Optimal allocation of sample size for randomization-based inference from 2^K factorial designs

Arun Ravichandran*, Nicole E. Pashley*, Brian Libgober** and Tirthankar Dasgupta*

*Department of Statistics, Rutgers University

**Department of Political Science and Law, Northwestern University

Abstract

Optimizing the allocation of units into treatment groups can help researchers improve the precision of causal estimators and decrease costs when running factorial experiments. However, existing optimal allocation results typically assume a super-population model and that the outcome data comes from a known family of distributions. Instead, we focus on randomization-based causal inference for the finite-population setting, which does not require model specifications for the data or sampling assumptions. We propose exact theoretical solutions for optimal allocation in 2^K factorial experiments under complete randomization with A-, D- and E-optimality criteria. We then extend this work to factorial designs with block randomization. We also derive results for optimal allocations when using cost-based constraints. To connect our theory to practice, we provide convenient integer-constrained programming solutions using a greedy optimization approach to find integer optimal allocation solutions for both complete and block randomization. The proposed methods are demonstrated using two real-life factorial experiments conducted by social scientists.

1. Introduction

Randomized 2^K factorial experiments are conducted to assess the marginal causal effects of K factors, each with two levels, along with their interactions on a response of interest. The two levels are often denoted as the "high level" and "low level" of the factor (Fisher 1935; Yates 1937). With K factors, there are 2^K unique treatment combinations to which units can be assigned. In the

twentieth century, factorial designs have mostly been discussed in an industrial setting, whereas in recent times, there has been a lot of interest in their application in the social, behavioral and biomedical sciences and randomization-based inference from such designs (e.g., Branson et al. 2016; Egami and Imai 2019).

Randomization-based inference is a useful methodology for drawing inference on causal effects of treatments in a finite population setting (e.g., Freedman 2006, 2008). A major advantage of randomization-based inference is that it applies even if the experimental units are not randomly sampled from a larger population, which is the case in most social science experiments (Abadie et al. 2020; Olsen et al. 2013). The theory, methods, and applications of randomization-based inference for two-level factorial experiments with a completely randomized treatment assignment mechanism have been developed and discussed (e.g., Dasgupta et al. 2015; Lu 2016). Further, randomization-based inference from experiments with more general factorial structures and complex assignment mechanisms have been discussed in Mukerjee et al. (2018). Connections between regression-based and randomization-based causal inference from factorial experiments have been studied by Zhao and Ding (2021).

Despite the growing literature in this area, most of the recent research on randomization-based inference of factorial experiments has been confined to the analysis side. On the design side, the main focus has been on rerandomization (Branson et al. 2016; Li et al. 2020; Morgan and Rubin 2012), which generalizes the idea of blocking by pre-defining an acceptable criterion for randomization based on covariate balance between treatment groups. There have also been extensions to fractional and incomplete factorial designs (Pashley and Bind 2022) and to the use of screening steps (Shi et al. 2023). However, the distribution of the total number of experimental units into the 2^K treatment groups has not received much attention. Balanced designs that assign equal number of units to the treatment groups are often the default choice, but it is unclear whether they are the "best" design under different conditions. One work that does discuss how to allocate units to optimize precision of factorial estimators from the randomization-based perspective is Blackwell et al. (2022). That work explores the advantages of Neyman-Allocation (Neyman 1934; Cochran 1977b) by extending the two-stage adaptive design in Hahn et al. (2011) to multiple treatment designs. Dai et al. (2023) similarly explores Neyman-Allocation but within sequential designs.

To motivate this problem, we consider an education experiment from Angrist et al. (2009),

conducted to assess the causal effects of two different interventions, a student support program (SSP) and a student fellowship program (SFP), on the academic performance of freshmen. This is a 2^2 factorial experiment in which each unit (freshman) can receive only one of four treatment combinations: control (neither of the two), SSP only, SFP only, and SFSP (both). The units were divided into two blocks based on their sex. Table 1 shows the allocation of units within each block to the four treatment combinations.

Table 1: Allocation of units to treatment combinations

Sex	Control	SFP	SSP	SFSP
Female	574	150	142	82
Male	432	100	108	68
Total	1006	250	250	150

Clearly, the design is unbalanced, with the highest number of units assigned to control and the fewest to the treatment SFSP. Such an allocation, among other reasons, could be motivated by the budget for experimentation. The question we investigate in this paper is the following: Under what assumptions, conditions, and requirements will such an allocation be the best possible one (in terms of being able to precisely answer scientific questions of interest)?

The problem of finding optimal designs in the context of model-based inference has been extensively studied in the twentieth century (see Atkinson et al. 2007, for example). In such settings, optimal designs depend on a postulated outcome model, that may be linear or non-linear. For example, for binary responses, optimal designs based on logistic models are likely to differ from those based on probit models, and depend on unknown model parameters (Yang et al. 2012; Yang and Mandal 2015; Yang et al. 2016). We aim to develop optimal designs that are tied to *model-free, randomization-based analysis* for finite and super populations. In addition to being robust to model assumptions, our approach works for continuous as well as binary outcomes as long the finite-population estimand is well-defined.

This paper is organized as follows: The next section introduces basic notation and estimands for factorial experiments using the potential outcomes framework. In Section 3 we derive optimal allocations of the N units in a population to different treatment combinations under three commonly used optimality criteria for a completely randomized design (CRD). In addition to theoretical results

for exact optimal allocations, we also provide numerical algorithms for obtaining integer solutions. In Section 4, we extend our results for CRDs to the setting of randomized block designs (RBDs). In Section 5, we derive optimality results under cost constraints. Two different applications of the proposed methodology, the motivating education experiment and an audit experiment conducted to assess discrimination, are described in Section 6. We conclude with a discussion, including opportunities for future work, in Section 7.

2. 2^K factorial experiments under the potential outcomes framework

Here, we introduce some key definitions and notation from Dasgupta et al. (2015). Consider a 2^K experiment with N units, in which the levels of each of the K factors are denoted by 0 and 1. Each treatment combination is of the form $\mathbf{z}_j = (z_1 \ldots, z_K)$, where $z_k \in \{0, 1\}$ for $k \in 1, \ldots, K$. There are $J = 2^K$ treatment combinations arranged in lexicographic order $1, \ldots, J$, where treatment combination $\mathbf{z}_j$ is such that $j = 2^{K-1}z_1 + 2^{K-2}z_2 + \ldots + z_K + 1$. In other words, $(z_1 \ldots z_K)$ is a binary representation of integer $j - 1$. Thus, for example, in a 2^2 experiment, the four treatment combinations 00, 01, 10 and 11 are numbered as $j = 1, 2, 3$ and 4 respectively, and the 8^{th} treatment combination in a 2^4 experiment is 0111. We will just refer to the treatment combination by its number (j) in notation below.

For $i = 1, \ldots, N$, under the Stable Unit Treatment Value Assumption or SUTVA (Rubin 1980), the i^{th} unit has $J = 2^K$ potential outcomes $Y_i(1), \ldots, Y_i(J)$ corresponding to the J treatment combinations $\mathbf{z}_1, \ldots, \mathbf{z}_J$. Let $\mathbf{Y}_i$ denote the $J \times 1$ vector of potential outcomes for unit i. For unit i, the unit-level main effect of factor $k = 1, \ldots, K$ is defined as the difference between the averages of potential outcomes for unit i for which the levels of factor k are at levels 0 versus 1. Mathematically, it is a contrast of the form $2^{-(K-1)}\boldsymbol{\lambda}_k^{\mathrm{T}}\mathbf{Y}_i = 2^{-(K-1)}\sum_{j=1}^{J}\lambda_{jk}\mathbf{Y}_i(j)$, where $\mathbf{x}^{\mathrm{T}}$ denotes the transpose of vector $\mathbf{x}$, $\boldsymbol{\lambda}_k$ is a $J \times 1$ column-vector with coefficient λ_{jk} such that $\lambda_{jk} = -1$ if the level of factor k in j^{th} treatment combination is 0, and $\lambda_{jk} = 1$ otherwise. For all $k = 1, \ldots, K$, $\boldsymbol{\lambda}_k$ is a contrast vector, i.e., $\sum_{j=1}^{J}\lambda_{jk} = 0$.

Proceeding along the lines of Dasgupta et al. (2015), for unit i, we can define $\binom{K}{2}$ two-factor interactions, $\binom{K}{3}$ three-factor interactions, and finally one K-factor interaction as contrasts of the form $2^{-(K-1)}\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Y}_i$, where the contrast vector $\boldsymbol{\lambda}$ for any interaction can be derived by element-wise

multiplication of the contrast vectors of the corresponding main effects $\boldsymbol{\lambda}_k$, for factors involved in the interaction. Denoting the $J-1=2^K-1$ contrast vectors for the $J-1$ unit-level factorial effects $\tau_{1i},\ldots,\tau_{J-1,i}$ by $\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_{J-1}$ respectively, we define the $J\times J$ matrix as

$$\boldsymbol{L}=(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_{J-1}), \tag{1}$$

where $\boldsymbol{\lambda}_0$ is the $J\times 1$ vector with all elements equal to one. We note that $\mathbf{L}$ is an orthogonal matrix with $\mathbf{L}\mathbf{L}^{\mathrm{T}}=\mathbf{L}^{\mathrm{T}}\mathbf{L}=2^{K-1}\mathbf{I}_J$, where $\mathbf{I}_J$ denotes the identity matrix of order J. For $i=1,\ldots,N$, let $\boldsymbol{\tau}_i=(2\tau_{0i},\tau_{1i},\ldots,\tau_{J-1,i})^{\mathrm{T}}$, where τ_{0i} denotes the average of all potential outcomes for unit i. The linear transform between the vector of unit-level potential outcomes $\mathbf{Y}_i$ and the vector of unit-level factorial effects $\boldsymbol{\tau}_i$ can be expressed as

$$\boldsymbol{\tau}_i=2^{-(K-1)}\mathbf{L}^{\mathrm{T}}\mathbf{Y}_i. \tag{2}$$

Having defined unit-level factorial effects, we now move to their population-level counterparts. Let $\overline{\mathbf{Y}}=N^{-1}\sum_{i=1}\mathbf{Y}_i$ and $\boldsymbol{\tau}=N^{-1}\sum_{i=1}\boldsymbol{\tau}_i$ respectively denote the $J\times 1$ vectors of average potential outcomes and the average factorial effects. Then, averaging (2) over $i=1,\ldots,N$, the vector of population-level factorial effects is given by

$$\boldsymbol{\tau}=2^{-(K-1)}\mathbf{L}^{\mathrm{T}}\overline{\mathbf{Y}}. \tag{3}$$

Note that the first element of $\boldsymbol{\tau}$ is twice the average of all potential outcomes.

In a CRD, a pre-assigned numbers of units, N_j, are randomly assigned to treatment j. The experiment generates an $N\times 1$ vector of observed outcomes data from which the vector of factorial effects $\boldsymbol{\tau}$ can be unbiasedly estimated. We examine the properties of $\widehat{\boldsymbol{\tau}}$, the unbiased estimator of $\boldsymbol{\tau}$, with respect to its randomization distribution, and formulate the problem of optimally allocating the N units to the J treatment combinations in Section 3.

Remark 1 (A super population perspective)**.** While the finite-population perspective does not depend on any hypothetical data generating process for the outcomes, alternative approaches assume that the potential outcomes are drawn from a, possibly hypothetical, super population. Assuming that $\mathbf{Y}_1,\ldots,\mathbf{Y}_N$ are independent and identically distributed random vectors with $E[\mathbf{Y}_i]=\boldsymbol{\mu}$,

factorial effects at a super population level are defined as

$$\boldsymbol{\tau}^{\mathrm{SP}} = 2^{-(K-1)}\mathbf{L}^{\mathrm{T}}\boldsymbol{\mu}.$$

Ding et al. (2017) discussed the conceptual and mathematical connections between finite- and super-population inference, showing that while the same estimator commonly used to estimate $\boldsymbol{\tau}$ unbiasedly is also an unbiased estimator of $\boldsymbol{\tau}^{\mathrm{SP}}$, its sampling variances under the two perspectives are different.

