Optimal allocation to treatments in a sequential multiple assignment randomized trial

Andrea Morciano[1] and Mirjam Moerbeek[2] (iD)

Abstract

One of the main questions in the design of a trial is how many subjects should be assigned to each treatment condition. Previous research has shown that equal randomization is not necessarily the best choice. We study the optimal allocation for a novel trial design, the sequential multiple assignment randomized trial, where subjects receive a sequence of treatments across various stages. A subject's randomization probabilities to treatments in the next stage depend on whether he or she responded to treatment in the current stage. We consider a prototypical sequential multiple assignment randomized trial design with two stages. Within such a design, many pairwise comparisons of treatment sequences can be made, and a multiple-objective optimal design strategy is proposed to consider all such comparisons simultaneously. The optimal design is sought under either a fixed total sample size or a fixed budget. A Shiny App is made available to find the optimal allocations and to evaluate the efficiency of competing designs. As the optimal design depends on the response rates to first-stage treatments, maximin optimal design methodology is used to find robust optimal designs. The proposed methodology is illustrated using a sequential multiple assignment randomized trial example on weight loss management.

Keywords

sequential multiple assignment randomized trial trials, optimal allocation, efficiency, response rates, cost constraint, maximin designs

Introduction

In many randomized controlled trials, participants are equally allocated to intervention arms. Such a design is consistent with the view of clinical equipoise that must exist before the start of the trial.[1] However, it may be preferable to allocate more participants to one arm than to another, for instance, when variances and/or costs vary across the treatment arms,[1–5] or when outcomes are categorical rather than quantitative.[6–10] The derivation of the optimal allocation of units to treatment conditions has not only been done for individually randomized trials, but also for more complex trial designs such as cluster-randomized trials,[11–16] and trials with partially nested data.[17–19] From a statistical point of view, it is more efficient to assign more subjects to the condition with the lowest costs and highest variance. Other, more practical, reasons to use unequal allocation over equal allocation include resource constraints, administrative, political or ethical concerns or when the aim is to gain experience from an intervention and to study its feasibility.[5,20]

The focus of these references is on trials where subjects are randomized to either one single treatment or a combination of treatments, but do not change their assigned treatments during the course of the trial. This is a drawback since in real research practice some subjects may benefit more from one treatment and others more from another. Adaptive treatment strategies (ATSs), which are also called dynamic treatment regimens or adaptive interventions, are more flexible in the

[1]Findomestic Banca, Florence, Italy
[2]Department of Methodology and Statistics, Utrecht University, the Netherlands

Corresponding author:
Mirjam Moerbeek, Department of Methodology and Statistics, Utrecht University, PO Box 80140, 3508 TC Utrecht, the Netherlands.
Email: m.moerbeek@uu.nl

sense that they allow changing treatments over time.[21–24] An ATS individualizes treatments to subjects via decision rules that adjust the type, intensity, dosage or delivery of a treatment and specify when, whether and how to proceed at certain critical clinical decisions. For instance, those subjects for whom their assigned treatment turns out to be beneficial may continue the same treatment, while those others may be assigned to another treatment. The use of sequential treatments is often necessary because of: (i) heterogeneous treatment outcomes across subjects, (ii) change in treatment goals over time, (iii) the need to balance potential risks and benefits or (iv) to reduce costs when intensive treatment is not necessary.[25,26] Also, the use of sequential treatments implies multiple clinical decisions to be taken throughout the course of the study. These clinical decisions are formalized through ATSs.

Based on the number of treatments and treatment switches, various competing ATSs may be developed and they may be compared to one another in a so-called sequential multiple assignment randomized trial (SMART).[25,27] SMARTs are multi-stage randomized trial designs that are used to inform on the development of multiple ATSs embedded in it. The use of SMART designs allows researchers to evaluate the timing, sequencing and adaptive selection of treatments by using randomization and developing the best sequence(s) of treatments that lead to the optimal outcomes in the long term. In SMARTs, participants are allowed to switch through multiple stages, where each stage corresponds to a clinical decision, and subjects may be randomized at each stage. Sequenced randomization ensures that at each decision point the groups of participants assigned to the intervention options are balanced in terms of patient characteristics. This adds flexibility, allowing participants to remain on those treatments that are having an effect and giving the possibility to switch away to patients being treated with less effective options. This has made SMART designs appealing in a broad variety of health care, behavioural and psychological settings.

Multiple ATSs are embedded in a SMART and the main question in the design phase of a SMART is how many subjects should be assigned to each ATS, and whether an unequal allocation is better than an equal allocation. Some recent papers studied the relation between sample size and power for SMART designs,[25,28–34] but did not study the optimal allocation of units to treatment sequences and the loss of efficiency of using equal rather than unequal allocation.

The aim of this paper is to derive optimal allocations of units for a prototypical SMART design. This is a two-stage design where all units are randomized to two treatment conditions in the first stage. Those who respond to their assigned treatment are not re-randomized in the second stage, while those who do not respond are re-randomized to two second-stage treatments. This design was considered earlier by NeCamp et al.[32] in the setting of a cluster-randomized trial. In our contribution, we focus on individual randomization. We focus on sample sizes to be used when comparing two ATSs that start with different first-stage treatments. Four of such pairwise comparisons can be made in their prototypical SMART design, and one comparison may be of more importance than another. We therefore use multiple-objective optimal design methodology to consider all comparisons simultaneously, while taking into account their relative importance.[35] Multiple-objective optimal designs are useful when the study has multiple and conflicting objectives, such multiple pairwise comparisons of marginal means of ATSs in a SMART. It combines these objectives in one optimality criterion and tries to seek a design that is highly efficient for each of these criteria. We provide a Shiny App to calculate the optimal allocation of units and to evaluate the efficiency of the design with equal allocation. We demonstrate our optimal design methodology on the basis of a SMART example that compares two different treatments, nutrition (NUT) and physical activity (PHY), for weight loss management. Our focus is on SMARTs with a quantitative outcome with individual randomization. In other words, we do not focus on cluster-randomized SMARTs or other complex SMART designs with clustered data.

The remainder of our contribution is organized as follows. Section 'Prototypical SMART design' further discusses the prototypical SMART design and its embedded ATSs. Furthermore, this section introduces the example of weight loss management. Section 'Derivation of the optimal design' derives the optimal allocation of units for studies in which either the total sample size or the budget is fixed. In the latter case, we consider the realistic situation where costs may vary across treatment conditions. The optimal allocation turns out to depend on the subjects' probabilities to respond to their first-stage treatment. We therefore also focus on maximin optimal designs that are robust to incorrect prior estimates of these probabilities. Furthermore, Section 'Derivation of the optimal design' introduces the Shiny App that we developed for finding the optimal design. Section 'A SMART example' demonstrates our optimal design methodology on the basis of the weight loss example. It shows how the optimal design is influenced by the costs per treatment, proportion of responders to first-stage treatments and the relative importance of the four pairwise comparisons. Section 'Discussion' summarizes our findings, discusses limitations of this contribution and gives directions for future research.

Prototypical SMART design

Before we focus on the prototypical SMART, we rehearse some general ingredients for arbitrary SMART (see for instance Ertefaie et al.,[36] but using different notation). The observed covariates and treatment assignment at stage k are denoted

O_k and X_k, respectively, and $\bar{O}_k$ and $\bar{X}_k$ denote the covariate and treatment histories up to and including stage k. Within a SMART multiple ATSs are embedded; these are denoted d_i, $i = 1, \ldots, I$. An ATS is basically a treatment trajectory and denoted by a vector of counterfactual treatment assignments for a given individual j. If the SMART has two stages, then $d_i = (X_1, X_2^R, X_2^{NR})$, where X_2^R is the treatment assignment in the second stage had the subject responded, and X_2^{NR} is the treatment assignment in the second stage had he or she not responded. So, for a subject who responds, X_2^{NR} is not observed, and for a subject who does not respond X_2^R, hence d_i is called a vector of counterfactual treatments. The observed treatment history only includes the treatments a subject has actually been assigned to $\bar{X}_2 = (X_1, X_2)$. At the end of each stage k, a tailoring variable is measured which determines if a subject has responded to the treatment in that stage or not. In other words, this variable determines which treatment the subject is assigned to in the subsequent stage. At the end of the study (i.e. at the end of the final stage) the continuous outcome variable Y_j is measured on each subject. These outcomes are then used to compare different ATSs to one another.

