-dotted line) and 2 (dashed line, long dashes)

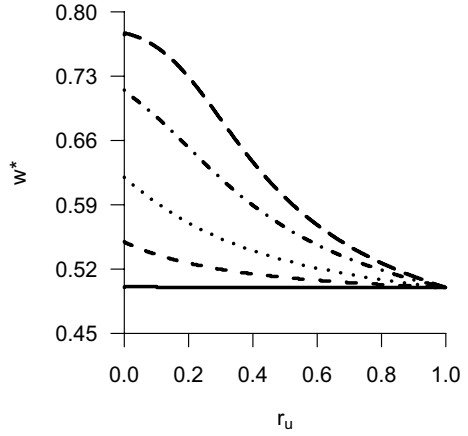


Fig. 3 A-efficiency of the equal allocation rate $w_0 = 0.5$ in dependence on the rescaled variance ratio $r_v = v/(1 + v)$ of the treatment effects for $u = 0.01$ (solid line), 0.1 (dashed line), 0.25 (dotted line), 0.5 (dashed-dotted line) and 1.5 (dashed line, long dashes)

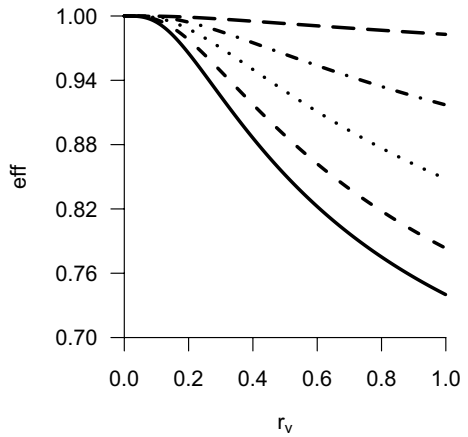
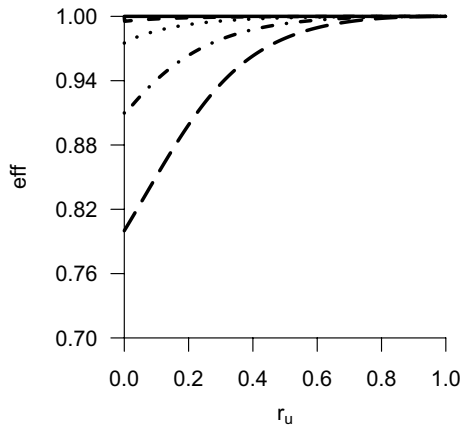


Fig. 4 A-efficiency of the equal allocation rate $w_0 = 0.5$ in dependence on the rescaled variance ratio $r_u = u/(1 + u)$ of the intercepts for $v = 0.01$ (solid line), 0.1 (dashed line), 0.25 (dotted line), 0.5 (dashed-dotted line) and 2 (dashed line, long dashes)



the variance of the treatment effects and increase with increasing values of the variance of the intercepts.

Example 2 Extending the illustrative example, we further consider trials with $K = 16$ or $K = 32$ clusters and $N = 6$ or $N = 4$ individuals in each cluster. Figure 5 exhibits the behaviour of the optimal allocation rate w^* to the treatment group in dependence on the variance ratio $q = v/u$ for some fixed values 0.01, 0.1, 0.25, 0.5 and 1.5 of the variance ratio u of the intercept. For reasons of presentation, we again plot the optimal allocation rate w^* against the rescaled variance ratio $r_q = q/(1 + q)$. As we can observe from the graphics, the dependence on the variance ratio q is more essential for $K = 32$ and $N = 4$ than for $K = 16$ and $N = 6$.

Figure 6 presents the efficiency of the equal allocation rate $w_0 = 0.5$ which is optimal in the fixed effects model. The efficiency is more sensible with respect to the variance ratio in the case $K = 32$ and $N = 4$ than for $K = 16$ and $N = 6$.

4 Discussion

In the present paper, we focused on A - and MV -optimality for prediction of the cluster treatment effects α . Alternatively, one may be tempted to employ the D -optimality criterion $\Phi_{D,\alpha}$ which is defined as the log determinant of the MSE matrix of the prediction $\hat{\alpha}$ of the cluster treatment effects α ,

$$\Phi_{D,\alpha}(w) = \log\left(\frac{1}{Nw(1-w)}\right) + (K-1)\log\left(\frac{v(Nu+1)}{(Nu+1)(Nwv+1) - N^2w^2uv}\right).$$