3. Optimal designs for completely randomized experiments

In a randomized experiment with N_j units assigned to treatment combination $j \in \{1, \ldots, J\}$, only one of the J potential outcomes is observed for unit i. This observed outcome is $y_i = Y_i(T_i)$ for $i = 1, \ldots, N$, where T_i is the random treatment assignment variable for unit i taking value j if unit i receives treatment j. In a CRD, the joint probability distribution of $(T_1, \ldots, T_N)$ is

$$P[(T_1, \ldots, T_N) = (t_1, \ldots, t_N)] = \begin{cases} (N_1! \ldots N_J!)/N! & \text{if } \sum_{i=1}^{N} \mathbb{1}_{\{t_i=j\}} = N_j \text{ for } j = 1, \ldots, J, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbb{1}_{\{A\}}$ denotes the indicator random variable for set A. Let $\overline{y}(j) = N_j^{-1} \sum_{i=1}^{N} \mathbb{1}_{\{t_i=j\}} Y_i(j)$ denote the average response for treatment j. Let $\overline{\mathbf{y}}$ denote the vector $(\overline{y}(1), \overline{y}(2), \ldots, \overline{y}(J))^{\mathrm{T}}$ of observed averages. Substituting $\overline{\mathbf{y}}$ in place of $\overline{\mathbf{Y}}$ in (3), we can unbiasedly estimate the vector of factorial effects as

$$\widehat{\boldsymbol{\tau}} = 2^{-(K-1)}\mathbf{L}^{\mathrm{T}}\overline{\mathbf{y}}. \tag{4}$$

Lu (2016) derived sampling properties of the estimator $\widehat{\boldsymbol{\tau}}$ with respect to its randomization distribution for the general case of unequal $N_1, \ldots, N_J$. Lu showed that $\widehat{\boldsymbol{\tau}}$ is an unbiased estimator of $\boldsymbol{\tau}$, and has the following finite-population covariance matrix:

$$\mathbf{V}_{\boldsymbol{\tau}} = \mathrm{Var}\left(\widehat{\boldsymbol{\tau}}\right) = \frac{1}{2^{2(K-1)}} \sum_{j=1}^{2^K} \frac{S_j^2}{N_j} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}} - \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\boldsymbol{\tau}_i - \boldsymbol{\tau}\right)\left(\boldsymbol{\tau}_i - \boldsymbol{\tau}\right)^{\mathrm{T}}, \tag{5}$$

where $\widetilde{\boldsymbol{\lambda}}_j$ represents the transpose of row j of the model matrix $\mathbf{L}$ defined in (1), $\boldsymbol{\tau}_i$ denotes the vector of unit-level factorial effects given by (2), $\boldsymbol{\tau}$ the vector of population-level factorial effects given by (3), and

$$S_j^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(Y_i(j) - \overline{Y}(j) \right)^2 \tag{6}$$

the variance of all N potential outcomes for treatment j with divisor $N-1$, where $\overline{Y}(j) = N^{-1} \sum_{i=1}^{N} Y_i(j)$.

In the spirit of classical optimal designs (Atkinson et al. 2007), we can define a design optimality criterion as a functional of the matrix $\mathbf{V}_{\boldsymbol{\tau}}$ defined in (5). For example, the D-optimality criterion, which aims to minimize the determinant of the covariance matrix, or the A-optimality criterion, which aims to minimize the trace of the covariance matrix, or the E-optimality criterion which aims to minimize the maximum eigenvalue of the covariance matrix, can be considered. However, the second term $1/(N(N-1)) \sum_{i=1}^{N} (\boldsymbol{\tau}_i - \boldsymbol{\tau}) (\boldsymbol{\tau}_i - \boldsymbol{\tau})^{\mathrm{T}}$, which is a measure of heterogeneity of treatment effects, cannot be estimated from observed data, because none of the unit-level treatment effects $\boldsymbol{\tau}_i$ are estimable without additional assumptions due to the missing potential outcomes. Because $1/(N(N-1)) \sum_{i=1}^{N} (\boldsymbol{\tau}_i - \boldsymbol{\tau}) (\boldsymbol{\tau}_i - \boldsymbol{\tau})^{\mathrm{T}}$ is positive semi-definite, the first term of (5) can be considered an upper bound of $\mathbf{V}_{\boldsymbol{\tau}}$, which is attained under specific restrictions on the potential outcomes (e.g., treatment effect homogeneity). Thus, we propose optimizing a functional of the first term of (5), which in turn is equivalent to optimizing a functional of the positive definite matrix

$$\widetilde{\mathbf{V}} = \sum_{j=1}^{J} \frac{S_j^2}{N_j} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}} = \mathbf{L}^{\mathrm{T}} \mathbf{A} \mathbf{L},$$

instead, where $\mathbf{A} = \mathrm{diag}(S_1^2/N_1, \ldots, S_J^2/N_J)$.

Another justification for using a functional of the matrix $\widetilde{\mathbf{V}}$ as an optimality criterion comes from the super-population perspective mentioned in Section 1. Ding et al. (2017) showed that the estimator $\widehat{\boldsymbol{\tau}}$ defined earlier is also an unbiased estimator of the super-population estimand $\boldsymbol{\tau}^{\mathrm{SP}}$. Further, extending their argument for a single factor with two levels to the case of 2^K factorial designs, if $V_j^2 = \mathrm{Var}[Y_i(j)]$, $j = 1, \ldots, J$, then a variance decomposition yields the following

sampling variance of $\widehat{\boldsymbol{\tau}}$

$$\mathrm{Var}^{\mathrm{SP}}\left(\widehat{\boldsymbol{\tau}}\right) = \frac{1}{2^{2(K-1)}} \sum_{j=1}^{2^K} \frac{V_j^2}{N_j} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}} = \frac{1}{2^{2(K-1)}} \sum_{j=1}^{2^K} \frac{E(S_j^2)}{N_j} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}},$$

where $\mathrm{Var}^{\mathrm{SP}}$ denotes variance over the random sampling from the super population and random assignment, and $V_j^2 = E(S_j^2)$ represents expectation of S_j^2 with respect to the distribution of the potential outcomes in the super population. This connection provides further motivation for the form of our optimization, but we focus on the finite-population setting going forward.

3.1. Exact optimal designs

The problem of finding an optimal design can be formulated as minimization of an appropriate functional $\psi(\widetilde{\mathbf{V}})$ subject to the constraint $\sum_{j=1}^{J} N_j = N$ or equivalently as $\sum_{j=1}^{J} p_j = 1$ in terms of the proportions of units $p_j = N_j/N$ to be assigned to treatment combination j. As discussed earlier, we consider three widely used functionals in optimal design literature: the D-optimality criterion where $\psi(\widetilde{\mathbf{V}}) = \left|\widetilde{\mathbf{V}}\right|$ and $|.|$ refers to the determinant, the A-optimality criterion where $\psi(\widetilde{\mathbf{V}}) = \mathrm{tr}\left(\widetilde{\mathbf{V}}\right)$, and the E-optimality criterion where $\psi(\widetilde{\mathbf{V}}) = \max\{\nu_1, \ldots, \nu_J\}$ and $\nu_1, \ldots, \nu_J$ are the eigenvalues of $\widetilde{\mathbf{V}}$. The following theorem, proved in Supplementary Material A, summarizes these optimality results.

Theorem 1. Let N units be allocated to J treatment groups such that $p_j = N_j/N$ proportion of units receive treatment j. Then, the optimal allocation of N units to J treatment groups on the basis of covariance matrix $\widetilde{\boldsymbol{V}}$ under

(a) A-optimality is proportional to the finite-population standard deviations of potential outcomes in the corresponding treatment groups, i.e., $p_j = S_j/(\sum_j S_j)$.

(b) D-optimality is balanced assignment to all J treatment groups, i.e., $p_j = 1/J$.

(c) E-optimality is proportional to the finite-population variances of potential outcomes in the corresponding treatment groups, i.e., $p_j = S_j^2/(\sum_j S_j^2)$.

Remark 2. The optimality results in Theorem 1 are similar in spirit to the determination of optimal sample sizes in stratified survey sampling (Cochran 1977a, Ch. 5). The A-optimality result

is the same as the Neyman-Allocation discussed in Blackwell et al. (2022), who motivate its use as reducing the identifiable portion of the finite-population variance of classical factorial estimators. Derivations of the A- and D- optimal designs are straightforward and use the Lagrangian multiplier-based optimization technique. The proof of the E-optimality result uses the idea of perturbing the eigenvalues of a scaled identity matrix to show that the E-optimal design is indeed characterized by equal eigenvalues of the matrix $\widetilde{\mathbf{V}}$.

Remark 3. In order to implement the results of Theorem 1, researchers need to "guess" the variances S_j^2, $j = 1, \ldots, J$ and substitute them into the expressions for p_j. This is similar to the application of optimal designs in non-linear models, where optimal designs are actually "locally optimal" (Chernoff 1953). It is often a common practice to conduct pilot studies to obtain some preliminary estimates of the S_j^2's, as done in finite-population survey sampling.

We now introduce two conditions associated with the matrix of potential outcomes under which Theorem 1 can be further simplified.

Condition 1 (Homoscedasticity)**.** We call an $N \times J$ matrix of potential outcomes *homoscedastic* if each column has the same variance i.e., $S_j^2 = S^2$ for $j = 1, \ldots, J$.

Condition 2 (Strict additivity)**.** Following Dasgupta et al. (2015), we call an $N \times J$ matrix of potential outcomes *strictly additive* if $Y_i(j) - Y_i(\widetilde{j}) = \tau(j, \widetilde{j})$ for all $j \neq \widetilde{j} \in \{1, 2, \ldots, J\}$. Potential outcomes satisfying this condition also satisfy Condition 1.

The following corollary of Theorem 1 is straightforward but useful:

Corollary 1. If the matrix of potential outcomes satisfies Condition 1, then the A-, D- and E-optimal designs are all balanced designs with $p_j = 1/J$ for $j = 1, \ldots, J$. Further, under Condition 2, optimizations based on $\widetilde{\boldsymbol{V}}$ and $\mathbf{V}_{\boldsymbol{\tau}}$ are equivalent.

While Theorem 1 provides results on exact optimal designs in terms of proportions p_j, experimenters need integer solutions in terms of N_j's satisfying $\sum_j N_j = N$. For example, while the exact D-optimal design is balanced, N is not necessarily a multiple of 2^K, and the result does not provide a D-optimal allocation of, say, 69 units to the 8 treatment combinations in a 2^3 factorial experiment. Thus we need approximate integer solutions to the optimization problem in which additional

constraints on the sample sizes assigned to specific treatment groups can also be introduced. Next, we discuss an integer programming approach to obtain such solutions.

3.2. Computation of exact optimal designs using an integer programming approach

Many sources in the integer programming literature address constrained optimization under integer space constraints (e.g. see Nemhauser and Wolsey 1988; Schrijver 1998; Khan 1995; Sofi et al. 2020). We adopt the methods proposed in Friedrich et al. (2015) that are designed for settings very similar to the ones we consider. Friedrich et al. (2015) consider the following integer programming problem:

$$\min_{N_1,\ldots,N_J} f(N_1,\ldots,N_J) \text{ s.t. } \begin{cases} \sum_{j=1}^{J} N_j = N, \\ l_j \leq N_j \leq u_j, \ \forall j = 1,2,\ldots,J, \\ N_j \in \mathbb{Z}_+^J, \ \forall j = 1,2,\ldots,J. \end{cases} \tag{7}$$

where, $\mathbb{Z}_+$ is the set of positive integers, (l_j, u_j) are the lower and upper bound constraints on N_j and $f : \mathbb{R}_+^J \to \mathbb{R}$ is a convex function. If $f(N_1,\ldots,N_J)$ is separable (i.e., can be expressed as $\sum_{j=1}^{J} f_j(N_j)$) then the greedy algorithm in Figure 1 finds the globally optimal integer solution of the minimization problem given in (7) under some regularity conditions that we show in Supplementary Material B.

From the proof of Theorem 1 in Supplementary Material A, it follows that the A-optimality and D-optimality criteria can respectively be expressed as $\sum_{j=1}^{J}(S_j^2/N_j)$ and $\sum_{j=1}^{J}(\log S_j^2/N_j)$. In Supplementary Material B, we show that the conditions for global convergence of the greedy algorithm are met and thus, convergence to the true integer optimal solution in the cases of A- and D-optimality are guaranteed if this algorithm is used. Hence, substitution of S_j^2/N_j and $\log S_j^2/N_j$ for $f_j(N_j)$ in the algorithm described in Figure 1 leads to optimal integer solutions for the A-optimality criterion and the D-optimality criterion, respectively. In our implementation of this algorithm, we take $l_j = 2$ for $j = 1,\ldots,J$ to guarantee at least two units are assigned to each treatment combination, allowing variance estimation within each treatment group.