The prototypical SMART design is visualized in Figure 1. This design has been used in various research fields; published examples of its use in the treatment and long-term management of many chronic conditions include weight loss,[26,37,38] substance abuse,[39,40] cancer research,[41,42] adolescent depression,[43] adolescent conduct problems,[44] suicide,[45] and attention-deficit/hyperactivity disorder.[46]

The prototypical SMART is a two-stage design with two first-stage treatments A and B; the proportions randomized to these treatments are denoted p_1 and $1 - p_1$, respectively. After some amount of time it is determined which subjects respond to their first-stage treatment, depending on some criterion such as a sufficient amount of weight loss or smoking cessation. The response rates to first-stage treatments A and B are equal to γ_1 and γ_2, respectively. Those subjects who respond to their first-stage treatment are not further randomized, but receive second-stage treatment C or F, depending on their first-stage treatment. This may be the same as the first-stage treatment, but may also be another treatment or discontinuation of treatment with or without further monitoring. Those subjects who do not respond to their first-stage
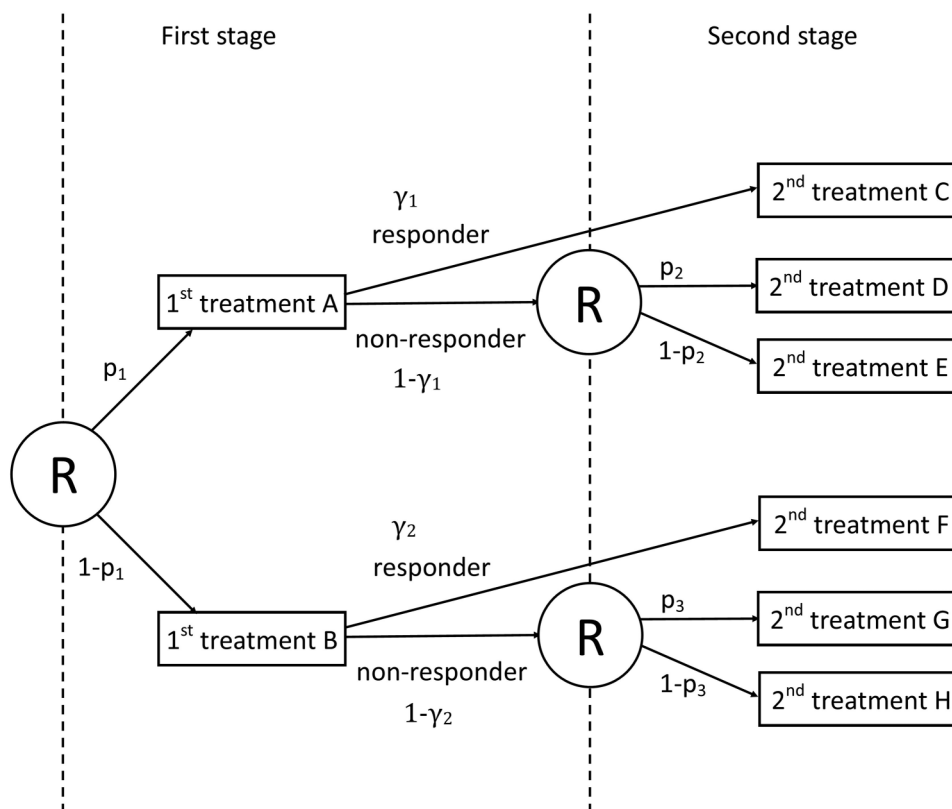

Figure 1. A scheme of the prototypical sequential multiple assignment randomized trial (SMART) design from NeCamp et al.[32] Circled 'R' denotes randomization at each stage. p_1 and $(1 - p_1)$ are, respectively, the proportions of subjects receiving first-stage treatments A and B. p_2 and $(1 - p_2)$ are, respectively, the proportions of subjects receiving second-stage treatments D and E for non-responders starting with first-stage treatment A. p_3 and $(1 - p_3)$ are, respectively, the proportions of subjects receiving second-stage treatments G and H for non-responders starting with first-stage treatment B. γ_1 and γ_2 indicate, respectively, response rates for the first-stage treatments A and B.

treatment are further randomized. Non-responders who received first-stage treatment A are randomized to second-stage treatments D and E, with proportions p_2 and $1 - p_2$, respectively. Such a second-stage treatment may be an intensified version of the first-stage treatment A, treatment A augmented with another treatment (which may be first-stage treatment B), first-stage treatment B, or an entirely different treatment. In the same manner, non-responders who received first-stage treatment B are randomized to two second-stage treatments G and H. This design includes eight different treatment conditions, where some of the second-stage treatments may be the same as the first-stage treatments or a combination of them.

Four ATSs are embedded in the prototypical SMART design, see Table 1. For instance, the first ATS, denoted d_1, assigns all subjects to first-stage treatment A. Responders receive second-stage treatment C while non-responders receive second-stage treatment D.

The primary analysis goal of a SMART design is usually one of the following: (i) comparing first-stage intervention options; (ii) comparing second-stage intervention options; (iii) comparing two or more embedded ATSs in the study starting with the same first-stage intervention option or (iv) comparing two or more embedded ATSs in the study starting with different first-stage intervention options.[31] In the derivation of our optimal design, we focus on embedded ATSs that start with different first-stage treatments, which is a common primary aim in SMARTs.[32]

Example: weight loss management

Bariatric surgery is an effective treatment for obese patients to lose weight. Given its costs, potentially harmful side effects and the risk of death, patients in the Netherlands are only considered eligible if they can demonstrate they have previously attempted other means to lose weight. Two treatments are an increase in PHY and a change in NUT.

Figure 2 visualises the example SMART design. All patients are first randomized to either PHY or NUT. Then, at the end of the first stage, subjects are categorized as responders or non-responders, according to some predefined definition of response, for example, a threshold for weight loss after a given period of time. Non-responders are then re-randomized to second-stage treatments, regardless of their treatment in the first-stage. They either switch to the other treatment or pursue with a combination of both treatments (NUT + PHY) in the second stage. Responders are not re-randomized and pursue with their first-stage treatment. This example is visualized in Figure 2. Four different ATSs are embedded within this prototypical SMART design: (i) $d_1 = (\text{PHY}, \text{PHY}^R, \text{NUT}^{NR})$, (ii) $d_2 = (\text{PHY}, \text{PHY}^R, (\text{NUT} + \text{PHY})^{NR})$, (iii) $d_3 = (\text{NUT}, \text{NUT}^R, \text{PHY}^{NR})$ and (iv) $d_4 = (\text{NUT}, \text{NUT}^R, (\text{NUT} + \text{PHY})^{NR})$. The superscript ^{R} refers to second-stage treatment assigned to responders, while the superscript ^{NR} denotes second-stage treatment assigned to non-responders.

The SMART design of this example is a simplification of the prototypical SMART design in the sense that just two treatments are involved. Responders continue with their first-stage treatment, while non-responders are randomized to the other treatment or a combination of both treatments. This specific SMART design was previously used for, among others, the treatment of anxiety disorder,[25] obsessive–compulsive disorder[47] and chronic pain.[48]

Derivation of the optimal design

Introduction

For a given ATS d_i, $i = 1, \ldots, 4$, let Y_j, $j = 1, \ldots, N_{d_i}$ be the continuous primary outcome of interest for the jth subject as measured at the end of stage 2, with N_{d_i} denoting the number of subjects whose treatment trajectories are

Table 1. The four ATSs embedded in the prototypical SMART design.

