Received: 19 May 2022 | Revised: 30 May 2024 | Accepted: 17 June 2024

DOI: 10.1002/sim.10161

RESEARCH ARTICLE

Statistics in Medicine WILEY

A sequential, multiple assignment, randomized trial design with a tailoring function

Holly Hartman[1] | **Matthew Schipper**[2] | **Kelley Kidwell**[2]

[1]Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA

[2]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Holly Hartman, Population and Quantitative Health Sciences, Case Western Reserve University, 2210 Circle Drive, Cleveland, OH 44106-7164, USA.
Email: heh39@case.edu

Funding information
Patient-Centered Outcomes Research Institute, Grant/Award Number: ME-1507-31108; National Cancer Institute, Grant/Award Number: P30-CA046592

We present a trial design for sequential multiple assignment randomized trials (SMARTs) that use a tailoring function instead of a binary tailoring variable allowing for simultaneous development of the tailoring variable and estimation of dynamic treatment regimens (DTRs). We apply methods for developing DTRs from observational data: tree-based regression learning and Q-learning. We compare this to a balanced randomized SMART with equal re-randomization probabilities and a typical SMART design where re-randomization depends on a binary tailoring variable and DTRs are analyzed with weighted and replicated regression. This project addresses a gap in clinical trial methodology by presenting SMARTs where second stage treatment is based on a continuous outcome removing the need for a binary tailoring variable. We demonstrate that data from a SMART using a tailoring function can be used to efficiently estimate DTRs and is more flexible under varying scenarios than a SMART using a tailoring variable.

KEYWORDS
clinical trials, dynamic treatment regimens, Q-learning, SMARTs, tailoring function, tailoring variable, tree based reinforcement learning

1 | INTRODUCTION

Dynamic treatment regimens (DTRs) are a sequence of decision rules that provide guidelines for treatments based on a patient's specific characteristics and response to treatment.[1] These treatment sequences are a form of personalized medicine where patients with the same condition may receive different treatments based on the evolution of their condition, their patient characteristics, and/or their response to prior treatments. A decision rule tailors treatment at each time point for each patient defined by a tailoring variable.[2,3] An example of a two-stage DTR is "Start all patients on treatment A. After 6 months, if the patient responds, maintain treatment A. Otherwise, give the patient treatment B." This DTR can be written as the treatment triplet (A, A, B) to indicate the first treatment, the second stage treatment for those who respond to the initial treatment, and the second stage treatment for those who do not respond. The tailoring variable here is response where response would have a predetermined clinical definition.

One way to develop DTRs is through observational studies[4-6] using rich treatment history data from electronic health records or disease registries. In observational studies, the second stage treatment is often chosen based on multiple factors and not a single binary variable as presented in the example above (ie, response/non-response to treatment).

Abbreviations: BR-SMART, balanced randomized SMART; DTR, dynamic treatment regimen; GEE, generalized estimating equations; IPTW, inverse probability treatment weighting; SMART, sequential, multiple assignment, randomized trials; TBRL, tree-based reinforcement learning; TBRL-IPW, tree-based reinforcement learning with inverse propensity weighting; TBRL-RF, tree-based reinforcement learning with random forest; TF-SMART, SMART with a tailoring function; TV-SMART, SMART with a tailoring variable; WRR, weighted and replicated regression.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Author(s). *Statistics in Medicine* published by John Wiley & Sons Ltd.

Thus, methods to estimate DTRs from observational data must determine the categorical variable(s) that dictate which treatment is optimal at each time point.

Another way to develop DTRs is through randomized trials, most often SMARTs. SMARTs integrate a tailoring variable into the design to assign a second stage treatment based on the patient's response to the previous treatment.[7] In SMARTs, DTRs are embedded into the design so that the effects of each DTR can be estimated from the resulting data. In a two-stage SMART, participants are randomly assigned to a first treatment and a first stage outcome or tailoring variable is measured. The second treatment is assigned probabilistically or deterministically based on the design and the selected tailoring variable. The final outcome is measured at the end of the second stage.

The tailoring variable in a SMART is often based on response or non-response to treatment, thus requiring that the trial protocol specify a binary rule for what qualifies as a "response." While there is some guidance in selecting the tailoring variable,[3] such as implementing a pilot study, there is a lack of guidance on how to proceed if a reasonable tailoring variable cannot be established. In many studies, the outcome of interest may not be a binary variable and selecting a tailoring variable may not be an easy or practical task.

Investigators may consider a balanced randomization, or unrestricted SMART.[8] This design re-randomizes all participants to subsequent treatment. The data can then be analyzed with subcovariates to define a tailoring variable and DTRs.