We provide a modified greedy algorithm described in Figure 2 for solving the E-optimality problem since the optimization problem cannot be written in a separable form as in A- or D-

Figure 1: Greedy algorithm for separable functions

1. Set $I \leftarrow \{1, ..., J\}$.
2. Let $t = 0$. Initialize $N_j^{(t)} \leftarrow l_j$ for $j = 1, \ldots, J$.
3. While $(\sum_{j=1}^{J} N_j^{(t)} \neq N \;\&\; I \neq \phi)$ do
   - for all $(j \in I)$, $\delta_j \leftarrow f_j(N_j^{(t)} + 1) - f_j(N_j^{(t)})$
   - choose $\mathcal{J} \leftarrow \text{argmin}_{j \in I}\, \delta_j$
   - $j^* \leftarrow \min \mathcal{J}$
   - if $N_{j^*}^{(t)} + 1 \leq u_{j^*}$, then $N_{j^*}^{(t+1)} \leftarrow N_{j^*}^{(t)} + 1$, $t \leftarrow t+1$ else $I \leftarrow I \setminus \{j^*\}$
4. Return optimal $(N_1, \ldots, N_J)$.

optimality criterion above. In this algorithm we take $f_j(N_j) = S_j^2/N_j$ and $l_j = 2$. In Section 4.2, the ability of this greedy algorithm to find E-optimal solutions is demonstrated empirically for blocked designs, discussed next.

Figure 2: Greedy algorithm for E-optimality

1. Set $I \leftarrow \{1, ..., J\}$.
2. Let $t = 0$. Initialize $N_j^{(t)} \leftarrow l_j$ for $j = 1, \ldots, J$.
3. While $(\sum_{j=1}^{J} N_j^{(t)} \neq N \;\&\; I \neq \phi)$ do
   - choose $\mathcal{J} \leftarrow \text{argmax}_{j \in I}\, f_j(N_j^{(t)})$
   - $j^* \leftarrow \min \mathcal{J}$
   - if $N_{j^*}^{(t)} + 1 \leq u_{j^*}$, then $N_{j^*}^{(t+1)} \leftarrow N_{j^*}^{(t)} + 1$, $t \leftarrow t+1$ else $I \leftarrow I \setminus \{j^*\}$
4. Return optimal $(N_1, \ldots, N_J)$.

4. Optimal allocation for factorial experiments with blocks

Consider a block-randomized 2^K factorial design with H blocks. That is, units are pre-assigned membership to one of h blocks based on some similarity metric (we do not consider how to form blocks here). Let M_h denote the size of block h, $h = 1, \ldots, H$. Also, let $M_{h,j}$ be the number of units in block h assigned to the treatment j $(j = 1, \ldots, J)$. Finally, let $b_i(h)$, $i = 1, \ldots, N$ be an

indicator variable taking value 1 if unit i belongs to block h and 0 otherwise. Treatment assignment under a factorial RBD is equivalent to performing an independent CRD, as described in Section 3, within each block.

The population average treatment effect $\boldsymbol{\tau}$ can be expressed as $\sum_h M_h \boldsymbol{\tau}_h / N$, where $\boldsymbol{\tau}_h$ is the block-level vector of factorial effects and its estimator $\widehat{\boldsymbol{\tau}}$ is a weighted average of $\widehat{\boldsymbol{\tau}}_h$, an unbiased estimator of $\boldsymbol{\tau}_h$ defined in the same way as in (4) for block h. Extending (5) by noting the independence of the assignment to treatment across blocks, the covariance matrix of $\widehat{\boldsymbol{\tau}}_h$ in a factorial RBD can thus be obtained as

$$\boldsymbol{V}_{\boldsymbol{\tau}_h} = \mathrm{Cov}\left(\widehat{\boldsymbol{\tau}}_h\right) = \frac{1}{2^{2(K-1)}} \sum_{j=1}^{2^K} \frac{1}{M_{h,j}} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}} S_{h,j}^2 - \frac{1}{M_h(M_h-1)} \sum_{i=1}^{N} b_i(h) \left(\boldsymbol{\tau}_i - \boldsymbol{\tau}_h\right) \left(\boldsymbol{\tau}_i - \boldsymbol{\tau}_h\right)^{\mathrm{T}}, \quad (8)$$

where $\widetilde{\boldsymbol{\lambda}}_j$ represents the transpose of row j of the model matrix $\mathbf{L}$ defined in (1) and $S_{h,j}^2$ denotes the variance of all M_h potential outcomes for units in block h under treatment j with divisor $M_h - 1$. The covariance matrix of $\widehat{\boldsymbol{\tau}}$ can then be expressed as $\mathrm{Cov}\left(\sum_h M_h \widehat{\boldsymbol{\tau}}_h / N\right)$. Because the block-level treatment estimators $\widehat{\boldsymbol{\tau}}_h$ are independent across blocks, we have

$$\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\tau}}\right) &= \sum_{h=1}^{H} \frac{M_h^2}{N^2} \boldsymbol{V}_{\boldsymbol{\tau}_h} \\
&= \frac{1}{2^{2(K-1)}} \sum_{j=1}^{2^K} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}} \left[\sum_{h=1}^{H} \frac{M_h^2}{N^2} \frac{S_{h,j}^2}{M_{h,j}}\right] - \sum_{h=1}^{H} \frac{M_h^2}{N^2} \frac{1}{M_h(M_h-1)} \sum_{i=1}^{N} b_i(h) \left(\boldsymbol{\tau}_i - \boldsymbol{\tau}_h\right) \left(\boldsymbol{\tau}_i - \boldsymbol{\tau}_h\right)^{\mathrm{T}}.
\end{aligned}$$

Writing $S_{\mathrm{blk},j}^2 = \sum_{h=1}^{H} (M_h^2/N^2)(S_{h,j}^2/M_{h,j})$ and proceeding along similar lines as in Section 3, we formulate a surrogate optimization problem to only optimize the first term, since the second term in the equation above is not identifiable. Then, choosing $M_{h,j}$'s to optimize some functional of $\mathrm{Cov}\left(\widehat{\boldsymbol{\tau}}\right)$ is equivalent to optimizing a functional of the matrix

$$\widetilde{\boldsymbol{V}}_{\mathrm{blk}} = \sum_{j=1}^{2^K} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}} S_{\mathrm{blk},j}^2 = \boldsymbol{L}^{\mathrm{T}} \boldsymbol{A}_{\mathrm{blk}} \boldsymbol{L}, \quad (9)$$

where $\boldsymbol{A}_{\mathrm{blk}} = \mathrm{diag}\left(S_{\mathrm{blk},1}^2, \ldots, S_{\mathrm{blk},J}^2\right)$.

4.1. Exact optimal designs

The problem of finding an optimal design for RBDs can be formulated as minimization of an appropriate functional $\psi(\widetilde{\boldsymbol{V}}_{\text{blk}})$ subject to the constraint $\sum_{j=1}^{J} M_{h,j} = M_h$ for each h or equivalently as $\sum_{j=1}^{J} p_{h,j} = 1$ in terms of the proportions of units to be assigned to treatment combination j in block h, $p_{h,j} = M_{h,j}/M_h$. While the A-optimality result is straightforward, finding exact D-optimal and E-optimal solutions in the setting with blocks is difficult without imposing restrictions on the potential outcomes. Before stating the optimality results, we first introduce two such restrictions that generalize Condition 1 to a block setting.

Condition 3 (Within-block homoscedasticity, WBH)**.** We call an $N \times J$ matrix of potential outcomes in H blocks to be *within-block homoscedastic (WBH)* if within each block, all treatment columns have the same variance, i.e., within block h, $S^2_{h,j} = S^2_{h\cdot}$ for $j = 1, \ldots, J$.

Condition 4 (Between-block homoscedasticity, BBH)**.** We call an $N \times J$ matrix of potential outcomes in H blocks to be *between-block homoscedastic (BBH)* if for each treatment column j, the variance of potential outcomes in each block is the same, i.e., $S^2_{h,j} = S^2_{\cdot,j}$ for $h = 1, \ldots, H$ and $j = 1, \ldots, J$.

The following theorem now summarizes the optimality results for blocked designs.

Theorem 2. Let N units be distributed across H blocks such that there are M_h units in block h and $N = \sum_{h=1}^{H} M_h$. Let each set of M_h units be allocated to J treatment groups such that $p_{h,j}$ proportion of the units are allocated to treatment j in block h, and let $S^2_{h,j}$ as defined in (8). Then, optimal allocation of M_h units to the J treatment groups on the basis of covariance matrix $\widetilde{\boldsymbol{V}}_{\text{blk}}$ under different optimality criteria can be summarized as follows.

(a) The A-optimal allocation is the same as the A-optimal CRD allocation within each block, i.e., $p_{h,j} = S_{h,j}/(\sum_{j=1}^{J} S_{h,j})$ for each h.

(b) If either (or both) of Conditions 3 (WBH) and 4 (BBH) hold, the D-optimal allocation is the balanced assignment within each block, i.e., $p_{h,j} = 1/J$ for each h.

(c) If Condition 3 (WBH) holds, the E-optimal allocation is the balanced design within each block, i.e., $p_{h,j} = 1/J$ for each h.

Remark 4. Condition 2 for all N units implies both WBH and BBH. Consequently, by Theorem 2, the D- and E-optimal allocation for strictly additive potential outcomes in randomized block designs is a balanced assignment within each block.

Remark 5. The A-optimal allocation under WBH is balanced allocation within each block (same as D- and E-optimal allocations). However, under BBH, the A-optimal allocation is different from the D- and E-optimal allocations, and is proportional to the standard deviations of the treatments $S_{\cdot,j}$ that is constant across blocks.

4.2. Computation of exact optimal designs for factorial RBDS using an integer programming approach

As in the case of completely randomized factorial designs, whereas Theorem 2 provides results on exact optimal designs in terms of proportions $p_{h,j}$, experimenters need integer solutions in terms of $M_{h,j}$'s in which additional constraints on the sample sizes assigned to specific treatment groups can also be introduced. Further, Theorem 2 provides D- and E-optimal solutions only under specific conditions like WBH and BBH. Thus, we discuss an integer programming approach to obtain integer solutions for settings covered and not covered by Theorem 2.

We can use the same algorithm in Figure 1 within each block to obtain the optimal integer solutions for A-optimality under the RBD by replacing the function $f_j(.)$ by $f_{h,j}(M_{h,j}) = (M_h^2/N^2)(S_{h,j}^2/M_{h,j})$. For D- and E-optimality, we extend the greedy idea from the algorithm in Figure 2 with minor changes to the function $f_j(.)$. We take $f_{h,j}(M_{h,j}) = \log(\sum_{h=1}^{H}(M_h^2/N^2)(S_{h,j}^2/M_{h,j}))$ for D-optimality and $f_{h,j}(M_{h,j}) = (M_h^2/N^2)(S_{h,j}^2/M_{h,j})$ for E-optimality. The exact algorithms taking the structure of the blocks into account are given in Supplementary Material D. The main difference between the algorithm used in Section 3 and the one proposed here lies in the fact that now we have to allocate the next best unit at iteration t over an $H \times J$ matrix $((M_{h,j}^{(t)}))$ with upper bounds on row sums, rather than a vector $(N_1^{(t)}, \ldots, N_J^{(t)})$ with an upper bound on the sum of the elements, as in the case of the CRD. Note that, by Friedrich et al. (2015), the greedy algorithm finds the correct solution in the case of A-optimality for the block design. It, however, does not extend to the D- and E-optimality in the block case, due to nature of the objective functions.

We now conduct an empirical exploration of the performances of the greedy algorithm in terms

of its ability to find E-optimal solutions. Five different settings of 2^2 factorial designs in two blocks, each corresponding to a specific type of potential outcome matrix, are considered. Each setting is defined by the block sizes M_1 and M_2, and the 4×2 matrix of variances $S^2_{h,j}$ as shown in columns 3 and 4 of Table 2, respectively.