ATS label	First-stage treatment	Status at the end of first-stage	Second-stage treatment
d_1	A	Responder	C
		Non-responder	D
d_2	A	Responder	C
		Non-responder	E
d_3	B	Responder	F
		Non-responder	G
d_4	B	Responder	F
		Non-responder	H

ATS: adaptive treatment strategy; SMART: sequential multiple assignment randomized trial.
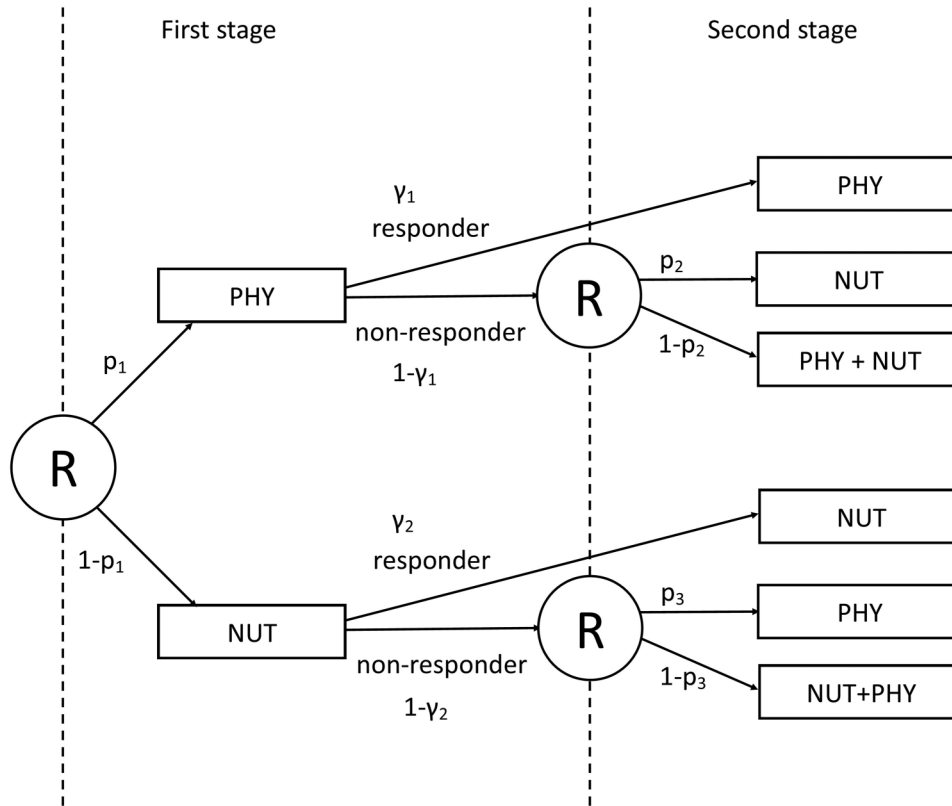
Figure 2. A scheme of the example SMART design on weight loss. Circled 'R' denotes randomization at each stage. p_1 and $(1-p_1)$ are, respectively, the proportions of subjects receiving the two first-stage treatments: PHY and NUT. p_2 and $(1-p_2)$ are, respectively, the proportions of subjects receiving second-stage treatments NUT and NUT + PHY for subjects starting with PHY as first-stage treatment. p_3 and $(1-p_3)$ are, respectively, the proportions of subjects receiving second-stage treatments PHY and NUT + PHY for subjects starting with NUT as first-stage treatment. γ_1 and γ_2 indicate, respectively, response rates for the first-stage treatments PHY and NUT. SMART: sequential multiple assignment randomized trial; PHY: physical activity; NUT: nutrition.

consistent with the ATS d_i. Y_j is supposed to have $E(Y_j) = \mu_i$ and $Var(Y_j) = \sigma^2$, for all $j = 1, \ldots, N_{d_i}$. We assume common variance σ^2 across all four ATSs. The target parameter μ_i, the marginal mean outcome expected under ATS d_i, depends on the proportion of responders to first-stage treatment in ATS d_i in the population. It is estimated by a weighted average of the observed outcomes of subjects whose treatment trajectories are consistent with d_i.[31]

The weights follow from the fact that there is a structural imbalance between responders and non-responders: the non-responders are re-randomized but the responders are not. For instance, for ATS d_1, responders have a probability of p_1 of receiving the treatment sequence they actually received, and their subject-specific weights are $W_j = 1/p_1$. For non-responders, this probability is $p_1 p_2$ and hence their weight is $W_j = 1/p_1 p_2$. Here p_1 is the randomization probability to treatment A in the first-stage and p_2 is the randomization probability to treatment C in the second-stage. The weights are the inverse of the probabilities, hence the weighting is called inverse probability weighting. By using these weights, the relative contribution of the responders and non-responders in the calculation of the weighted mean outcome in ATS d_1 is the same as when this ATS had not been embedded in a SMART. In other words, since the ATS is embedded in a SMART, the non-responders have a higher weight than the responders to account for the fact that some of them are randomized to treatment E, rather than treatment D. This is a generalization from the work of Ghosh et al.[31] in the sense that we allow the proportions p_1 and p_2 to be unequal to 0.5. For the other ATSs, subject-specific weights can be obtained in a similar way.

The weighted mean for the continuous primary outcome of interest for ATS d_i is equal to

$$\bar{Y}_{d_i} = \frac{\sum_j^{N_{d_i}} W_j Y_j}{\sum_j^{N_{d_i}} W_j} \tag{1}$$

The expected value of this weighted mean is given by

$$E(\bar{Y}_{d_i}) = \frac{\sum_j^{N_{d_i}} W_j E(Y_j)}{\sum_j^{N_{d_i}} W_j} = \mu_i \tag{2}$$

Equation (2) shows that the weighted mean is an unbiased estimator of the marginal mean. The variance of the weighted mean is equal to

$$\text{Var}(\bar{Y}_{d_i}) = \sigma^2 \frac{\sum_j^{N_{d_i}} W_j^2}{\left(\sum_j^{N_{d_i}} W_j \right)^2} \tag{3}$$

For each ATS d_i, the variance of the weighted mean is computed using the subject-specific weights. First, the expected number of people in the trial whose treatment trajectories are consistent with d_i is computed for each ATS. For d_1, this is equal to

$$E(N_{d_1}) = Np_1\gamma_1 + Np_1(1 - \gamma_1)p_2, \tag{4}$$

with the first term on the right side representing the expected number of responders and the second being the expected number of non-responders. The proportions p_1 and p_2 are defined as above, while N is the total sample size of the SMART and γ_1 is the response rate to first-stage treatment A.

Following from (4), we obtain

$$\sum_j^{N_{d_1}} W_j = \frac{1}{p_1} Np_1\gamma_1 + \frac{1}{p_1 p_2} Np_1(1 - \gamma_1)p_2 = N \tag{5}$$

$$\sum_j^{N_{d_1}} W_j^2 = \frac{1}{p_1^2} Np_1\gamma_1 + \frac{1}{(p_1 p_2)^2} Np_1(1 - \gamma_1)p_2 = \frac{N\gamma_1}{p_1} + \frac{N(1 - \gamma_1)}{p_1 p_2} \tag{6}$$

The variance for the weighted mean $\bar{Y}_{d_1}$, for ATS d_1, is obtained by plugging (5) and (6) into (3):

$$\text{Var}(\bar{Y}_{d_1}) = \frac{\sigma^2}{N} \frac{\gamma_1 p_2 + (1 - \gamma_1)}{p_1 p_2} \tag{7}$$

The right side of (7) consists of two factors. The first is the common variance of a mean, while the second is used to account for the fact that subjects may be re-randomized. This second factor is a function of the response rate γ_1 to first-stage treatment A.

Using their respective subject-specific weights, formulae for the variance of the weighted mean $\bar{Y}_{d_i}$ for the other ATSs are obtained in a similar way; these are shown in Table 2.