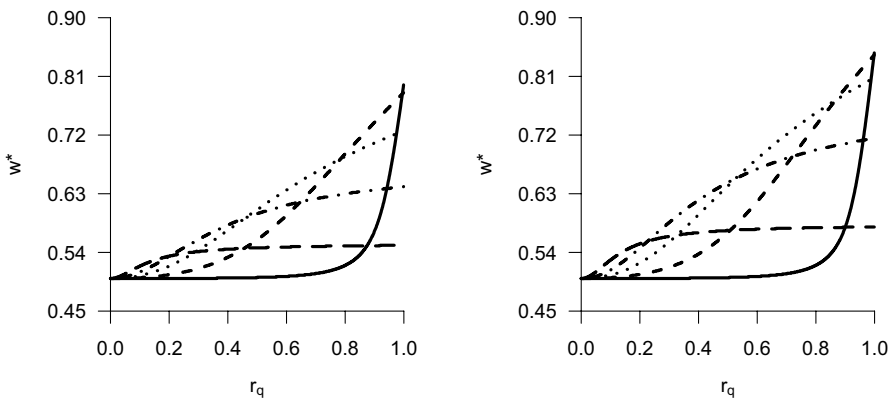


Fig. 5 A -optimal allocation rate w^* to the treatment group in dependence on the rescaled variance ratio $r_q = q/(1 + q)$ of the treatment effects for $u = 0.01$ (solid line), 0.1 (dashed line), 0.25 (dotted line), 0.5 (dashed-dotted line), 1.5 (dashed line, long dashes) and for $K = 16$ and $N = 6$ (left panel), $K = 32$ and $N = 4$ (right panel)

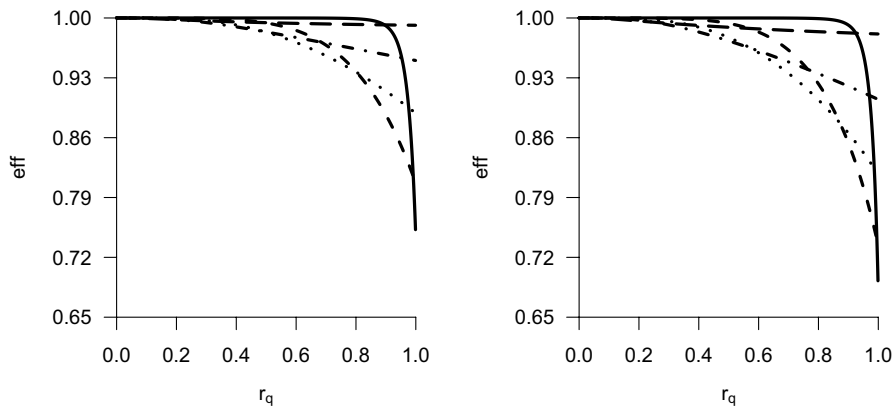


Fig. 6 A-efficiency of the equal allocation rate $w_0 = 0.5$ in dependence on the rescaled variance ratio $r_q = q/(1 + q)$ of the intercepts for $u = 0.01$ (solid line), 0.1 (dashed line), 0.25 (dotted line), 0.5 (dashed-dotted line) and 1.5 (dashed line, long dashes) and for $K = 16$ and $N = 6$ (left panel), $K = 32$ and $N = 4$ (right panel)

The D -criterion is commonly used in the case of estimation of fixed effects and measures essentially the volume of a simultaneous confidence ellipsoid for the parameters of interest taking correlations between the parameter estimates into account. In the present setting of prediction of cluster treatment effects, especially when considering treatment effects in patient subpopulations, stand-alone interpretations for the given clusters are targeted and the simultaneous prediction ellipsoid corresponding to the D -criterion appears less relevant for the clinical interpretation. From a practical point of view, simultaneous prediction intervals based on a Bonferroni-like approach would better fit to the intended application. The corresponding R -optimality criterion (see e.g. [1]) measures the volume of these multidimensional intervals and is defined as the product of the diagonal elements of the MSE matrix. In the present situation of equal diagonal elements, the R -criterion $\Phi_{R,\alpha}$ is related to the A - and MV -criterion by $\Phi_{R,\alpha} = (\Phi_{MV,\alpha})^K = (\frac{1}{K}\Phi_{A,\alpha})^K$, and hence, the A -optimal designs are also R -optimal.

As illustrated in the examples, the larger the between-unit (between-cluster) variability of the treatment effects, the more the optimal weight deviates from equal allocation, especially if the variance of the treatment effects is large compared to the variance of the intercepts of the units. An increasing heterogeneity in the treatment effect leads to a decreased precision of the design that is optimal for overall population parameters: A balanced design is far from optimal if the treatment effects vary strongly as compared to the residual error and more subjects should be recruited to the active (new) treatment in multi-cluster trials. Nevertheless, efficiency loss may still be limited as in the examples resulting in a total sample size increase of about 10–20% in the considered scenarios if individual predictions are foreseen with a less efficient balanced allocation. If between-unit variability of treatment effects is considered to be small, equal allocation may suffice. However, using the results given in the paper, specific settings with different expectations can be assessed properly, in

order to make optimally use of a limited number of patients or sample units to predict random effects of individual units. Basket trials with few patients in a number of biomarker defined subpopulations could be designed as efficient as possible in order to better facilitate the development of new targeted drugs.