The first setting considers blocks of equal sizes with potential outcomes satisfying Condition 2 (strict additivity), leading to an E-optimal design that is balanced within each block as per Remark 4. The second setting considers equal block sizes with potential outcomes satisfying Condition 3 (WBH), leading to a balanced design by Theorem 2. In this setting and the previous one, the exact optimal designs provide optimal integer solutions. This is not the case in the third setting, which considers unequal block sizes with potential outcomes satisfying Condition 4 (BBH) but not Condition 3 (WBH). Theorem 2 does not apply directly for E-optimality, but the greedy algorithm identifies the unique true E-optimal allocation determined by the exhaustive search. The fourth setting is similar the third case above, but is one where the exhaustive search provides multiple solutions, identifying four different allocations, each of which is optimal. In this case, Theorem 2 does not apply directly and the greedy algorithm identifies one of these solutions. The fifth setting neither satisfies Condition 3 (WBH) nor Condition 4 (BBH), and consequently Theorem 2 cannot provide an exact E-optimal solution. However, the greedy algorithm identifies one of the two (identified through exhaustive search) true optimal integer allocations. A quick note on our greedy algorithm is that, due to the nature of the algorithm in Figures 3 and 4 of Supplementary Material D, ties are broken deterministically using minimum index when the greedy step returns more than one solution. Thus, our greedy solutions will always achieve the same solution for a given set of inputs without regard for the plurality of solutions (such as the ones in the fourth and fifth setting above).

A similar exploration performed for the D-optimal allocation (shown in Supplementary Material C) provides evidence that the greedy algorithm can identify the true optimal integer solution when it is unique, and *one* of the true optimal solutions when multiple optimal integer solutions exist.

S.No.	Case	Block Size (M_h)	Variances ($S^2_{h,j}$)	Exhaustive search E-optimal solution	Greedy solution
1.	Equal blocks with equal variances	$\begin{bmatrix}40\\40\end{bmatrix}$	$\begin{bmatrix}1&1&1&1\\1&1&1&1\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$
2.	Equal blocks with equal variances for all treatments within block	$\begin{bmatrix}40\\40\end{bmatrix}$	$\begin{bmatrix}4&4&4&4\\1&1&1&1\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$
3.	Unequal blocks with equal variances across blocks for each treatment	$\begin{bmatrix}40\\20\end{bmatrix}$	$\begin{bmatrix}1&2&3&4\\1&2&3&4\end{bmatrix}$	$\begin{bmatrix}4&8&12&16\\2&4&6&8\end{bmatrix}$	$\begin{bmatrix}4&8&12&16\\2&4&6&8\end{bmatrix}$
4.	Unequal blocks with equal variances but exact solution is non-integer	$\begin{bmatrix}40\\20\end{bmatrix}$	$\begin{bmatrix}1&2&3&5\\1&2&3&5\end{bmatrix}$	$\left[\begin{bmatrix}4&8&11&17\\2&3&5&10\end{bmatrix}\right.$ $\begin{bmatrix}4&7&11&18\\2&4&5&9\end{bmatrix}$ $\begin{bmatrix}3&8&11&18\\3&3&5&9\end{bmatrix}$ $\left.\begin{bmatrix}3&7&11&19\\3&4&5&8\end{bmatrix}\right]$	$\begin{bmatrix}4&7&11&18\\2&4&5&9\end{bmatrix}$
5.	Equal blocks with unequal variances	$\begin{bmatrix}40\\40\end{bmatrix}$	$\begin{bmatrix}1&2&3&4\\4&3&2&1\end{bmatrix}$	$\left[\begin{bmatrix}6&10&11&13\\13&11&10&6\end{bmatrix}\right.$ $\left.\begin{bmatrix}6&9&12&13\\13&12&9&6\end{bmatrix}\right]$	$\begin{bmatrix}6&9&12&13\\13&12&9&6\end{bmatrix}$

Table 2: Summary of Greedy algorithm solutions for E-optimality for $H = 2, K = 2$

5. Optimal allocation driven by cost constraints

So far, we have considered optimality criteria that are based on the covariance matrix of the estimated factorial effects, implicitly assuming that all treatment combinations are equally expensive (with respect to cost and/or time). However, such assumptions may not be true in many practical situations and cost constraints can play an important role in determining optimal allocation. Thus it is worthwhile to explore solutions to optimal allocation under cost constraints.

We consider the optimal allocation for 2^K factorial CRDs. Let the cost of assigning treatment combination j to one unit be $C_j > 0$, and the total available budget be C. In the new optimization problem, we replace the constraint $\sum_j N_j = N$ in the original problem described in Section 3.1 by the cost constraint $\sum_j C_j N_j \leq C$. The new optimization problem is therefore:

$$\min_{N_j} \psi\left(\widetilde{\mathbf{V}}\right) \text{ subject to } \sum_j C_j N_j \leq C, \tag{10}$$

where $\widetilde{\mathbf{V}} = \sum_{j=1}^{J} \frac{S_j^2}{N_j} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}}$ and $\psi(\widetilde{\mathbf{V}})$ is a functional of $\widetilde{\mathbf{V}}$. A straightforward approach to incorporate this new constraint into our previous setting is to re-write the constraint as $\sum_j \widetilde{N}_j \leq C$ where $\widetilde{N}_j = N_j C_j$ is the total cost for the suggested allocation to treatment arm j. Under this one-to-one transformation $\tilde{N}_j = C_j N_j$, the optimization problem in (10) is equivalent to minimizing the objective function over $\widetilde{N}_j$ (Boyd and Vandenberghe 2004), and can be written as:

$$\begin{aligned} \min_{\tilde{N}_j} \psi\left(\widetilde{\mathbf{V}}\right) &= \min_{\tilde{N}_j} \psi\left(\sum_{j=1}^{J} \frac{S_j^2}{N_j} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}}\right) = \min_{\tilde{N}_j} \psi\left(\sum_{j=1}^{J} \frac{S_j^2}{\tilde{N}_j / C_j} \widetilde{\boldsymbol{\lambda}}_j \widetilde{\boldsymbol{\lambda}}_j^{\mathrm{T}}\right), \\ &\text{subject to } \sum_j \tilde{N}_j \leq C. \end{aligned}$$

Because the optimal solution of the above optimization problem is attained at $\sum_j \widetilde{N}_j = C$, the inequality constraint can be replaced by the equality constraint. Then, proceeding along the lines of Theorem 1, one can obtain the cost for the optimal allocation to treatment arm j as $\widetilde{N}_j \propto S_j \sqrt{(C_j)}$, $\tilde{N}_j = C/J$ and $\widetilde{N}_j \propto S_j^2 C_j$ as the A-, D- and E- optimal solutions. These results are formalized in terms of the optimal proportion of the budget allocated to each treatment arm, which can be used to determine the number of units to assign to each treatment arm, in the theorem below.

Theorem 3. Let C be total budget for the whole experiment and the cost of allocating one experimental unit to treatment j be $C_j > 0$. Let $\pi_j = C_j N_j / C$ denote the proportion of the total budget assigned to treatment j with $\sum_j \pi_j \leq 1$. Then, the

(a) A-optimal cost-based allocation to the J treatment groups on the basis of covariance matrix $\widetilde{\boldsymbol{V}}$ is $\pi_j = (S_j \sqrt{C_j})/(\sum_j S_j \sqrt{C_j})$.

(b) D-optimal cost-based allocation to the J treatment groups on the basis of covariance matrix $\widetilde{\boldsymbol{V}}$ is $\pi_j = 1/J$.

(c) E-optimal cost-based allocation to the J treatment groups on the basis of covariance matrix $\widetilde{\boldsymbol{V}}$ is $\pi_j = (S_j^2 C_j)/(\sum_j S_j^2 C_j)$.

Remark 6. Theorem 3 can be extended to the case of block designs along the lines of Theorem 2.

Remark 7. If for $j = 1, \ldots, J$, the costs C_j in Theorem 3 are the same and equal to C_0, then the constraint $\sum_j C_j N_j \leq C$ reduces to $\sum_j N_j \leq N$ where $N = C/C_0$. Also, $\pi_j = C_0 N_j / C \equiv p_j$. Thus

the optimization problem becomes the same as the one in Theorem 1, making it a special case of Theorem 3.

We use an example to demonstrate the applicability of Theorem 3. Consider a 2^2 factorial design, with $C = 100$, per unit cost vector $(C_1, \ldots, C_4) = (0.1, 4, 4, 9)$. This set up represents many common scenarios where treatment arm 1 represents the control group 00 and involves a per-unit cost that is negligible compared to the ones with at least one active treatment. On the other hand, treatment arm 4 has both treatments at active level and involves the highest cost. Table 3 shows the A-, D- and E- optimal proportions of total cost π_j's for two different vectors of variances $(S_1^2, \ldots, S_4^2)$. In one setting, we take the vector as $(1, 1, 1, 1)$ and in another, set it to $(1, 2, 3, 4)$. For the sake of completeness, we also add a column of equal cost $(1, 1, 1, 1)$, under which the p_j's of Theorem 1 and π_j's of Theorem 3 become identical, as explained in Remark 7. Thus, the optimal allocations in the first column of Table 3 can also be derived from Theorem 1 with $N = C = 100$.

Table 3: Optimal π_j's under cost constraints obtained from Theorem 3

Variance Vector	Type of Optimality	Cost vector $(C_1, \ldots, C_4)$ $(1,1,1,1)$	$(0.1,4,4,9)$
(1,1,1,1)	A	(0.250,0.250,0.250,0.250)	(0.043,0.273,0.273,0.410)
	D	(0.250,0.250,0.250,0.250)	(0.250,0.250,0.250,0.250)
	E	(0.250,0.250,0.250,0.250)	(0.006,0.234,0.234,0.526)
(1,2,3,4)	A	(0.163,0.230,0.282,0.325)	(0.025,0.224.0.275,0.476)
	D	(0.250,0.250,0.250,0.250)	(0.250,0.250,0.250,0.250)
	E	(0.100,0.200,0.300,0.400)	(0.002,0.143,0.214,0.642)

One can obtain the number of units N_j's the A-, D- and E-optimal allocations of N_j's by substituting the optimal π_js from Theorem 3 into $N_j = (C\pi_j)/C_j$. However, rounding these optimal N_j's into nearest integers may lead to violation of the constraint $\sum_j C_j N_j \leq C$. To avoid such possibilities, one can consider the optimal values of $\lfloor (C\pi_j)/C_j \rfloor$ as approximate integer solutions, where $\lfloor x \rfloor$ denotes the largest integer contained in x.

Researchers may decide to impose an additional constraint on the optimization problem (11) that forces the sum of N_j's to be exactly equal to a predetermined N. Such a problem would give optimal allocation under fixed N, unlike Theorem 3. However, imposing this additional constraint may force the set of feasible solutions to the optimization problem to be empty. For example, suppose for all j, $C_j > C/N$. Then, $\sum_j C_j N_j > \sum_j (C/N) N_j$, which exceeds the allowable cost C if the restriction $\sum_j N_j = N$ is imposed. Thus, additional conditions are necessary to guarantee

that the feasible set is non-empty. Obtaining closed-form solutions under such conditions may not be straightforward and one may need to rely on numerical methods to obtain such solutions.

6. Applications in real experiments

In this section, we demonstrate applications of the results and algorithms developed to two real-life experiments. First, we re-visit the education example from Angrist et al. (2009) described in Section 1. Second, we discuss a pilot audit experiment reported in Libgober (2020) conducted to identify how perceptions of race, gender and affluence affect access to lawyers, and demonstrate how the proposed methodology can be used to design follow-up experiments in similar populations.

6.1. Education experiment

In the experiment described in Angrist et al. (2009), the authors use a CRD to allocate the $N = 1656$ units to the 2^2 treatments. Theorem 1 can directly inform us of the optimal allocation without costs, but there is more structure that we can exploit. For instance, there are potentially two blocks of experimental units or subjects representing female (block 1) and male (block 2) students, with block sizes $M_1 = 948$ and $M_2 = 708$, which can be used to improve their design. Theorem 2 will give us the optimal designs in this case.