We consider pairwise comparisons of ATSs that start with different first-stage treatments. The expected difference in weighted means of two such ATSs d_i and $d_{i'}$ (with $i = 1$ or 2 and $i' = 3$ or 4) is $\mu_i - \mu_{i'}$ with the corresponding variance

$$\text{Var}(\bar{Y}_{d_i} - \bar{Y}_{d_{i'}}) = \text{Var}(\bar{Y}_{d_i}) + \text{Var}(\bar{Y}_{d_{i'}}) + 2\text{Cov}(\bar{Y}_{d_i}, \bar{Y}_{d_{i'}}) = \text{Var}(\bar{Y}_{d_i}) + \text{Var}(\bar{Y}_{d_{i'}}) \tag{8}$$

where $\text{Cov}(\bar{Y}_{d_i}, \bar{Y}_{d_{i'}}) = 0$ if we assume that weighted means of ATSs that start with different first-stage treatments are independent. This assumption holds as long as outcomes of subjects from ATSs that start different first-stage treatments are independent

Table 2. Variance for the weighted mean $\bar{Y}_{d_i}$ for the four adaptive treatment strategies (ATSs) embedded.

d_i	$\text{Var}(\bar{Y}_{d_i})$
$d_1 = (A, C^R, D^{NR})$	$\sigma^2 \frac{\gamma_1 p_2 + (1 - \gamma_1)}{Np_1 p_2}$
$d_2 = (A, C^R, E^{NR})$	$\sigma^2 \frac{\gamma_1 (1 - p_2) + (1 - \gamma_1)}{Np_1(1 - p_2)}$
$d_3 = (B, F^R, G^{NR})$	$\sigma^2 \frac{\gamma_2 p_3 + (1 - \gamma_2)}{N(1 - p_1)p_3}$
$d_4 = (B, F^R, H^{NR})$	$\sigma^2 \frac{\gamma_2 (1 - p_3) + (1 - \gamma_2)}{N(1 - p_1)(1 - p_3)}$

Considering the ATSs embedded in our example, four possible pairwise comparisons exist, with corresponding variances: (i) $\Phi_{13} = \text{Var}(\bar{Y}_{d_1} - \bar{Y}_{d_3})$; (ii) $\Phi_{14} = \text{Var}(\bar{Y}_{d_1} - \bar{Y}_{d_4})$; (iii) $\Phi_{23} = \text{Var}(\bar{Y}_{d_2} - \bar{Y}_{d_3})$ and (iv) $\Phi_{24} = \text{Var}(\bar{Y}_{d_2} - \bar{Y}_{d_4})$, with $\bar{Y}_{d_i}$ being the weighted mean for the continuous primary outcome variable of interest for the ATS d_i, $i = 1, \ldots, 4$. Formulae for the variance of these comparisons can be derived by plugging in the variances of the single ATSs as reported in Table 2.

The optimal design $\xi^*_{ii'} = (p^*_1, p^*_2, p^*_3)$ for objective $\Phi_{ii'}$ is the one among all designs $\xi = (p_1, p_2, p_3)$ in the design space $\Omega = (0 \leq p_1 \leq 1, 0 \leq p_2 \leq 1, 0 \leq p_3 \leq 1)$ for which $\text{Var}(\bar{Y}_{d_i} - \bar{Y}_{d_{i'}})$ is minimized. Each objective has its own optimal design. For instance, the optimal design for Φ_{13} is $\xi^*_{13} = (0.5, 1, 1)$, which implies both first-stage treatments have randomization probability 0.5, all non-responders in first-stage treatment A receive second-stage treatment D, and all non-responders to first-stage treatment B receive second-stage treatment G. The optimal designs for the other objectives are $\xi^*_{14} = (0.5, 1, 0)$, $\xi^*_{23} = (0.5, 0, 1)$ and $\xi^*_{24} = (0.5, 0, 0)$. The optimal design for one objective does not only hold for the other single objectives, but it may also perform poorly.[49] For that reason, a multiple-objective optimal design is used, so that all of the four pairwise comparisons are taken into account simultaneously. We do so by using a weighted sum of the four objectives, where weights are to be chosen by the user. The use of weights allows placing more emphasis on the one objective than another, subject to the researcher's interests and the goals of the study. A constraint is put on the weights such that their sum is equal to 1. The optimal design problem becomes a multiple-objective optimal design problem. The aim is to minimize the optimality criterion

$$\Phi = \Phi_{13}\lambda_{13} + \Phi_{14}\lambda_{14} + \Phi_{23}\lambda_{23} + \Phi_{24}\lambda_{24} \tag{9}$$

with $\lambda_{ii'}$ being the weight assigned to the respective objective $\Phi_{ii'}$. The corresponding optimal design is a so-called compound-optimal design.

Optimal design under a fixed total sample size

In this scenario, the optimal design is sought under a fixed total sample size N. This is a realistic scenario when studying treatments for a rare disease or condition, but it can also be used when resource constraints allow recruiting a fixed number of subjects. It is assumed that a priori estimates of the response rates γ_1 and γ_2 are available.

The optimal design minimizes the objective in (9); it is found by taking the gradient of (9) with respect to p_1, p_2 and p_3. The optimal proportions for the second-stage treatments are given by

$$p^*_2 = \frac{\sqrt{(\lambda_{13} + \lambda_{14})}}{\sqrt{(\lambda_{23} + \lambda_{24})} + \sqrt{(\lambda_{13} + \lambda_{14})}} \tag{10}$$

and

$$p^*_3 = \frac{\sqrt{(\lambda_{13} + \lambda_{23})}}{\sqrt{(\lambda_{13} + \lambda_{23})} + \sqrt{(\lambda_{14} + \lambda_{24})}} \tag{11}$$

It is worth noting that the optimal second-stage proportions p^*_2 and p^*_3 do not depend on the response rates γ_1 and γ_2, or on the total sample size N, but only on the choice of the weights. In particular, p^*_2 increases as λ_{13} and/or λ_{14} increase. This is obvious since objectives Φ_{13} and Φ_{14} are comparisons that include treatment D, and more efficient comparisons can be made if more subjects are assigned to this treatment. Similarly, p^*_3 increases when λ_{13} and/or λ_{23} increase. This is also obvious since objectives Φ_{13} and Φ_{23} are comparisons that include treatment G, and more efficient comparisons can be made if more subjects are assigned to this treatment.

The optimal randomization probability for the first-stage treatment A takes on a more complicated form:

$$p^*_1 = \frac{\sqrt{\varphi}}{\sqrt{\psi} + \sqrt{\varphi}} \tag{12}$$

where

$$\psi = [(\gamma_2 p^*_3 + (1 - \gamma_2))(\lambda_{13} + \lambda_{23})(1 - p^*_3) + (\gamma_2(1 - p^*_3) + (1 - \gamma_2))(\lambda_{14} + \lambda_{24})p^*_3]p^*_2(1 - p^*_2) \tag{13}$$

$$\varphi = [(\gamma_1 p^*_2 + (1 - \gamma_1))(\lambda_{13} + \lambda_{14})(1 - p^*_2) + (\gamma_1(1 - p^*_2) + (1 - \gamma_1))(\lambda_{23} + \lambda_{24})p^*_2]p^*_3(1 - p^*_3) \tag{14}$$

p^*_1 depends on both γ_1 and γ_2, and on the optimal proportions p^*_2 and p^*_3, while it does not depend on N. A detailed derivation of the optimal design is given in the online supplement.

Optimal design under a fixed budget

In this scenario, we consider a budgetary constraint: the total costs C for treating subjects should not exceed the budget B. The costs are calculated as

$$C = c_A N_A + c_B N_B + \cdots + c_H N_H \tag{15}$$

where c_A are the costs per subject in treatment A and N_A are the number of subjects who receive treatment A, and similarly for the other treatments B to H. The costs may vary across subjects and are assumed to be known beforehand. The sample sizes are stochastic since they depend on the proportions p_1, p_2 and p_3 and response rates γ_1 and γ_2. In the derivation of the optimal design, we use their expected values. For the first-stage treatments, we have

$$E(N_A) = p_1 N \quad \text{and} \quad E(N_B) = (1 - p_1)N \tag{16}$$

for the second-stage treatments C, D and E, we have

$$E(N_C) = \gamma_1 p_1 N \text{ and } E(N_D) = \gamma_1 p_1 p_2 N \text{ and } E(N_E) = \gamma_1 p_1 (1 - p_2)N \tag{17}$$

and for the second-stage treatments F, G and H, we have

$$E(N_F) = \gamma_2 (1 - p_1)N \text{ and } E(N_G) = \gamma_2 (1 - p_1)p_3 N \text{ and } E(N_H) = \gamma_2 (1 - p_1)(1 - p_3)N \tag{18}$$

For a given budget, the total sample size N that can be used decreases when the costs increase. This implies that a design is not only determined by the proportions but also by the total sample size: $\xi = (p_1, p_2, p_3, N)$. The optimal design is found in a numerical manner through a domain search algorithm, see the online supplement for more details.