If, however, investigators have a continuous intermediate outcome of interest and want to allow for different treatment options for those doing well on the initial treatment versus those who are not, neither design is fully appropriate. A traditional SMART using a predefined tailoring variable (TV-SMART) requires dichotomization of the intermediate continuous variable and a balance randomized SMART (BR-SMART) does not take into consideration the patient's response to the initial treatment.

Thus, we propose a SMART design where the second stage treatment is assigned based on a tailoring function. A tailoring function, similar to a mapping function in Hartman et al,[9] takes a continuous outcome at the end of the first stage and assigns a probability of receiving each possible second stage treatment. In the design presented here, we assume a binary tailoring variable is not available, and thus data analysis must identify the tailoring variable and estimate the effects of DTRs. This trial design may be most useful in a Phase II setting where the goal is to identify the best DTR to use in a larger Phase III trial. This project addresses a gap in clinical trial methodology by presenting a SMART design where second stage treatment is based on a continuous outcome.

In Section 2, we present the proposed SMART design, a a SMART which uses a tailoring function (TF-SMART), with a goal of estimating DTRs in flexible manner without needing to pre-specify a tailoring variable. Next, in Section 3, we discuss several methods for developing DTRs from observational studies in the context of a SMART thereby eliminating the need for a pilot study to select a definition of response and non-response. We then present simulation designs to compare our new TF-SMART design to a TV-SMART and a BR-SMART in Section 4. Lastly, in Section 5, we show results from these simulations to compare estimated DTRs and patient outcomes. This new TF-SMART design simultaneously defines decision rules and the appropriate sequence of treatments within the context of randomized trial which is a novel approach to using a SMART design to develop DTRs.

2 | TRIAL DESIGNS

2.1 | SMART notation

In a SMART design, patients are randomly assigned to one first stage treatment, typically done so that there is an equal number of patients in all first stage treatment arms. Assume there are s possible first stage treatments and t possible second stage treatments. Then all possible first stage treatments are $\mathbf{A_1} = \{a_{11}, \ldots, a_{1s}\}$ and all possible second stage treatments are $\mathbf{A_2} = \{a_{21}, \ldots, a_{2t}\}$. Then the first and second treatments a patient receives (A_1 and A_2 respectively) are such that $A_1 \in \mathbf{A_1}$ and $A_2 \in \mathbf{A_2}$. Each treatment $a_{11}, \ldots, a_{1s}$ has its own mean treatment effect on the first stage outcome, Y_1.

The average second stage outcome, Y_2, is dependent on the first stage outcome (Y_1) and the first and second treatments. Assuming a linear relationship between Y_1 and Y_2 given A_1 and A_2, the mean of Y_2, conditional on Y_1, A_1, and A_2 would be:

$$\mu(A_2, Y_1, A_1) = \sum_{i=1}^{s}\sum_{j=1}^{t}\beta_{a_{1i}a_{2j}} I(A_1 = a_{1i})I(A_2 = a_{2i}) + \sum_{i=1}^{s}\sum_{j=1}^{t}\gamma_{a_{1i}a_{2j}} I(A_1 = a_{1i})I(A_2 = a_{2i})Y_1 \quad (1)$$

where the βs are the intercept of the (Y_1, Y_2) line and the γs are the slope.

2.2 | Proposed designs

We examine two TF-SMART designs presented in Figure 1. In these designs, patients are first randomized with equal probability to one of two treatments. Then, a continuous outcome is measured at the end of the first stage, Y_1. The second stage treatment is assigned using a tailoring function (more details in Section 2.3). The tailoring function takes Y_1 as input and outputs a probability. The output probability is then used to assign probabilities for each second stage treatment for the patient. In Figure 1, we present one possible option for treatments in the second stage. However, these designs could accommodate a variety of treatments. Second stage treatment options may include continuing the same treatment, alternate treatments, a change in dosage, a combination of treatments, or maintenance therapy. More general designs with distinct second stage treatments are shown in Web Figure 1.

In Design I (Figure 1A), there are two first stage treatments, A and B, and two second stage treatments options for each arm, continue or add. At the end of the first stage, a patient is randomly assigned to one of the two second stage treatments based on Y_1. If a tailoring variable is used in place of a tailoring function, patients are assigned to either continue or add based on their response to the first stage treatment and there is not a second randomization point. That is, $f(Y_1) = 1$ or 0 so the second stage treatments are assigned deterministically based on Y_1. Thus, if a tailoring variable is used, this design