Assuming that there is no prior information about the variances of potential outcomes (GPAs after year 1), we assume that the variances are equal within and across blocks (Conditions 3 and 4). Then, optimal allocations under both CRD and RBD, from Theorem 1 and Theorem 2 respectively, are shown in Tables 4 and 5.

Table 4: A-, D- & E-optimal allocations under CRD assuming Condition 1

	Treatment combination			
	00	01	10	11
N= 1656	414	414	414	414

Table 5: A-, D- & E-optimal allocations under RBD assuming Conditions 3 & 4

Block	Block size	Treatment combination			
		00	01	10	11
1	$M_1 = 948$	237	237	237	237
2	$M_2 = 708$	177	177	177	177

Now let us consider a hypothetical situation where the number of units N is not prespecified, but there is a budget constraint that depends on the costs associated with the four treatment combinations in this experiment. The treatment combinations 01 (SFP but not SSP) and 10 (SSP but not SFP), each involve cost associated with one of two programs. Angrist et al. (2009) report about \$5,000 for individual students that were allocated to treatment 10 (SSP). Per unit cost for treatment combination 01(SFP) is not mentioned but if we assume a similar cost as with SSP, then, we can infer that the cost to allocate a student to the treatment combination 11 (SFSP) would be the sum total of the individual costs (\$10,000). The control, representing the treatment combination 00, is possibly the cheapest to allocate units to, because it would involve only administrative cost, which we assume to be \$500. Then under the original allocation (1106, 250, 250, 150) in the actual experiment as shown in Table 1, the cost of the experiment would be approximately \$4.5 million. Assuming this amount to be our budget constraint C, the A-, D- and E-optimal allocations for two different variance vectors obtained from Theorem 3 are shown in Table 6. The first row shows the allocation of the proportions π_j's of the total budget to the four treatment arms, and the second shows the corresponding approximate integer solution for N_j as $\lfloor C\pi_j/C_j \rfloor$.

Table 6: Optimal allocations (π_j and $N_j = \lfloor C\pi_j/C_j \rfloor$) with cost vector (500,5000,5000,10000) and total budget of $C =$ \$4.5 million

Variance vector $(S_1^2, \ldots, S_4^2)$	Type of Optimality		Treatment combination 00	01	10	11
(1,1,1,1)	A	π_j	0.085	0.268	0.268	0.379
		N_j	762	241	241	170
	D	π_j	0.25	0.25	0.25	0.25
		N_j	2250	225	225	112
	E	π_j	0.024	0.244	0.244	0.488
		N_j	219	219	219	219
(1,2,2,2)	A	π_j	0.062	0.275	0.275	0.389
		N_j	553	247	247	174
	D	π_j	0.25	0.25	0.25	0.25
		N_j	2250	225	225	112
	E	π_j	0.012	0.245	0.245	0.494
		N_j	111	222	222	222

6.2. Audit experiment

Libgober (2020) reported an audit study in which the experimental units were 96 lawyers randomly selected from lawyers in California with a certification in criminal law. Each lawyer in the

experiment received an email about a routine 'driving under influence' (DUI) case (a very common criminal matter). The email template suggested that the person sending the email was (i) either white or black (with a racially distinctive name being used to influence perceived race), (ii) either female or male (again cued via the email sender's name), and (iii) either relatively affluent or relatively lower-income description of client's earnings. Thus, this experiment had a 2^3 factorial structure. The response was recorded as a binary outcome taking value 1 if there was a response to the email and 0 otherwise. The experiment was replicated with 96 additional lawyers after a certain period of time. The estimated variances s_j^2 for $j = 1, \ldots, 8$ treatment groups for the individual replicates and their pooled values are shown in Table 7.

Table 7: Estimated variances for different treatment groups

Experiment	000	001	010	011	100	101	110	111
Replicate I	0.15	0.15	0.15	0.20	0.27	0.15	0.27	0.27
Replicate II	0.27	0.24	0.20	0.20	0.20	0.27	0.27	0.15
Pooled	0.21	0.20	0.18	0.20	0.23	0.21	0.27	0.21

If another completely randomized experiment is planned with lawyers selected from a similar pool with a sample size of 192, then based on the pooled estimated variances shown in Table 7, we can apply Theorem 1 to obtain the optimal designs given in Table 8.

Table 8: Optimal allocations for future CRD

Optimality	000	001	010	011	100	101	110	111
A	24	23	22	23	25	24	27	24
D	24	24	24	24	24	24	24	24
E	24	22	20	22	26	24	30	24

Now assume for illustration that (contrary to fact) instead of two replicates, the original experiment was conducted in two blocks, each block representing one type of lawyer (e.g., criminal and divorce), and suppose we want to obtain optimal allocations within each block for a future experiment. Further suppose that the variance estimates in row j of Table 7 represent estimates of the variances $S_{h,j}^2$ in block $j = 1, 2$ and block 2 respectively. Then, we can directly use part (a) of Theorem 2 to derive the A-optimal design. However, neither WBH or BBH appear to hold, and we cannot apply parts (b) and (c) of Theorem 2. Conveniently, we can obtain the D- and E-optimal designs using the greedy search algorithm proposed in Section 4.2.

Table 9: Optimal allocations for future RBD

Optimality	Block	000	001	010	011	100	101	110	111
A	I	11	11	10	12	14	10	14	14
	II	13	13	12	11	11	13	13	10
D	I	11	11	12	13	13	10	12	14
	II	13	13	13	12	11	13	11	10
E	I	10	10	10	12	15	10	16	13
	II	13	12	10	11	12	13	15	10

7. Discussion

In this paper, we consider optimal allocations of a finite population of experimental units to different treatment combinations of a 2^K factorial experiment under the potential outcomes model. Rather than invoking the standard assumption in the mainstream optimal design literature that outcome data comes from a known family of distributions, our work revolves around randomization-based causal inference for the finite-population setting. We find that for 2^K factorial designs with a completely randomized treatment assignment mechanism, D-optimal solutions are always balanced designs, while A- and E-optimal solutions are proportional to finite-population standard deviations and finite-population variances of the treatment groups, respectively. For blocked designs, our solution does not admit a closed form for D- or E-optimality without imposing specific restrictions on the potential outcomes, but the A-optimal allocation is equivalent to finding the A-optimal solution within each block. Convenient integer-constrained programming solutions using a greedy optimization approach to find integer optimal allocation solutions for both complete and block randomization are proposed. Optimal allocations are also derived under cost constraints.

While there is a large literature on model-based optimal designs, to the best of our knowledge, such designs have had very limited development for randomization-based inference for finite populations. The ideas explored and results developed in this paper exploit the connection between finite-population sampling and experimental design. This recondite connection has recently been emphasized, explored, and utilized in various contexts by several researchers in causal inference, as discussed in Mukerjee et al. (2018). This article attempts to further strengthen the bridge between finite-population survey sampling and experimental design by utilizing ideas from proportional and optimal allocation for stratified sampling in the context of optimal designs. While the optimal solutions are derived for a finite-population setting, they are readily applicable to a super-population

setting without making any assumptions about the probability distribution of the outcome variable.

A question that practitioners may ask is, which optimal design should be chosen for a given experiment? The answer would depend on the research goal of the experimenter. As our results have shown, strong assumptions like strict additivity lead to equivalence of A-, D- and E- optimal designs. However, under treatment effect heterogeneity, different criteria will lead to different allocations. Both A- and D-optimality criteria are associated with quality of estimated causal effects - whereas A-optimality minimizes the average variance of estimators, the D-optimality criterion minimizes the volume of the confidence ellipsoid around the parameters. Some researchers (e.g., Jones et al. 2021) have argued that in model-based settings, A-optimal designs exhibit better performance than D-optimal designs when the objective is screening of active effects from inactive ones. On the other hand, when the goal is to draw the most precise inference on the vector of estimated causal effects, D-optimal design may be a better choice. The goal of the E-optimal design is to minimize the maximum variance of all possible normalized linear combinations of estimated treatment effects. Thus the E-optimal design is useful when a large number of linear combinations of factorial effects are of interest. The E-optimal allocation, being a minimax strategy, is likely to provide a more conservative solution to the inference problem, but as shown by some researchers (e.g., Wong 1994) in other contexts, the E-optimal solution may be less sensitive to incorrect prior information or assumptions about potential outcomes in comparison to A- and D-optimal designs. However, more investigation is required along these lines in the randomization-based setting.

The work presented in this paper can be extended in several directions. One limitation of the proposed approach lies is the fact that the correlation among the potential outcomes under different treatment combinations is unidentifiable from the data, forcing us to ignore one term in the covariance matrix of estimated factorial effects while formulating the optimization problem. Basse and Airoldi (2018) proposed a model-based approach to overcome this problem in two-armed experiments, in which information on the correlation among the outcomes is available pre-intervention. Such an idea may be extended to the setting of factorial experiments.

Also, a natural extension of the randomization-based framework of causal inference is the Bayesian framework, in which the potential outcomes are assumed to follow a hierarchical probabilistic model containing hyperparameters with assumed prior distributions. The Bayesian framework proposed in Dasgupta et al. (2015) for drawing both super-population and finite-population

causal inference from 2^K factorial designs can be utilized to obtain Bayesian optimal deigns according to different criteria proposed in literature (e.g., Chaloner and Verdinelli 1995).

Another setting that has gained a lot of attention in recent times is when SUTVA is violated, for example, in the presence of interference between units. Extending the proposed results to such settings is a challenging, yet rewarding problem.

Finally, in certain situations, it is possible that instead of the traditional factorial effects defined by (3), the experimenter is interested in other contrasts of the treatment means or more general factorial effects. One such natural choice of contrast is one that compares the outcome of the control group with the average of all other groups that have at least one treatment. Optimal allocations under such a reformulated optimization problem would be an interesting problem to study.

Acknowledgement: This research was partially supported by National Science Foundation grant SES 2217522. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88:265–296.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–63.

Atkinson, A., Donev, A., and Tobias, R. (2007). *Optimum Experimental Designs, With SAS*. Oxford: Oxford University Press.

Basse, G. and Airoldi, E. (2018). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, 105(4).

Blackwell, M., Pashley, N. E., and Valentino, D. (2022). Batch adaptive designs to improve efficiency in social science experiments. `https://www.mattblackwell.org/files/papers/batch_adaptive.pdf`.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Branson, Z., Dasgupta, T., and Rubin, D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education High School Study. *The Annals of Applied Statistics*, 10(4):1958 – 1976.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3).

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 30:586–602.

Cochran, W. G. (1977a). *Sampling Techniques*. New York: John Wiley.

Cochran, W. G. (1977b). *Sampling Techniques, 3rd Edition.* John Wiley, New York.

Dai, J., Gradu, P., and Harshaw, C. (2023). Clip-ogd: An experimental design for adaptive neyman allocation in sequential experiments. *arXiv preprint arXiv:2305.17187*.

Dasgupta, T., Pillai, N., and Rubin, D. (2015). Causal inference for 2^K factorial designs by using potential outcomes. *Journal of the Royal Statisical Society (Series B)*, 77(4):727–753.

Ding, P., Li, X., and Miratrix, L. (2017). Bridging finite and super population causal inference. *Journal of Causal Inference*, 5:20160027.

Egami, N. and Imai, K. (2019). Causal interaction in factorial experiments: Application to conjoint analysis. *Journal of the American Statistical Association*, 114:526–540.

Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.

Freedman, D. A. (2006). Statistical models for causation: What inferential leverage do they provide? *Evaluation Review*, 30:691–713.

Freedman, D. A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40:180–193.

Friedrich, U., Münnich, R., de Vries, S., and Wagner, M. (2015). Fast integer-valued algorithms for optimal allocations under constraints in stratified sampling. *Computational Statistics & Data Analysis*, 92:1–12.

Hahn, J., Hirano, K., and Karlan, D. (2011). Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29(1):96–108.

Jones, B., Allen-Moyer, K., and Goos, P. (2021). A-optimal versus d-optimal design of screening experiments. *Journal of Quality Technology*, 53(4):369–382.

Khan, M. G. M. (1995). *Mathematical programming in sampling*. PhD thesis, Aligarh Muslim University.

Li, X., Ding, P., and Rubin, D. B. (2020). Rerandomization in 2^K factorial experiments. *The Annals of Statistics*, 48(1):43 – 63.

Libgober, B. (2020). Getting a lawyer while black: A field experiment. *Lewis & Clark Law Review.*, 24(1):53–108.