Robust optimal design

The optimal design depends on the response rates γ_1 and γ_2, hence the optimal design is locally optimal. These parameters are often unknown in the design stage of a SMART and an educated a priori guess based on expert opinions or findings in the literature should be used. There is, however, no guarantee that such a guess is correct and robust optimal design methodology may be used to protect against a loss of efficiency due to a misspecification of the response rates. We use maximin optimal design methodology[50] to allow specification of intervals, rather than point estimates, of the two response rates.

The maximin optimal design ξ_{MMD} maximizes the minimal relative efficiency (RE) among all designs in the design space Ω. In other words, it selects the best of the worst-case scenarios. The maximin optimal design can be found using the following three steps:

1. Define the parameter space for the response rates and the design space Ω for the proportions. For instance, the first response rate γ_1 may be between 0.2 and 0.3 and the second response rate γ_2 may be between 0.35 and 0.45. The design space is $\Omega = (0 \leq p_1 \leq 1, \ 0 \leq p_2 \leq 1, \ 0 \leq p_3 \leq 1)$.
2. For each possible combination of the two response rates in the parameter space, compute the locally optimal design ξ_{LOD}. Then compute the RE of each design ξ in Ω compared with the locally optimal design: $RE = \Phi(\xi_{LOD}) / \Phi(\xi)$.
3. For each design in Ω, find its smallest RE value within the parameter space. Then, select the design that has the highest minimum RE across all designs in the design space. This is the maximin optimal design ξ_{MMD} and its minimum RE is called the maximin value.

This procedure yields the design which is most robust to a misspecification of the response rates and it can be used when working under a fixed budget or under a fixed total sample size.

Statistical power for the optimal design

Once the optimal allocation to treatments has been derived, it makes sense to determine how much power the study has for each of the four pairwise comparisons of ATSs [51]. The following steps should be taken in such a power analysis:

1. Calculate the variance $\text{Var}(\bar{Y}_{d_i})$ for each of the four ATSs in the SMART. For the case of a fixed total sample size this can be done easily by plugging in the optimal proportions p_1^*, p_2^* and p_3^* and total sample size N into the equations of Table 2. For the case of a fixed budget, first, the total sample size N has to be calculated from the budget, costs and optimal proportions. This can be done on the basis of equations (15) to (18), as is further explained in the online supplement.

2. For each of the four pairwise comparisons of ATSs: calculate $\text{Var}(\bar{Y}_{d_i} - \bar{Y}_{d_{i'}})$ from equation (8).

3. For each of the four pairwise comparisons of ATSs, get a prior estimate of the expected difference in marginal means $\mu_i - \mu_{i'}$. A prior estimate may be obtained from the literature or an expert's expectations. As an alternative, one may use the minimal relevant effect size, that is, the smallest effect size that is considered to be relevant.

4. For each of the four pairwise comparisons of ATSs, select the type I error rate α and decide whether a one-sided or two-sided test has to be performed.

5. For each of the four pairwise comparisons of ATSs, calculate the power. For a one-sided alternative use the following equation:

$$1 - \beta = z^{-1}\left(\frac{|\mu_i - \mu_{i'}|}{\text{Var}(\bar{Y}_{d_i} - \bar{Y}_{d_{i'}})} - z_{1-\alpha}\right)$$

where $\text{Var}(\bar{Y}_{d_i})$ and $\text{Var}(\bar{Y}_{d_{i'}})$ are the variances of the two ATSs to be compared, $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of the standard normal distribution and z^{-1} is the inverse of the standard normal distribution. For a two-sided alternative, α has to be replaced by $\alpha / 2$.

Shiny app

We developed a Shiny app[52] to facilitate finding the optimal design; it is available from https://andreamorciano.shinyapps.io/OptimalSMART/. It calculates locally optimal designs for a fixed total sample size as well as a fixed budget. In the first case, the user should specify the total sample size, in the latter case, he or she should specify the costs per treatment along with the budget. Furthermore, an a priori estimate of the two response rates should be specified to find the locally optimal design. The numerical algorithm that finds the optimal design for the budgetary constraint has a precision of 0.00002 for the optimal proportions.

The Shiny app can also be used to find the maximin optimal design. It that case intervals $[\gamma_1 - 0.05, \gamma_1 + 0.05]$ and $[\gamma_2 - 0.05, \gamma_2 + 0.05]$ are considered around the user-specified values γ_1 and γ_2. These intervals are continuous; in our algorithm, we use a step size of 0.01 to discretize these intervals, while a step size of 0.05 is used for the response rates. In the case the reader is interested in using a different step size, he/she can contact the first author.

A SMART example

Introduction

We apply the optimal design methodology to the example of the weight loss management study of Figure 2. Participants are randomized to two first-stage treatments: PHY and NUT. A response is defined as a (absolute or relative) loss in body weight that exceeds a user-selected threshold value. We use three sets of a priori guesses for the two response rates of the two first-stage treatments: $(\gamma_1, \gamma_2) = (0.15, 0.25)$, $(\gamma_1, \gamma_2) = (0.25, 0.40)$ and $(\gamma_1, \gamma_2) = (0.40, 0.55)$. In each case, we choose a larger value for NUT than for PHY, as previous research has demonstrated that PHY produces smaller bodyweight loss than diet (NUT).[53] For the first set of response rates, the definition of a response is most stringent, resulting in the smallest response rates, and for the third it is most lenient, resulting in the highest response rates.

We consider three sets of weights for the multiple-objective optimal design (9). The first considers each comparison to be of equal importance, which implies that equal weights are used: $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.25, 0.25, 0.25, 0.25)$. The second puts more emphasis on those comparisons where second-stage treatments are either PHY or NUT, but not a combination of the two. In this case, researchers are mainly interested in the comparison between $d_1 = (\text{PHY}, \text{PHY}^R, \text{NUT}^{NR})$ and $d_3 = (\text{NUT}, \text{NUT}^R, \text{PHY}^{NR})$ rather than the other ones. Designs with a single second-stage treatment are less expensive, they may be easier to implement by the researchers and easier to adhere to by the participants. As an illustration we use $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.70, 0.10, 0.10, 0.10)$. The third set of weights puts more emphasis on those second-stage treatments that are a combination of NUT and PHY, for instance, because there is a believe combined treatment is more effective. In that case the main focus is on the comparison between $d_2 = (\text{PHY}, \text{PHY}^R, (\text{NUT} + \text{PHY})^{NR})$ and $d_4 = (\text{NUT}, \text{NUT}^R, (\text{NUT} + \text{PHY})^{NR})$. As an illustration we use $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.10, 0.10, 0.10, 0.70)$.

For this specific example, we developed another version of our Shiny app; this is available at https://andreamorciano.shinyapps.io/OptimalSMART2/.

Locally optimal design under a fixed total sample size

For each combination of (γ_1, γ_2) and $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24})$, the optimal design is given in Table 3, along with the RE of the balanced design (where $p_1 = p_2 = p_3 = 0.50$) as compared to the optimal design. We observe the optimal design hardly depends on the response rates, but it does depend on the weights. For each set of weights, the optimal design dictates (about) equal randomization to first-stage treatments. For the first set of weights, the optimal design is (almost) equal to the balanced design and the RE of the balanced design is 1. For the second set of weights more than half (two-thirds) of participants are randomized to single second-stage treatments. This is obvious because the chosen weights put more emphasis on the comparison of single second-stage treatments. For the third set of weights, less than half (one-third) of participants are randomized to single second-stage treatments. This is also obvious because the chosen weights put more emphasis on the comparison of combined second-stage treatments. The optimal proportions p_2^* and p_3^* for the second set of weights are the complement of those for the third set of weights. In all cases, the RE of the balanced design is above 0.9, which implies it performs rather well as compared to the optimal design.

The results do not necessarily apply to other combinations of weights and response rates, so a researcher who is planning a SMART is advised to use our Shiny app to derive the optimal design for the trial at hand, and to do a sensitivity analysis to study how the optimal design is influenced using by various realistic combinations of weights and response rates.