Appendix: Derivations of the Formulas for the BLUPs for the Cluster Parameters

We use formula (4) for computing the BLUPs. First note

$$\begin{aligned}
 (\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1})^{-1} &= \frac{1}{(Nu + 1)(nv + 1) - n^2uv} \begin{pmatrix} u(nv + 1) & -nuv \\ -nuv & v(Nu + 1) \end{pmatrix}, \\
 \mathbf{F}^\top \mathbf{F} \hat{\boldsymbol{\beta}}_{i;\text{within}} &= \mathbf{F}^\top \mathbf{Y}_i = \begin{pmatrix} n\bar{Y}_{i\cdot}^{(T)} + (N - n)\bar{Y}_{i\cdot}^{(C)} \\ n\bar{Y}_{i\cdot}^{(T)} \end{pmatrix}, \\
 (\mathbf{F}^\top \mathbf{F})^{-1} &= \frac{1}{N - n} \begin{pmatrix} 1 & -1 \\ -1 & \frac{N}{n} \end{pmatrix}, \\
 \mathbf{F}^\top \bar{\mathbf{Y}} &= \begin{pmatrix} n\bar{Y}_{\cdot\cdot}^{(T)} + (N - n)\bar{Y}_{\cdot\cdot}^{(C)} \\ n\bar{Y}_{\cdot\cdot}^{(T)} \end{pmatrix}, \\
 \hat{\boldsymbol{\beta}}_0 &= \begin{pmatrix} \bar{Y}_{\cdot\cdot}^{(C)} \\ \bar{Y}_{\cdot\cdot}^{(T)} - \bar{Y}_{\cdot\cdot}^{(C)} \end{pmatrix}
 \end{aligned}$$

and

$$\mathbf{D}^{-1} \hat{\boldsymbol{\beta}}_0 = \begin{pmatrix} \frac{1}{v} \bar{Y}_{\cdot\cdot}^{(C)} \\ \frac{1}{v} (\bar{Y}_{\cdot\cdot}^{(T)} - \bar{Y}_{\cdot\cdot}^{(C)}) \end{pmatrix}.$$

This implies

$$\mathbf{F}^\top \mathbf{F} \hat{\boldsymbol{\beta}}_{i;\text{within}} + \mathbf{D}^{-1} \hat{\boldsymbol{\beta}}_0 = \begin{pmatrix} n\bar{Y}_{i\cdot}^{(T)} + (N - n)\bar{Y}_{i\cdot}^{(C)} + \frac{1}{v} \bar{Y}_{\cdot\cdot}^{(C)} \\ n\bar{Y}_{i\cdot}^{(T)} + \frac{1}{v} (\bar{Y}_{\cdot\cdot}^{(T)} - \bar{Y}_{\cdot\cdot}^{(C)}) \end{pmatrix}$$

and

$$\hat{\boldsymbol{\beta}}_i = \begin{pmatrix} c_0(\bar{Y}_{i\cdot}^{(T)} - \bar{Y}_{\cdot\cdot}^{(T)}) + c\bar{Y}_{i\cdot}^{(C)} + (1 - c)\bar{Y}_{\cdot\cdot}^{(C)} \\ c_1\bar{Y}_{i\cdot}^{(T)} + (1 - c_1)\bar{Y}_{\cdot\cdot}^{(T)} - c_2\bar{Y}_{i\cdot}^{(C)} - (1 - c_2)\bar{Y}_{\cdot\cdot}^{(C)} \end{pmatrix}$$

with constants c_0 , c , c_1 , and c_2 as given in (5) and (6).

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Dette H (1997) Designing experiments with respect to 'standardized' optimality criteria. *J R Stat Soc B* 59:97–110
2. Fedorov V, Jones B (2005) The design of multicentre trials. *Stat Methods Med Res* 14:205–248
3. Fedorov V, Leonov S (2013) Optimal design for nonlinear response models. CRC Press, Boca Raton
4. Lemme F, van Breukelen GJP, Berger MPF (2015) Efficient treatment allocation in two-way nested designs. *Stat Methods Med Res* 24:494–512
5. Liski E, Louma A, Mandal N, Sinha B (1998) Optimal designs for prediction in random coefficient linear regression models. *J Combin Inf Syst Sci (J. N. Srivastava Felicitation Volume)* 23:1–16
6. Liski E, Mandal N, Shah K, Sinha B (2002) Topics in optimal designs, vol 163. Lecture notes in statistics. Springer, New York
7. Prus M, Schwabe R (2016) Optimal designs for the prediction of individual parameters in hierarchical models. *J R Stat Soc B* 78:175–191
8. Schwabe R, Schmelter T (2008) On optimal designs in random intercept models. *Tatra Mt Math Publ* 39:145–153

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.