Lu, J. (2016). On randomization-based and regression-based inferences for 2^K factorial designs. *Statistics and Probability Letters*, 112:72–78.

Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2):1263–1282.

Mukerjee, R., Dasgupta, T., and Rubin, D. B. (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *Journal of the American Statistical Association*, 113:868–881.

Nemhauser, G. and Wolsey, L. (1988). The scope of integer and combinatorial optimization. *Integer and combinatorial optimization*, pages 1–26.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):558–625.

Olsen, R., Orr, L., Bell, S., and Stuart, E. (2013). External validity in policy evaluations that choose sites purposively. *Journal of Policy Analysis and Management*, 32:107–121.

Pashley, N. E. and Bind, M.-A. C. (2022). Causal inference for multiple treatments using fractional factorial designs. *Canadian Journal of Statistics*.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593.

Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.

Shi, L., Wang, J., and Ding, P. (2023). Forward screening and post-screening inference in factorial designs. *arXiv preprint arXiv:2301.12045*.

Sofi, N., Ahmad, A., Maqbool, D. S., and Ahmad, B. (2020). A branch and bound approach to optimal allocation in stratified sampling.

Wong, W. K. (1994). Comparing robust properties of a-, d- e- and g-optimal designs. *Computational Statistics and Data Analysis*, 18:441–448.

Yang, J. and Mandal, A. (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics*, 44:2264–2277.

Yang, J., Mandal, A., and Majumdar, D. (2012). Optimal designs for two-level factorial experiments with binary response. *Statistica Sinica*, 22:885–907.

Yang, J., Mandal, A., and Majumdar, D. (2016). Optimal designs for 2^k factorial experiments with binary response. *Statistica Sinica*, 26:385–411.

Yates, F. (1937). The design and analysis of factorial experiments. Imperial Bureau of Soil Sciences - Technical communication 35.

Zhao, A. and Ding, P. (2021). Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties. *Biometrika*, 109(3):799–815.

Supplementary Material

for

"Optimal allocation of sample size for randomization-based inference from 2^K factorial designs"

Supplementary Material A Proof of results

Proof of Theorem 1:

We first state and prove a lemma about the matrix $\widetilde{\boldsymbol{V}}$.

Lemma 1. The matrix $\widetilde{\boldsymbol{V}}$ has J non-zero eigenvalues $J(S_1^2/N_1), \ldots, J(S_J^2/N_J)$.

Proof: Note that $\widetilde{\boldsymbol{V}}$ can be written as $\left(\sqrt{J}^{-1}\boldsymbol{L}^{\mathrm{T}}\right)(J\boldsymbol{A})\left(\sqrt{J}^{-1}\boldsymbol{L}\right)$. Because the rows of $\sqrt{(}J)^{-1}\boldsymbol{L}$ form an orthonormal basis of vectors in $\mathbb{R}^J$, the above expression is a spectral decomposition of $\widetilde{\boldsymbol{V}}$. Thus, the eigenvalues of $\widetilde{\boldsymbol{V}}$ are its diagonal elements $J(S_1^2/N_1), \ldots, J(S_J^2/N_J)$. □

Now we prove the three parts of Theorem 1.

Part (a): Because the trace of a matrix is the sum of its eigenvalues, from Lemma 1, it follows that $\mathrm{tr}(\widetilde{\boldsymbol{V}}) = J\sum_{j=1}^{J}(S_j^2/N_j)$. The problem of minimizing $\mathrm{tr}(\widetilde{\boldsymbol{V}})$ can be expressed as:

$$\text{minimize} \sum_{j=1}^{J}(S_j^2/N_j) \;\; \text{subject to} \sum_{j=1}^{J} N_j = N$$

This constrained optimization problem can be obtained using the method of Lagrange multipliers, by solving

$$\min \left[\sum_{j=1}^{J} S_j^2/N_j + \lambda\left(\sum_{j=1}^{J} N_j - N\right)\right],$$

where λ is a Lagrangian multiplier. Taking partial derivative of the objective function with respect to N_j and setting it to zero, we get $-S_j^2/N_j^2 + \lambda = 0$, which implies $N_j = S_j/\sqrt{\lambda}$. Solving for the constraint $\sum_{j=1}^{J} N_j = N$, we get,

$$\sqrt{\lambda} = \frac{\sum_{j=1}^{J} S_j}{N} \Rightarrow N_j = S_j\left(\frac{N}{\sum_{j=1}^{J} S_j}\right) \Rightarrow p_j \propto S_j. \quad \square$$

Part (b): Because the determinant of a matrix is the product of the eigenvalues, from Lemma 1, we have $|\widetilde{\boldsymbol{V}}| = J^J \prod_{j=1}^J (S_j^2/N_j)$. The problem of minimizing the determinant can thus be equivalently expressed as

$$\text{minimize} \sum_{j=1}^{J} \log(S_j^2/N_j) \;\; \text{subject to} \;\; \sum_{j=1}^{J} N_j = N.$$

Taking partial derivative of the objective function with respect to N_j and setting it to zero, we get $-(S_j^2/N_j^2)(N_j/S_j^2) + \lambda = -1/N_j + \lambda = 0$. It is straightforward to see that the optimal solution is $N_j = N/J$ or equivalently $p_j = 1/J$ for $j = 1, \ldots, J$. □

Part (c): Let $(N_1^{\mathcal{D}}, \ldots, N_J^{\mathcal{D}})$ denote the allocation vector of a design $\mathcal{D}$. Also, let the J eigenvalues of $\widetilde{\boldsymbol{V}}$ for design $\mathcal{D}$ be $\nu_1^{\mathcal{D}}, \ldots, \nu_J^{\mathcal{D}}$, and $\nu_{(1)}^{\mathcal{D}} \leq \ldots \leq \nu_{(J)}^{\mathcal{D}}$ denote the ordered eigenvalues. We will show that the design $\mathcal{D}^*$ in which $N_j^{\mathcal{D}^*} = (NS_j^2)/\left(\sum_{j=1}^J S_j^2\right)$ for $j = 1, \ldots, J$ is the E-optimal design. From Lemma 1, design $\mathcal{D}^*$ can be characterized as a design in which all eigenvalues are equal to $J\left(\sum_{j=1}^J S_j^2\right)/N = \nu^{\mathcal{D}^*}$. Recall the definition of E-optimality of minimizing the maximum eigenvalue of the design matrix. Thus, it suffices to show that any design in which all eigenvalues of $\widetilde{\boldsymbol{V}}$ are not equal cannot be E-optimal because the solution $\mathcal{D}^*$ with all eigenvalues equal to $J\left(\sum_{j=1}^J S_j^2\right)/N$ is the only solution with all eigenvalues equal in the feasible space of the optimization problem.

We will proceed by proof by contradiction. Assume that a design $\widetilde{\mathcal{D}}$ for which all eigenvalues of $\widetilde{\boldsymbol{V}}$ are not equal is E-optimal, i.e.,

$$\nu_{(J)}^{\widetilde{\mathcal{D}}} \leq \; \nu_{(J)}^{\mathcal{D}}, \tag{11}$$

for any design $\mathcal{D}$.

Let $\mathcal{M}$ denote the set of $m \geq 1$ equal maximum eigenvalues of $\widetilde{\boldsymbol{V}}$ for design $\widetilde{\mathcal{D}}$ ($m = 1$ indicate a unique maximum and because not all eigenvalues are equal we must have $m < J$). Then for any $j_1 \in \mathcal{M}$ and any $j_2 \notin \mathcal{M}$, $\nu_{j_1}^{\widetilde{\mathcal{D}}} > \nu_{j_2}^{\widetilde{\mathcal{D}}}$. Construct a new design $\mathcal{D}'$ by perturbing only the allocations

for *all* treatments $j_1 \in \mathcal{M}$ and *one specific* $j_2 \notin \mathcal{M}$ as follows:

$$\begin{aligned} N_{j_1}^{\mathcal{D}'} &= S_{j_1}^2 \frac{\sum_{j\in\mathcal{M}} N_j^{\widetilde{\mathcal{D}}} + N_{j_2}^{\widetilde{\mathcal{D}}}}{\sum_{j\in\mathcal{M}} S_j^2 + S_{j_2}^2} \text{ for all } j_1 \in \mathcal{M} \\ N_{j_2}^{\mathcal{D}'} &= S_{j_2}^2 \frac{\sum_{j\in\mathcal{M}} N_j^{\widetilde{\mathcal{D}}} + N_{j_2}^{\widetilde{\mathcal{D}}}}{\sum_{j\in\mathcal{M}} S_j^2 + S_{j_2}^2} \\ N_j^{\mathcal{D}'} &= N_j^{\widetilde{\mathcal{D}}}, \text{ for } j \notin \mathcal{M} \cup \{j_2\}. \quad \square \end{aligned}$$

Using Lemma 1, after a little algebra, it follows that $\nu_{j_1}^{\widetilde{\mathcal{D}}} > \nu_{j_1}^{\mathcal{D}'} = \nu_{j_2}^{\mathcal{D}'} > \nu_{j2}^{\widetilde{\mathcal{D}}}$ for all $j_1 \in \mathcal{M}$. Also, for all $j \notin \mathcal{M} \cup \{j_2\}$, $\nu_{j_1}^{\widetilde{\mathcal{D}}} > \nu_j^{\widetilde{\mathcal{D}}} = \nu_j^{\mathcal{D}'}$. Consequently $\nu_{(J)}^{\widetilde{\mathcal{D}}} > \nu_{(J)}^{\mathcal{D}'}$, and contradicts (11).

Proof of Theorem 2:

We need the following lemma regarding the eigenvalues of the covariance matrix $\widetilde{\boldsymbol{V}}_{\text{blk}}$ defined in (9) under a blocked design:

Lemma 2. The matrix $\widetilde{\boldsymbol{V}}_{\text{blk}}$ has J non-zero eigenvalues $JS_{\text{blk},1}^2, \ldots, JS_{\text{blk},J}^2$.

Proof: Along the same lines as Lemma 1, note that $\widetilde{\boldsymbol{V}}_{\text{blk}}$ can written as $\left(\sqrt{J}^{-1}\boldsymbol{L}^{\mathrm{T}}\right)(J\boldsymbol{A}_{\text{blk}})\left(\sqrt{J}^{-1}\boldsymbol{L}\right)$. Because the rows of $\sqrt{(}J)^{-1}\boldsymbol{L}$ forms an orthonormal basis of vectors in $\mathbb{R}^J$, the above expression is a spectral decomposition of $\widetilde{\boldsymbol{V}}_{\text{blk}}$. Thus, the eigenvalues of $\widetilde{\boldsymbol{V}}_{\text{blk}}$ are its diagonal elements $JS_{\text{blk},1}^2, \ldots, JS_{\text{blk},J}^2$. $\square$

Now we prove the three parts of the main theorem.