Locally optimal design under a fixed budget

To find the optimal design under a budgetary constraint, the costs for both treatments and the budget need to be defined. We assume both stages are of equal length, so the costs do not vary across stages. The costs for combined treatment are the sum of the costs for both single treatments. We consider two sets of costs for NUT (C_N) and PHY (C_P): $(C_N, C_P) = (300, 50)$ and $(C_N, C_P) = (300, 300)$. Let us assume the costs are expressed in euros and the length of each stage is one month. The costs for NUT are a reasonable amount to buy healthy food for one participant per month in the Netherlands. The costs for PHY in the first set cover a subscription to the local gym for one month, those in the second set also include personal training by a fitness coach. Furthermore, the budget is $B = 100,000$. For the response rates and the weights, we consider the same sets of values as in Section 'Locally optimal design under a fixed total sample size'.

For $(C_N, C_P) = (300, 50)$, the optimal proportion p_1^* is somewhat above 0.5, which implies that in the first stage more subjects are randomized to the least expensive treatment PHY than to the more expensive treatment NUT. The optimal proportion p_1^* hardly depends on the chosen weights, but it slightly increases with increasing response rates. Higher response rates imply more subjects receive the same treatment in stage 2 as they did in stage 1. It is therefore advantageous to already randomize more subjects to the least expensive treatment PHY in stage 1, so that more subjects receive this treatment in stage 2 as well. For $(C_N, C_P) = (300, 300)$, both first-stage treatments are equally expensive and the optimal proportion p_1^* is (about) 0.5. It hardly depends on the chosen weights and the response rates.

The optimal proportions p_2^* and p_3^* hardly depend on the response rates but they do depend on the chosen weights. For the first set of weights, $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.25, 0.25, 0.25, 0.25)$, somewhat more subjects are randomized to the single second-stage treatments NUT or PHY than to the combined second-stage treatment PHY + NUT. This is obvious since single second-stage treatments are less expensive than combined treatments. For the second set of weights, $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.70, 0.10, 0.10, 0.10)$, even more subjects are randomized to single second-stage treatments than for the first set of weights. This is also obvious because the second set of weights puts more emphasis on the comparison of those ATSs with single second-stage treatments. For the third set of weights, $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.10, 0.10, 0.10, 0.70)$, more subjects are randomized to combined second-stage treatments than

Table 3. Locally optimal design: optimal proportions for first-stage (p_1^*) and second-stage (p_2^*, p_3^*) treatments for three different sets of weights ($\lambda_{13}, \lambda_{23}, \lambda_{14}, \lambda_{24}$) for the multiple-objective optimal design, and for three different sets of response rates (γ_1, γ_2). The relative efficiency (RE) of the balanced design is also provided. The optimal proportions are derived under a fixed total sample size.

| | | $(\lambda_{13}, \lambda_{23}, \lambda_{14}, \lambda_{24}) =$ (0.25, 0.25, 0.25, 0.25) | | | | $(\lambda_{13}, \lambda_{23}, \lambda_{14}, \lambda_{24}) =$ (0.70, 0.10, 0.10, 0.10) | | | | $(\lambda_{13}, \lambda_{23}, \lambda_{14}, \lambda_{24}) =$ (0.10, 0.10, 0.10, 0.70) | | | |
γ_1	γ_2	p_1^*	p_2^*	p_3^*	RE	p_1^*	p_2^*	p_3^*	RE	p_1^*	p_2^*	p_3^*	RE
0.15	0.25	0.50	0.50	0.50	1	0.50	0.67	0.67	0.91	0.50	0.33	0.33	0.91
0.25	0.40	0.51	0.50	0.50	1	0.51	0.67	0.67	0.92	0.51	0.33	0.33	0.92
0.40	0.55	0.51	0.50	0.50	1	0.51	0.67	0.67	0.93	0.51	0.33	0.33	0.93

to single second-stage treatments, which is also obvious because this set of weights puts more emphasis on the comparison of ATSs with combined second-stage treatments.

The optimal total sample size N^* depends on the combination of costs (C_N, C_P). As is obvious, fewer subjects can be included for $C_P = 300$ than for $C_P = 50$. Furthermore, N^* depends on the weights: most subjects can be included for the second set of weights and fewest for the third set of weights. For the second set of weights, more subjects are randomized to the least expensive single second-stage treatments, hence a larger total number of subjects can be included. Finally, more subjects can be included when the response rates increase. Subjects who respond to treatment are not re-randomized, hence they receive a single treatment in the second-stage. Single treatments are less expensive than combined treatments, hence more subjects can be included.

The RE of the balanced design slightly depends on the response rates. It is also related to the weights. The RE is highest for the first set of weights, since the optimal proportions are nearest to those of the balanced design. Slightly lower relative efficiencies are found for the third set of weights, but these relative efficiencies are still above 0.9. The lowest relative efficiencies are observed for the second set of weights as the optimal proportions deviate most from those of the balanced design. The lowest RE is $RE = 0.85$, which implies that the balanced design requires $100\%[1 - (1 / 0.85)] = 17\%$ more subjects than the optimal design.

Robust optimal design

The optimal designs that were presented in subsections 'Locally optimal design under a fixed total sample size' and 'Locally optimal design under a fixed budget' are locally optimal since they depend on the response rates γ_1 and γ_2. Such response rates are often unknown in the design phase of a SMART and an educated a priori guess must be given. There is, however, no guarantee such a guess is correct, and an incorrect guess may result in a suboptimal design. This problem may be overcome by using robust optimal design methodology; here we use the maximin optimal design methodology as described in section 'Robust optimal design'.

Tables 5 and 6 in the online supplement show maximin optimal designs using the same sets of weights and combinations of costs as in Tables 3 and 4. The ranges used for the response rates are $[\gamma_1 - 0.05, \gamma_1 + 0.05]$ and $[\gamma_2 - 0.05, \gamma_2 + 0.05]$, where γ_1 and γ_2 are the values in Tables 3 and 4.

A comparison of Table 3 and Table 5 of Supplemental material, and Table 4 and Table 6 of Supplemental material shows the locally optimal designs and maximin optimal designs are (almost) identical for the chosen sets of weights, response rates and costs. As a result, the minimal RE of the balanced design as given in Tables 5 and 6 of Supplemental material is almost equal to that of the RE of the balanced design in Tables 3 and 4. This result is not surprising since in Sections 'Locally optimal design under a fixed total sample size' and 'Locally optimal design under a fixed budget', it was shown that the optimal design hardly depends on the response rates. Of course, this finding does not necessarily hold for all combinations of responses rates, weights and costs. The user is therefore encouraged to apply maximin optimal design methodology in the case the response rates are likely to be misspecified.

Table 4. Locally optimal design: optimal proportions for first-stage (p_1^*) and second-stage (p_2^*, p_3^*) treatments for three different sets of weights (λ_{13}, λ_{23}, λ_{14}, λ_{24}) for the multiple-objective optimal design and for three different sets of response rates (γ_1, γ_2). The relative efficiency (RE) of the balanced design is also provided. The optimal proportions are derived under a fixed budget with C = 100, 000 and for two different sets of costs (C_P, C_N).