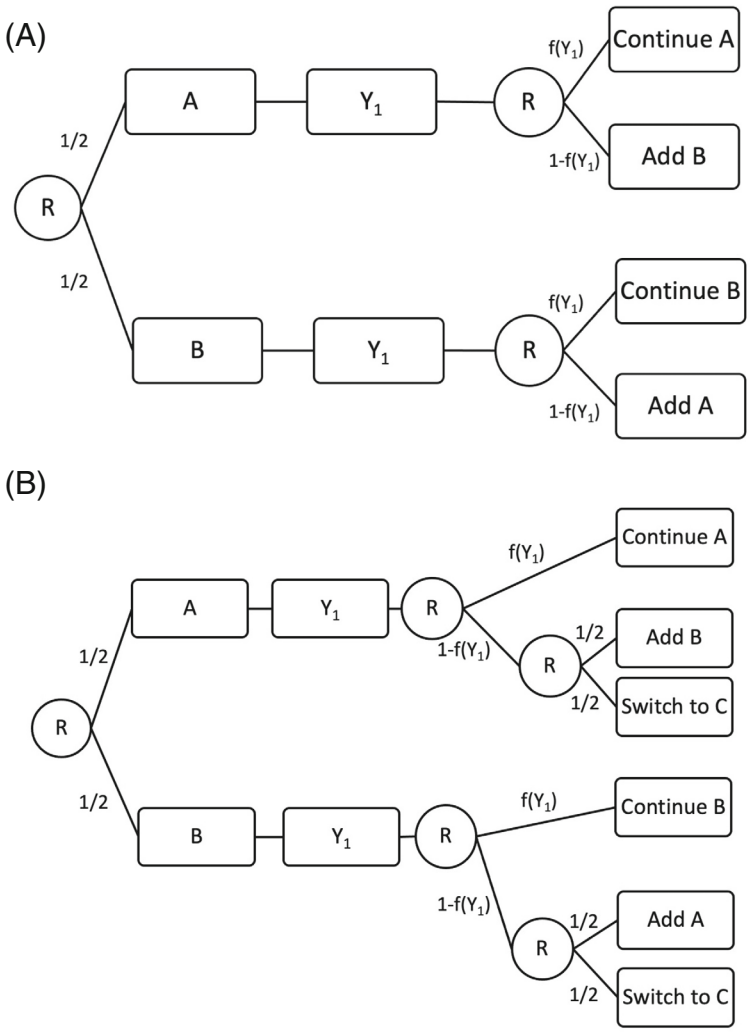


FIGURE 1 A, B, and C are treatments, Y_1 is the first stage outcome, R indicates a randomization point, $f(Y_1)$ is the tailoring function, and the expressions along the lines indicates the randomization probability of the treatment following. A. Design I: Patients with more favorable responses are more likely to continue the same treatment and patients with less favorable responses add to the other treatment. B. Design II: Patients with more favorable responses are more likely to continue the same treatment and patients with less favorable responses are randomized to either add the other first stage treatment or switch to a new treatment.

is not a SMART as there is only one randomization point. However, there are two DTRs embedded in the design so the design can still be analyzed using standard analytic methods for SMARTs by adjusting the weights to reflect the single randomization point.

In Design II (Figure 1B), patients are assigned a second stage treatment in two steps. First, patients are randomly assigned to either continue the same treatment or to the two options of of either add a treatment or switch treatments. If patients are assigned to the two treatment arms (add or switch), then the treatment is randomly selected from the two options with equal probability. This design is a SMART regardless of if a tailoring variable or tailoring function is used as there are two randomization points in sequence in both cases.

2.3 | Tailoring function details

Tailoring functions are a generalization of tailoring variables and are mechanistically the same as mapping functions presented in Hartman et al.[9] However, interest in this paper is in developing DTRs in a larger size trial which differs from previous work where the goal was to estimate the best first stage treatment in a small sample. A tailoring function takes a continuous first stage outcome, Y_1, and maps it to randomization probabilities of each second stage treatment. For designs such as those in Figure 1, this function can map better outcomes to values closer to 1 so that patients are more likely to stay on the treatment that worked well for them. Conversely, patients that are not doing well on the first stage treatment are more likely to switch to a new treatment or add an additional treatment with the goal of improving the second stage outcome.

There are many options for selecting a tailoring function and any cumulative distribution function could be used. More guidance is provided in Hartman et al,[9] which we briefly summarize. To maximize statistical efficiency, it is ideal to aim for equal numbers of patients in each arm. A function may also be chosen in a more patient-centered manner to optimize patient outcomes within the trial. However, the tailoring function should be selected so that at the conclusion of the study, there are patients in each arm of the study to ensure the treatment effects of all pathways are observed, all treatment pathway effects are estimable so that the positivity assumption is satisfied.