Part (a): We can proceed very similarly to the proof of Theorem 1(a). Because the trace of a matrix is the sum of its eigenvalues, using Lemma 2, $\text{tr}(\widetilde{\boldsymbol{V}}_{\text{blk}}) = J\sum_{j=1}^J S_{\text{blk},j}^2$. The problem of minimizing $\text{tr}(\widetilde{\boldsymbol{V}}_{\text{blk}})$ can be expressed as

$$\text{minimize} \sum_{j=1}^J S_{\text{blk},j}^2 \text{ subject to} \sum_{j=1}^J M_{h,j} = M_h \text{ for } h = 1, \ldots, H.$$

We can again solve this using the method of Lagrange multipliers,

$$\min\left[\sum_{j=1}^J S_{\text{blk},j}^2 + \sum_{h=1}^H \lambda_h \left(\sum_{j=1}^J M_{h,j} - M_h\right)\right] = \min\left[\sum_{j=1}^J \sum_{h=1}^H \frac{M_h^2}{N^2}\frac{S_{h,j}^2}{M_{h,j}} + \sum_{h=1}^H \lambda_h \left(\sum_{j=1}^J M_{h,j} - M_h\right)\right],$$

where λ_h are Lagrangian multipliers. Taking partial derivative of the objective function with respect to $M_{h,j}$ and setting it to zero, we get $-\frac{M_h^2}{N^2}\frac{S_{h,j}^2}{M_{h,j}^2} + \lambda_h = 0$, which implies

$$M_{h,j} = \frac{M_h}{N}\frac{S_{h,j}}{\sqrt{\lambda_h}}.$$

Solving for the constraint $\sum_{j=1}^{J} M_{h,j} = M_h$, we get,

$$\sqrt{\lambda_h} = \frac{\sum_{j=1}^{J} S_{h,j}}{N} \Rightarrow M_{h,j} = M_h \frac{S_{h,j}}{\sum_{j=1}^{J} S_{h,j}} \Rightarrow p_{h,j} = \frac{S_{h,j}}{\sum_{j=1}^{J} S_{h,j}}. \quad \square$$

Part (b): Because the determinant of a matrix is the product of the eigenvalues, from Lemma 2, we have

$$|\widetilde{V}| = J^J \prod_{j=1}^{J} S_{\text{blk},j}^2 \quad .$$

The problem of minimizing the determinant can thus be equivalently expressed as

$$\text{minimize} \sum_{j=1}^{J} \log(S_{\text{blk},j}^2) \;\; \text{subject to} \;\; \sum_{j=1}^{J} M_{h,j} = M_h \quad .$$

To prove the special cases, we use the Lagrangian multiplier based optimization approach as in (a). So, we need to solve

$$\begin{aligned} &\min \left[\sum_{j=1}^{J} \log(S_{\text{blk},j}^2) + \sum_{h=1}^{H} \lambda_h \left(\sum_{j=1}^{J} M_{h,j} - M_h \right) \right] \\ &= \min \left[\sum_{j=1}^{J} \log \left(\sum_{h=1}^{H} \frac{M_h^2}{N^2} \frac{S_{h,j}^2}{M_{h,j}} \right) + \sum_{h=1}^{H} \lambda_h \left(\sum_{j=1}^{J} M_{h,j} - M_h \right) \right], \end{aligned}$$

where λ_h are Lagrangian multipliers.

Taking the derivative with respect to $M_{h,j}$ and setting equal to 0, we have

$$0 = -M_h^2 \frac{S_{h,j}^2}{M_{h,j}^2} \left(\sum_{k=1}^{H} M_k^2 \frac{S_{k,j}^2}{M_{k,j}} \right)^{-1} + \lambda_h \quad . \tag{12}$$

Under special case (i), WBH: variances of potential outcomes are the same within

each block (such that $S^2_{h,j} = S^2_{h,\cdot}$ for all $j = 1, \ldots, J$).

Substitution of $S^2_{h,j} = S^2_{h,\cdot}$ into (12) yields

$$0 = -M_h^2 \frac{S^2_{h,\cdot}}{M^2_{h,j}} \left(\sum_{k=1}^{H} M_k^2 \frac{S^2_{k,j}}{M_{k,j}} \right)^{-1} + \lambda_h.$$

After a little algebra, we get

$$M_{h,j} = M_h \frac{S_{h,\cdot}}{\sqrt{\lambda_h}} c_j \quad \text{where } c_j = \sqrt{\left(\sum_{k=1}^{H} M_k^2 \frac{S^2_{k,\cdot}}{M_{k,j}} \right)^{-1}} \tag{13}$$

Now, summing over j in (13) and applying the second constraint $\sum_{j=1}^{J} M_{h,j} = M_h$ we have:

$$\begin{aligned} M_h &= \sum_{j=1}^{J} M_{h,j} = \sum_{j=1}^{J} M_h \frac{S_{h,\cdot}}{\sqrt{\lambda_h}} c_j = \frac{S_{h,\cdot} M_h}{\sqrt{\lambda_h}} \sum_{j=1}^{J} c_j \\ \Rightarrow \quad \frac{S_{h,\cdot}}{\sqrt{\lambda_h}} &= \left(\sum_{j=1}^{J} c_j \right)^{-1} \end{aligned} \tag{14}$$

Thus, substituting (14) in (13),

$$M_{h,j} = M_h \frac{c_j}{\sum_{j=1}^{J} c_j} = M_h \delta_j \text{ where } \delta_j = \frac{c_j}{\sum_{j=1}^{J} c_j} \ .$$

Substituting this back into definition of c_j in (13) gives us,

$$\begin{aligned} c_j &= \sqrt{\left(\sum_{k=1}^{H} M_k^2 \frac{S^2_{k,\cdot}}{M_k \delta_j} \right)^{-1}} = \sqrt{\left(\sum_{k=1}^{H} M_k \frac{S^2_{k,\cdot}}{\delta_j} \right)^{-1}} = \sqrt{\delta_j} \sqrt{\left(\sum_{k=1}^{H} M_k S^2_{k,\cdot} \right)^{-1}} \\ &= \sqrt{\delta_j}\alpha \quad \text{where} \quad \alpha = \sqrt{\left(\sum_{k=1}^{H} M_k S^2_{k,\cdot} \right)^{-1}} \quad \text{is a constant} \\ &= \frac{\sqrt{c_j}}{\sqrt{\sum_{j=1}^{J} c_j}} \alpha \\ &\implies c_j = \frac{\alpha^2}{\sum_{j=1}^{J} c_j} = \beta, \quad \text{a constant free of } j. \end{aligned}$$

Thus, $\delta_j = c_j / \sum_j c_j = 1/J$ and $M_{h,j} = M_h/J$ and $p_{h,j} = 1/J$. □

Under special case (ii), BBH: variances are the same across blocks for each treatment (such that $S^2_{h,j} = S^2_{\cdot,j}$ for all $h = 1, \ldots, H$).

Substituting $S^2_{h,j} = S^2_{\cdot,j}$ in (12), we get,

$$
\begin{aligned}
0 &= -\frac{M_h^2}{M_{h,j}^2}\left(\sum_{k=1}^{H}\frac{M_k^2}{M_{k,j}}\right)^{-1} + \lambda_h \\
\implies M_{h,j} &= \frac{M_h}{\sqrt{\lambda_h}}\sqrt{\left(\sum_{k=1}^{H}\frac{M_k^2}{M_{k,j}}\right)^{-1}} = \frac{M_h}{\sqrt{\lambda_h}}c_j \quad \text{where } c_j = \sqrt{\left(\sum_{k=1}^{H}\frac{M_k^2}{M_{k,j}}\right)^{-1}}
\end{aligned} \tag{15}
$$

Now, summing over j and applying the second constraint $\sum_{j=1}^{J} M_{h,j}$ yields:

$$
\begin{aligned}
M_h &= \sum_{j=1}^{J} M_{h,j} = \sum_{j=1}^{J}\frac{M_h}{\sqrt{\lambda_h}}c_j = \frac{M_h}{\sqrt{\lambda_h}}\sum_{j=1}^{J} c_j \\
\implies \frac{M_h}{\sqrt{\lambda_h}} &= \frac{M_h}{\sum_{j=1}^{J} c_j}.
\end{aligned}
$$

Proceeding similarly as in the proof of the previous part, we can show that $\delta_j = c_j/(\sum_{j=1}^{J} c_j) = 1/J$ for all $j = 1, \ldots, J$ and hence $M_{h,j} = M_h/J$ and $p_{h,j} = 1/J$. □

Part (c): Recall from Section 4, $S^2_{blk,j} = \sum_{h=1}^{H}(M_h/N)^2(S^2_{h,j}/M_{h,j})$.

We need the following lemmas to prove the theorem.

Lemma 3. For any $r > 0$ and $J > 1$, $\delta_j > -r$ such that $\sum_{j=1}^{J}\delta_j = 0$ and $\delta_j \neq 0\ \forall j = 1, 2, \ldots, J$,

$$\sum_j \left(\frac{1}{r+\delta_j} - \frac{1}{r}\right) > 0$$

Proof. Let $j^* = \text{argmin}_{\delta_j > 0}\,\delta_j$. Then,

- when $\delta_j > 0$ and $j \neq j^*$, $(r+\delta_j^*) < (r+\delta_j) \implies 1/(r+\delta_j^*) > 1/(r+\delta_j) \implies -\delta_j/(r+\delta_j^*) < -\delta_j/(r+\delta_j)$
- when $\delta_j < 0$, $(r+\delta_j^*) > (r+\delta_j) \implies 1/(r+\delta_j^*) < 1/(r+\delta_j) \implies -\delta_j/(r+\delta_j^*) < -\delta_j/(r+\delta_j)$
- when $\delta_j = 0$, $-\delta_j/(r+\delta_j^*) = -\delta_j/(r+\delta_j)$
- when $j = j^*$, $-\delta_j/(r+\delta_j^*) = -\delta_j/(r+\delta_j)$

Thus, it holds that,

$$\sum_j \frac{-\delta_j}{r+\delta_j} > \sum_j \frac{-\delta_j}{r+\delta_{j^*}}$$

Then,

$$\sum_j \left(\frac{1}{r+\delta_j} - \frac{1}{r}\right) = \sum_j \frac{-\delta_j}{r(r+\delta_j)} > \sum_j \frac{-\delta_j}{r(r+\delta_{j^*})} = \frac{-\sum_j \delta_j}{r(r+\max_j \delta_{j^*})} = 0$$

□

Lemma 4. Given integers M_h for $h = 1, \ldots, H$, let $M_{h,j}$ denote any allocation of M_h into $J > 1$ groups such that $\sum_j M_{h,j} = M_h$. Let $a_h > 0$, for $h \in \{1, \ldots, H\}$, be fixed. Then,

$$\max_j \sum_{h=1}^{H} \frac{a_h}{M_{h,j}} \geq \sum_h \frac{a_h}{(\frac{M_h}{J})}.$$

That is, the allocation that minimizes $\max_j \sum_{h=1}^{H} a_h/M_{h,j}$ is $M_{h,j} = M_h/J$.

Proof. For each $h = 1, \ldots, H$, using an argument similar to the one in the proof of Theorem 1, we can write for $h = 1, \ldots, H$,

$$\max_j \frac{a_h}{M_{h,j}} \geq \frac{a_h}{\frac{M_h}{J}}$$

Suppose there exists an allocation $\tilde{M}_{h,j} = M_h/J + \delta_j^h$ that maximizes $\max_j \sum_{h=1}^{H} a_h/M_{h,j}$, where for each h, $\exists j$ such that $\delta_j^h \neq 0$. Then, $\sum_{j=1}^{J} \tilde{M}_{h,j} = M_h \implies \sum_{j=1}^{J} \delta_j^h = 0 \; \forall h = 1, \ldots, H$.

Further, by our assumption on $\tilde{M}_{h,j}$,

$$\begin{aligned}
\max_j \sum_{h=1}^{H} \frac{a_h}{\tilde{M}_{h,j}} \leq \sum_{h=1}^{H} \frac{a_h}{\frac{M_h}{J}} &\iff \max_j \sum_{h=1}^{H} \left(\frac{a_h}{\tilde{M}_{h,j}} - \frac{a_h}{\frac{M_h}{J}} \right) \leq 0 \iff \max_j \sum_{h=1}^{H} \left(\frac{a_h}{(\frac{M_h}{J} + \delta_j^h)} - \frac{a_h}{\frac{M_h}{J}} \right) \leq 0 \\
&\iff \max_j \sum_{h=1}^{H} \epsilon_j^h \leq 0 \text{ (where } \epsilon_j^h = \frac{a_h}{(\frac{M_h}{J} + \delta_j^h)} - \frac{a_h}{\frac{M_h}{J}}) \\
&\iff \sum_{h=1}^{H} \epsilon_j^h \leq 0 \quad \forall j \\
&\implies \sum_{j=1}^{J} \sum_{h=1}^{H} \epsilon_j^h \leq 0 \\
&\iff \sum_{h=1}^{H} a_h \sum_{j=1}^{J} \left(\frac{1}{(\frac{M_h}{J} + \delta_j^h)} - \frac{1}{\frac{M_h}{J}} \right) \leq 0
\end{aligned}$$

which is a contradiction by taking $r = M_h/J$ and $\delta_j = \delta_j^h$ for each h in Lemma 3.

Thus, we have that,

$$\max_j \sum_h \frac{a_h}{M_{h,j}} \geq \sum_h \frac{a_h}{\frac{M_h}{J}}$$

□

Under special case, WBH: variances of potential outcomes are the same within each block (such that $S^2_{h,j} = S^2_{h,\cdot}$ for all $j = 1, \ldots, J$).