		(λ_{13}, λ_{23}, λ_{14}, λ_{24}) = (0.25, 0.25, 0.25, 0.25)					(λ_{13}, λ_{23}, λ_{14}, λ_{24}) = (0.70, 0.10, 0.10, 0.10)					(λ_{13}, λ_{23}, λ_{14}, λ_{24}) = (0.10, 0.10, 0.10, 0.70)				
γ_1	γ_2	p_1^*	p_2^*	p_3^*	N^*	RE	p_1^*	p_2^*	p_3^*	N^*	RE	p_1^*	p_2^*	p_3^*	N^*	RE
(C_N, C_P) = (300, 50)																
0.15	0.25	0.56	0.52	0.56	243	0.98	0.55	0.68	0.72	255	0.85	0.56	0.35	0.39	232	0.93
0.25	0.40	0.58	0.52	0.56	250	0.97	0.57	0.68	0.72	259	0.86	0.58	0.35	0.39	241	0.93
0.40	0.55	0.60	0.52	0.55	265	0.96	0.60	0.68	0.71	272	0.86	0.60	0.35	0.38	257	0.92
(C_N, C_P) = (300, 300)																
0.15	0.25	0.50	0.55	0.55	141	0.99	0.50	0.71	0.71	149	0.85	0.50	0.38	0.38	133	0.96
0.25	0.40	0.51	0.55	0.54	144	0.99	0.51	0.71	0.71	152	0.87	0.51	0.38	0.38	138	0.96
0.40	0.55	0.51	0.54	0.54	149	1	0.51	0.70	0.70	155	0.89	0.51	0.38	0.37	143	0.96

Discussion

Considering our example of a prototypical SMART design, we derived the optimal design $\xi^* = (p_1^*, p_2^*, p_3^*)$ both under a fixed sample size and budget constraint. Under a fixed sample size, we found that the optimal probability in the first-stage p_1^* is mostly influenced by the weights chosen for the multiple-objective optimal design, while it is only slightly influenced by the response rates. On the other hand, second-stage optimal probabilities are only influenced by the choice of the weights. When considering the second set of weights $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.70, 0.10, 0.10, 0.10)$ or the third set, $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.10, 0.10, 0.10, 0.70)$, which, respectively, put more emphasis on the use of single and combined treatments, the optimal design ξ^* performs better than the balanced design $\xi_b = (0.50, 0.50, 0.50)$, although the latter still achieves a RE above 0.90. When equal weights are used, ξ^* and ξ_b perform almost identically in terms of RE.

Under a fixed budget, the optimal proportions are influenced also by the cost of treatments, besides the aforementioned weights and response rates. When including cost of treatments into account, the performance in terms of RE of the optimal design ξ^*, with respect to ξ_b, improves. The reason might be that unequal allocation of patients to intervention options seems to work better under a fixed budget than under a fixed sample size, as was also previously stated in the literature.[2,3]

It is especially advised to use the optimal design rather than the balanced design when the second set of weights, $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.70, 0.10, 0.10, 0.10)$, is used. For this set, ξ_b may have a RE as low as 0.86. When using equal weights for the multiple-objective optimal design, ξ_b achieves a RE with respect to ξ^* above 0.95. When using the third set of weights, $(\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}) = (0.10, 0.10, 0.10, 0.70)$, ξ_b achieves a RE above 0.90.

It should be mentioned that the optimal designs are locally optimal, as they depend on the two unknown response rates γ_1 and γ_2. One way to address this issue is using maximin optimal design methodology. In our example, the maximin optimal designs are quite similar to the locally optimal designs. In other words, the locally optimal designs are rather robust with respect to mild misspecification of the response rates. However, this finding does not always hold and it is advocated to derive a maximin optimal design if there is uncertainty about the a priori guesses of the response rates.

We derived our optimal design under the assumption that outcomes of subjects in ATSs that start with different first-stage treatments are independent of each other, resulting in a zero correlation between weighted mean outcomes of ATSs starting with different first-stage treatments. There are situations in which this assumption may be violated. Consider for instance the situation in our weight loss example where just a limited number of personal trainers is available. It may then occur, a personal trainer trains subjects from ATSs starting with different first-stage treatments. In such a case, the outcomes of subjects who have been trained by the same personal trainer become dependent because of the trainer's skills, enthusiasm, experience, etc. In such a case, the assumption of independence is violated and hence our optimal design is not applicable. Such a problem can be easily solved by letting each personal trainer only train subjects from ATSs that start with the same first-stage treatment.

One limitation of this study is that it does not take clustered data structures into account, while such data may also occur in SMARTs.[54,55] Clustered data occur, among others, in cluster-randomized trials and multicentre trials. In such studies not only the total number of subjects in each treatment sequence needs to be determined, but also the number of clusters and cluster size.[56] The optimal design will depend on the intraclass correlation coefficient, which measures the degree of dependence of outcomes within the same cluster.

Another limitation of this study is that formulae and methodology only apply to the prototypical SMART designs in Figures 1 and 2. Based on the number of treatments, stages and randomizations, different SMART designs can be developed, of which many examples exist in the literature[57,58] and online.[59] It would be necessary to study optimal designs for such other types of SMART designs.

To our knowledge, this is the first paper that studies optimal allocation to treatments in SMARTs. Our Shiny App allows researchers in the fields of biomedical, health and social sciences to derive the optimal design for their SMART and to calculate the efficiency of a balanced design. We hope that this paper will further contribute to the development and implementation of SMARTs.

ORCID iD

Mirjam Moerbeek ID https://orcid.org/0000-0001-5537-1237

Supplemental material

Supplemental material for this article is available online.

References

1. Sverdlov O and Ryeznik Y. Implementing unequal randomization in clinical trials with heterogeneous treatment costs. *Stat Med*. Epub ahead of print 2019; **38**: 2905–2927. DOI: 10.1002/sim.8160.
2. Schouten HJA. Sample size formula with a continuous outcome for unequal group sizes and unequal variances. *Stat Med* 1999; **18**: 87–91.
3. Singer J. A simple procedure to compute the sample size needed to compare two independent groups when the population variances are unequal. *Stat Med* 2001; **20**: 1089–1095.
4. Jan SL and Shieh G. Optimal sample sizes for Welch's test under various allocation and cost considerations. *Behav Res Methods* 2011; **43**: 1014–1022.
5. Dumville JC, Hahn S, Miles JNV, et al. The use of unequal randomisation ratios in clinical trials: A review. *Contemp Clin Trials* 2006; **27**: 1–12.
6. Demidenko E. Sample size determination for logistic regression revisited. *Stat Med* 2007; **26**: 3385–3397.
7. Demidenko E. Sample size and optimal design for logistic regression with binary interaction. *Stat Med* 2008; **27**: 36–46.
8. Dette H. On robust and efficient designs for risk estimation in epidemiological studies. *Scand J Stat* 2004; **31**: 319–331.
9. Zhu W and Wong WK. Optimum treatment allocation for dual-objective clinical trials with binary outcomes. *Commun Stat Theory Methods* 2000; **29**: 957–974.
10. Yang J, Mandal A and Majumdar D. Optimal designs for two-level factorial experiments with binary response. *Stat Sin* 2012; **22**: 885–907.
11. Lemme F, van Breukelen GJP and Berger MPF. Efficient treatment allocation in 2×2 cluster randomized trials, when costs and variances are heterogeneous. *Stat Med* 2016; **35**: 4320–4334.
12. van Breukelen GJP and Candel MJJM. Efficient design of cluster randomized trials with treatment-dependent costs and treatment-dependent unknown variances. *Stat Med* 2018; **37**: 3027–3046.
13. Lemme F, Van Breukelen GJP, Candel MJJM, et al. The effect of heterogeneous variance on efficiency and power of cluster randomized trials with a balanced 2×2 factorial design. *Stat Methods Med Res* 2015; **24**: 574–593.
14. Candel MJJM and Van Breukelen GJP. Sample size calculation for treatment effects in randomized trials with fixed cluster sizes and heterogeneous intraclass correlations and variances. *Stat Methods Med Res* 2015; **24**: 557–573.
15. Lemme F, Van Breukelen GJP and Berger MPF. Efficient treatment allocation in two-way nested designs. *Stat Methods Med Res* 2015; **24**: 494–512.
16. Moerbeek M. Optimal designs for group randomized trials and group administered treatments with outcomes at the subject and group level. *Stat Methods Med Res*. Epub ahead of print 2020; **29**: 797–810. DOI: 10.1177/0962280219846149.
17. Heo M, Litwin AH, Blackstock O, et al. Sample size determinations for group-based randomized clinical trials with different levels of data hierarchy between experimental and control arms. *Stat Methods Med Res* 2017; **26**: 399–413.
18. Moerbeek M and Wong WK. Sample size formulae for trials comparing group and individual treatments in a multilevel model. *Stat Med* 2008; **27**: 2850–2864.
19. Roberts C and Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clin Trials* 2005; **2**: 152–162.
20. Torgerson DJ and Torgerson CJ. Unequal randomisation. In: Torgerson DJ and Torgerson CJ (eds) *Designing randomised trials in health, education and the social sciences*. London: Palgrave Macmillan, 2008, pp.108–113.
21. Kosorok MR and Moodie EEM. *Adaptive treatment strategies in practice: Planning trials and analyzing data for personalized medicine*. Philadelphia: SIAM, 2016.
22. Antognini AB and Giovagnoli A. *Adaptive designs for sequential treatment allocation*. Boca Raton: CRC Press, 2015.
23. Chakraborty B and Moodie EEM. *Statistical methods for dynamic treatment regimes: Reinforcement learning, causal inference, and personalized medicine*. New York: Springer, 2013.
24. Tsiatis AA, Davidian M, Holloway ST, et al. *Dynamic treatment regimes: Statistical methods for precision medicine*. Boca Raton: Chapman and Hall/CRC, 2019.
25. Almirall D, Compton SN, Gunlicks-Stoessel M, et al. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Stat Med* 2012; **31**: 1887–1902.
26. Naar-King S, Ellis DA, Idalski Carcone A, et al. Sequential multiple assignment randomized trial (SMART) to construct weight loss interventions for African American adolescents. *J Clin Child Adolesc Psychol* 2016; **45**: 428–441.
27. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat Med* 2005; **24**: 1455–1481.
28. Seewald NJ, Kidwell KM, Nahum-Shani I, et al. Sample size considerations for comparing dynamic treatment regimens in a sequential multiple-assignment randomized trial with a continuous longitudinal outcome. *Stat Methods Med Res* 2020; **29**: 1891–1912.
29. Kidwell KM, Seewald NJ, Tran Q, et al. Design and analysis considerations for comparing dynamic treatment regimens with binary outcomes from sequential multiple assignment randomized trials. *J Appl Stat* 2018; **45**: 1628–1651.