We have, $S^2_{blk,j} = \sum_{h=1}^{H} (M_h/N)^2 (S^2_{h,\cdot}/M_{h,j})$. We can rewrite this equation as $S^2_{blk,j} = \sum_{h=1}^{H} a_h/M_{h,j}$ where $a_h = (M_h/N)^2 S^2_{h,\cdot}$ does not depend on j.

By Lemma 4, we get $\max_j \sum_{h=1}^{H} a_h/M_{h,j} \geq J \sum_h a_h/M_h$.

We immediately see that for $M^*_{h,j} = M_h/J$, the left hand side of the inequality attains the lower bound, which is minimax. Hence, $p^*_{h,j} = 1/J$ for each treatment j within block h, a balanced design within each block. □

Supplementary Material B Conditions for Greedy Algorithm from Friedrich et al. (2015)

Theorem 3.2 of Friedrich et al. (2015) have the following conditions to be satisfied for global convergence of the greedy algorithm to the true optimum. We restate the theorem here for reference.

Theorem 3.2 (Friedrich et al. (2015)). The globally optimal integer solution of the problem

$$\min_{(N_1,\ldots,N_J)} \left\{ f(N_1,\ldots,N_J) \;\middle|\; N_j \geq 0 \;\forall j,\; \sum_{j\in A} N_j \leq \varphi(A) \;\forall A \subset E \right\},$$

is found by a Greedy algorithm if

1. E is a finite set,
2. $\varphi : 2^E \to \mathbb{Z}_+$ is submodular, monotone and satisfies $\varphi(\phi) = 0$,
3. $f : \mathbb{R}_+^E \to \mathbb{R}$ is separable and convex with continuous components.

In the case of a CRD, we can identify the following in the theorem above:

- $E = \{1, 2, ..J\}$ (finite)
- $A \in \{\{\emptyset\}, \{1\}, ..., \{J\}, \{1,2\}, .., \{1, ..., J\}\} = 2^E$
- $\varphi(A) = \min(\sum_{j\in A}(u_j - l_j), N - \sum_{j\in E} l_j)$
- $f(N_1, \ldots, N_J)$ defined as in Section 3, $\sum_j S_j^2/N_j$ for A-optimality and $\sum_j \log(S_j^2/N_j)$ for D-optimality respectively (separable and convex) and each continuous in their individual components, i.e., $f_j(N_j)$ are continuous in N_j ($1/x$ and $\log(1/x)$ is continuous in x for $x \in \mathbb{R}_+^E$).

Condition 1 is already satisfied as noted above. Condition 3 is straightforward since all our real-valued objective functions are convex and finite-dimensional and hence separable. Thus, it suffices to show that Condition 2 above in Theorem 3.2 is satisfied in our case for the submodular set function $\varphi(.)$. Again, we give the definition of a submodular function as in Friedrich et al. (2015).

Definition (Submodular function)**.** $\varphi : 2^E \to \mathcal{Z}_+$ is submodular if

$$\varphi(X \cap Y) + \varphi(X \cup Y) \leq \varphi(X) + \varphi(Y),\; \forall X, Y \subset E$$

Thus, in the case of a CRD, take A to be defined as above, then $g(A) = \sum_{j\in A}(u_j - l_j)$, $h(A) = N - \sum_{j\in A} l_j$ as corresponding set functions for the following argument. Linear set functions are always submodular as can be quickly shown: $g(X \cup Y) = \sum_{j\in X}(u_j - l_j) + \sum_{j\in Y}(u_j - l_j) -$

$\sum_{j\in X\cap Y}(u_j - l_j) = g(X) + g(Y) - g(X \cap Y)$. Similarly for h, $h(X \cup Y) = (N - \sum_{j\in X} u_j) + (N - \sum_{j\in Y} u_j) - (N - \sum_{j\in X\cap Y} u_j) = h(X) + h(Y) - h(X \cap Y)$. Submodularity of our $\phi(.)$ function follows directly from Friedrich et al. (2015) as $\min(g, h)$ is submodular if g, h are submodular and $g - h$ is monotone. g is submodular and monotone (defined as $\forall T, S \subset E$, s.t. $T \subset S \Rightarrow f(T) \leq f(S)$) because $u_j - l_j > 0$, $\forall j \in J$. And, h is submodular. Finally, $(g-h)(A) = \sum_{j\in A} u_j - N$ is monotone since $g - h$ is linear in A.

Supplementary Material C Empirical evidence of greedy algorithm for D-optimality in RBD

Following from Section 4, we show the performance of the greedy algorithms for finding D-optimal solutions empirically.

Five different settings of 2^2 factorial designs in two blocks, each corresponding to a specific type of potential outcome matrix, are considered. Each setting is defined by the block sizes M_1 and M_2, and the 4×2 matrix of variances $S^2_{h,j}$ as shown in columns 3 and 4 of Table 10, respectively.

The first setting considers blocks of equal sizes with potential outcomes satisfying Condition 2 (strict additivity), leading to a D-optimal design that is balanced within each block as per Remark 4. The second setting considers equal block sizes with potential outcomes satisfying Condition 3 (WBH). The third setting considers unequal block sizes with potential outcomes satisfying Condition 4 (BBH) but not Condition 3 (WBH). Note that, in the above cases, the exact solution as given by Theorem 2 is indeed an integer solution, due to the choice of M_h and J. The fourth setting is similar to the third case above, but is one where Theorem 2 provides an exact solution that is not an integer solution. An exhaustive search leads to identification of six different allocations, each of which is optimal. In this case, the greedy algorithm identifies one of these solutions. The fifth setting satisfies neither Condition 3 (WBH) nor Condition 4 (BBH), and consequently Theorem 2 cannot provide an exact D-optimal solution. However, the greedy algorithm identifies the true optimal integer allocation (where truth is identified through exhaustive search).

S.No.	Case	Block Size (M_h)	Variances ($S^2_{h,j}$)	Exhaustive search E-optimal solution	Greedy solution
1.	Equal blocks with equal variances	$\begin{bmatrix}40\\40\end{bmatrix}$	$\begin{bmatrix}1&1&1&1\\1&1&1&1\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$
2.	Equal blocks with equal variances for all treatments within block	$\begin{bmatrix}40\\40\end{bmatrix}$	$\begin{bmatrix}4&4&4&4\\1&1&1&1\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\10&10&10&10\end{bmatrix}$
3.	Unequal blocks with equal variances across blocks for each treatment	$\begin{bmatrix}40\\20\end{bmatrix}$	$\begin{bmatrix}1&2&3&4\\1&2&3&4\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\5&5&5&5\end{bmatrix}$	$\begin{bmatrix}10&10&10&10\\5&5&5&5\end{bmatrix}$
4.	Unequal blocks with equal variances but exact solution is non-integer	$\begin{bmatrix}40\\20\end{bmatrix}$	$\begin{bmatrix}1&2&3&5\\1&2&3&5\end{bmatrix}$	$\left[\begin{bmatrix}10&10&10&10\\8&8&7&7\end{bmatrix}\begin{bmatrix}10&10&10&10\\8&7&8&7\end{bmatrix}\begin{bmatrix}10&10&10&10\\7&8&8&7\end{bmatrix}\begin{bmatrix}10&10&10&10\\8&7&7&8\end{bmatrix}\begin{bmatrix}10&10&10&10\\7&8&7&8\end{bmatrix}\begin{bmatrix}10&10&10&10\\7&7&8&8\end{bmatrix}\right]$	$\begin{bmatrix}10&10&10&10\\8&7&8&7\end{bmatrix}$
5.	Equal blocks with unequal variances	$\begin{bmatrix}40\\20\end{bmatrix}$	$\begin{bmatrix}1&2&3&4\\4&3&2&1\end{bmatrix}$	$\begin{bmatrix}7&10&11&12\\7&6&4&3\end{bmatrix}$	$\begin{bmatrix}7&10&11&12\\7&6&4&3\end{bmatrix}$

Table 10: Summary of Greedy algorithm solutions for D-optimality for $H = 2, K = 2$

Supplementary Material D Greedy Algorithm for RBD

Using the methods in Section 3.2, we can use the same algorithm in Figure 1 to obtain the optimal integer solutions for the A-optimality case under block design by taking $f_{h,j}(M_{h,j}) = (M_h^2/N^2)(S^2_{h,j}/M_{h,j})$. The exact algorithm taking the structure of the blocks into account is given in Figure 3.

For D- and E-optimality, we extend the greedy idea from the previous parts and provide the algorithm in Figure 3 and 4. For D-optimality, take $f_{h,j}(M_{h,j}) = \log(\sum_{h=1}^{H}(M_h^2/N^2)(S^2_{h,j}/M_{h,j}))$. For E-optimality, take $f_{h,j}(M_{h,j}) = (M_h^2/N^2)(S^2_{h,j}/M_{h,j})$.

Figure 3: Greedy algorithm for A- & D-optimality

1. Set $I \leftarrow \begin{bmatrix} 1 & ... & J \\ & ... & \\ 1 & ... & J \end{bmatrix}_{H \times J}$
2. $I_1 \leftarrow vec(I)$
3. Initialize $M_{h,j} \leftarrow l_{h,j}$ for $j = 1, \ldots, J, h = 1, ..., H$.
4. While $(\sum_{h=1}^{H} \sum_{j=1}^{J} M_{h,j} \neq N \ \& \ I_1 \neq \phi)$ do
   - for each $(h \in \{1, ..., H\})$
     - for $(j \in I[h,])$, $\delta_{h,j} \leftarrow f_{h,j}(M_{h,j} + 1) - f_{h,j}(M_{h,j})$
   - For A-optimality:
     - for each $(h \in \{1, ..., H\})$
       (i) choose $j^* \leftarrow \min \operatorname{argmin}_{j \in I[h,]} \delta_{h,j}$.
       (ii) if $(M_{h,j^*} < u_{h,j^*})$, then $M_{h,j^*} \leftarrow M_{h,j^*} + 1$,
       else $I[h, j^*] \leftarrow NA$
   - For D-optimality:
     - choose $(h^*, j^*) \leftarrow \operatorname{argmin}_{(h,j) \in I} \delta_{h,j}$. If non-unique, choose minimum index.
     - if $(\sum_j M_{h^*,j} < M_h^*)$
       (i) if $(M_{h^*,j^*} < u_{h,j^*}$ and $I[h^*, j^*] \neq NA)$, then $M_{h^*,j^*} \leftarrow M_{*h,j^*} + 1$, else $I[h^*, j^*] \leftarrow NA$
       else $I[h^*,] \leftarrow NA$
   - $I_1 \leftarrow vec(I[!is.na(I)])$
5. Return optimal $(M_{1,1}, \ldots, M_{H,J})$.

Figure 4: Greedy algorithm for E-optimality in the Block Case

1. Set $I \leftarrow \begin{bmatrix} 1 & ... & J \\ & ... & \\ 1 & ... & J \end{bmatrix}_{H \times J}$
2. $I_1 \leftarrow vec(I)$
3. Initialize $M_{h,j} \leftarrow l_{h,j}$ for $j = 1, \ldots, J, h = 1, ..., H$.
4. While $(\sum_{h=1}^{H} \sum_{j=1}^{J} M_{h,j} \neq N \ \& \ I_1 \neq \phi)$ do
   - for each $(h \in \{1, ..., H\})$
     - for $(j \in I[h,])$, $\delta_{h,j} \leftarrow f_{h,j}(M_{h,j} + 1) - f_{h,j}(M_{h,j})$
   - choose $j^* \leftarrow \min \operatorname{argmin}_{j \in \{1,...,J\}} \sum_{h=1}^{H} f_{h,j}$
   - choose $h^* \leftarrow \min \operatorname{argmin}_{h \in \{1,...,H\}} \delta_{h,j^*}$
   - if $(\sum_j M_{h^*,j} < M_h^*)$
     - if $(M_{h^*,j^*} < u_{h,j^*}$ and $I[h^*, j^*] \neq NA)$, then $M_{h^*,j^*} \leftarrow M_{*h,j^*} + 1$, else $I[h^*, j^*] \leftarrow NA$

     else $I[h^*,] \leftarrow NA$
   - $I_1 \leftarrow vec(I[!is.na(I)])$
5. Return optimal $(M_{1,1}, \ldots, M_{H,J})$.