30. Nahum-Shani I, Wu T, Artman WJ, et al. Power analysis in a SMART design: Sample size estimation for determining the best embedded dynamic treatment regime. *Biostatistics* 2018; **21**: 1–17.

31. Ghosh P, Cheung YK and Chakraborty B. Sample size calculations for clustered SMART designs. In: Kosorok MR and Moodie EEM (eds) *Adaptive treatment strategies in practice*. Alexandria, VSA: Society for Industrial and Applied Mathematics, 2015, pp. 55–70.

32. Necamp T, Kilbourne A and Almirall D. Comparing cluster-level dynamic treatment regimens using sequential, multiple assignment, randomized trials: Regression estimation and sample size considerations. *Stat Methods Med Res* 2017; **26**: 1572–1589.

33. Oetting A, Levy J, Weiss R, et al. Statistical methodology for a SMART design in the development of adaptive treatment strategies. In: Shrout P, Keyes K and Ornstein K (eds) *Causality and psychopathology: Finding the determinants of disorders and their cures*. Arlington, VA: Oxford University Press, 2010, pp. 734–763.

34. Artman WJ, Nahum-Shani I, Wu T, et al. Power analysis in a SMART design: Sample size estimation for determining the best embedded dynamic treatment regime. *Biostatistics* 2020; **21**: 432–448.

35. Cook RD and Wong WK. On the equivalence of constrained and compound optimal designs. *J Am Stat Assoc* 1994; **89**: 687.

36. Ertefaie A, Shortreed S and Chakraborty B. Q-learning residual analysis: Application to the effectiveness of sequences of antipsychotic medications for patients with schizophrenia. *Stat Med* 2016; **35**: 2221–2234.

37. Pfammatter AF, Nahum-Shani I, DeZelar M, et al. SMART: Study protocol for a sequential multiple assignment randomized controlled trial to optimize weight loss management. *Contemp Clin Trials* 2019; **82**: 36–45.

38. Sherwood NE, Butryn ML, Forman EM, et al. The BestFIT trial: A SMART approach to developing individualized weight loss treatments. *Contemp Clin Trials* 2016; **47**: 209–216.

39. Schmitz JM, Stotts AL, Vujanovic AA, et al. A sequential multiple assignment randomized trial for cocaine cessation and relapse prevention: Tailoring treatment to the individual. *Contemp Clin Trials* 2018; **65**: 109–115.

40. Nahum-Shani I, Almirall D, Yap JRT, et al. SMART Longitudinal analysis: A tutorial for using repeated outcome measures from SMART studies to compare adaptive interventions. *Psychol Methods*. Epub ahead of print 2020; **25**: 1–29. DOI: 10.1037/met0000219.

41. Kidwell KM. SMART Designs in cancer research: Past, present, and future. *Clin Trials* 2014; **11**: 445–456.

42. Sikorskii A, Wyatt G, Lehto R, et al. Using SMART design to improve symptom management among cancer patients: A study protocol. *Res Nurs Heal* 2017; **40**: 501–511.

43. Gunlicks-Stoessel M, Mufson L, Westervelt A, et al. A pilot SMART for developing an adaptive treatment strategy for adolescent depression. *J Clin Child Adolesc Psychol* 2016; **45**: 480–494.

44. August GJ, Piehler TF and Bloomquist ML. Being "SMART" about adolescent conduct problems prevention: Executing a SMART pilot study in a juvenile diversion agency. *J Clin Child Adolesc Psychol* 2016; **45**: 495–509.

45. Pistorello J, Jobes DA, Compton SN, et al. Developing adaptive treatment strategies to address suicidal risk in college students: A pilot sequential, multiple assignment, randomized trial (SMART). *Arch Suicide Res* 2018; **22**: 644–664.

46. Pelham WE, Fabiano GA, Waxmonsky JG, et al. Treatment sequencing for childhood ADHD: A multiple-randomization study of adaptive medication and behavioral interventions. *J Clin Child Adolesc Psychol* 2016; **45**: 396–415.

47. Fatori D, de Bragança Pereira CA, Asbahr FR, et al. Adaptive treatment strategies for children and adolescents with obsessive-compulsive disorder: A sequential multiple assignment randomized trial. *J Anxiety Disord* 2018; **58**: 42–50.

48. Flynn D, Eaton LH, Langford DJ, et al. A SMART design to determine the optimal treatment of chronic pain among military personnel. *Contemp Clin Trials* 2018; **73**: 68–74.

49. Cheng Q and Yang M. On multiple-objective optimal designs. *J Stat Plan Inference* 2019; **200**: 87–101.

50. Ouwens MJNM, Tan FES and Berger MPF. Maximin D-optimal designs for longitudinal mixed effects models. *Biometrics*. Epub ahead of print 2002; **58**: 735–741. DOI: 10.1111/j.0006-341X.2002.00735.x.

51. Cohen J. A power primer. *Psychol Bull*. 1992; **112**: 155–159.

52. Chang W, Cheng J, Allaire J, et al. Shiny: Web application framework for R. 2016. https://cran.r-project.org/package=shiny.

53. Verheggen RJHM, Maessen MFH, Green DJ, et al. A systematic review and meta-analysis on the effects of exercise training versus hypocaloric diet: Distinct effects on body weight and visceral adipose tissue. *Obes Rev* 2016; **17**: 664–690.

54. Kilbourne AM, Smith SN, Choi SY, et al. Adaptive School-based Implementation of CBT (ASIC): Clustered-SMART for building an optimized adaptive implementation intervention to improve uptake of mental health interventions in schools. *Implement Sci* 2018; **13**: 1–15.

55. Kilbourne AM, Almirall D, Eisenberg D, et al. Protocol: Adaptive implementation of effective programs trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. *Implement Sci* 2014; **9**: 1–14.

56. Moerbeek M and Teerenstra T. *Power analysis of trials with multilevel data*. Boca Raton: CRC Press, 2016.

57. Evans SR, Follmann D, Liu Y, et al. Sequential, multiple-assignment, randomized trials for comparing personalized antibiotic strategies (SMART-COMPASS). *Clin Infect Dis* 2019; **68**: 1961–1967.

58. Seewald NJ, Kidwell KM, Nahum-Shani I, et al. Sample size considerations for comparing dynamic treatment regimens in a sequential multiple-assignment randomized trial with a continuous longitudinal outcome. *Stat Methods Med Res*. Epub ahead of print 2020. **29**: 1891-1912.. DOI: 10.1177/0962280219877520.

59. Penn State College of Health and Human Development. Projects using SMARTs. https://www.methodology.psu.edu/ra/adap-inter/projects/ (accessed 9 June 2020).