Assume higher values of the outcome Y_1 indicate better outcomes. That is, higher values of Y_1 are mapped closer to 1. One simple option is to use a tailoring function that is a linear function between the minimum and maximum values of Y_1, $f(Y_1) = (Y_1 - Y_{min})/(Y_{max} - Y_{min})$. This linear function can be modified to implement thresholds to limit extreme circumstances by setting Y_{max} and Y_{min} to reflect safety and ethical considerations. Additionally, the linear tailoring function can be modified using a power function, $f(Y_1)^k$, to have different properties. For $k > 1$, more people will switch treatments than for $k = 1$. This may be helpful if the outcome is left skewed or if having more people switch is of interest. For $0 < k < 1$, fewer people will switch treatments which may be useful if the outcome is right skewed. For more guidance in selecting a tailoring function and examples, refer to Hartman et al.[9]

Tailoring variables can be thought of as a special case of tailoring functions where $f(Y_1) = 0$ for $Y_1 < x$ and $f(Y_1) = 1$ for $Y_1 \geq x$ for some x which is a pre-selected cut off. Using a tailoring variable defined as a step-wise function in this manner generates a TV-SMART design with a binary tailoring variable that assigns second stage treatment.

Further, a design where $f(Y_1) = p$ for some constant p, $0 \leq p \leq 1$, can be considered a special case of the tailoring function. This design is a BR-SMART when $p = 1/K_2$ where K_2 is the number of second stage treatments. BR-SMARTs do not incorporate the first stage outcome and randomly assign the second stage treatment with all treatments having equal probability. Since treatment assignment is not based on first stage outcomes, first and second stage treatments can be assigned at the baseline randomization.

3 | DTR ESTIMATION METHODS

In estimating the effects of potential DTRs, we want to maximize the expected second stage outcome when applied to the target population. In a traditional SMART design with a binary intermediate outcome, weighted and replicated regression (WRR) with generalized estimating equations (GEE) is commonly used to analyze data and estimate the embedded DTRs.[7,10-12] Traditional SMART designs and analytic methods assume a predefined tailoring variable exists. Q-learning may also be used to analyze data from a SMART to develop more personalized DTRs that include baseline and time-varying covariates,[11] but the tailoring variable is set.

Published methods for DTR estimation from observational data include Q-learning,[5,11,13] tree-based reinforcement learning (TBRL),[6,14] Bayesian machine learning,[15] and outcome weighted learning (OWL).[16] Here, we present WRR, Q-learning, and TBRL to estimate DTR effects in the context of the TF-SMART designs shown in Figure 1 and compare their results to DTR effects estimated from a TV-SMART using WRR.

Let Y_1 and Y_2 be the first and second stage outcomes, respectively. Let A_1 and A_2 be the first and second stage treatment assignments, respectively. The goal of any DTR analysis is to identify the DTR which maximizes the expected second stage outcome, Y_2, for patients in the target population.

3.1 | Weights

For Q-learning, tree-based reinforcement learning with inverse propensity weighting (TBRL-IPW), and WRR we use inverse probability of treatment weighting (IPTW). Weights for the first stage analyses in both Designs I and II (Figure 1) are $w_1 = \frac{1}{2}$ due to the equal stage randomization to the two first stage treatments. Since all participants are weighted equally, we do not need to weight at this stage. The weights for participants vary in the second stage and are defined for Design I as:

$$w_2 = \frac{2}{f(y_1)} I(A_2 \in (\text{Continue } A, \text{Continue } B)) + \frac{2}{(1-f(y_1))} I(A_2 \in (\text{Add } B, \text{Add } A)) \tag{2}$$

and Design II as:

$$w_2 = \frac{2}{f(y_1)} I(A_2 \in (\text{Continue } A, \text{Continue } B)) + \frac{4}{(1-f(y_1))} I(A_2 \in (\text{Add } B, \text{Switch to } C, \text{Add } A, \text{Switch to } C)). \tag{3}$$

3.2 | Weighted and replicated regression with generalized estimating equations

The WRR method to analyze SMART data weights observations based on their probability of receiving the treatment regimen they received using IPTW. In a traditional SMART, there are multiple DTRs embedded within the trial design. Some people receive a sequence of treatments that is a part of more than one DTR. The data from these patients are replicated and each replicate is assigned to one of the possible DTRs they could be following. This replication allows standard software to simultaneously estimate the effects of all embedded DTRs using one regression model. A fully saturated model is often used for analysis.[17]

Due to BR-SMARTs and our proposed TF-SMART design not having pre-specified decision rules for defining a binary tailoring variable, this method cannot be used to estimate DTRs for the TF-SMART or BR-SMART design. The analytic methods we use for our new TF-SMART design and the BR-SMART design must both develop the tailoring variable and define the treatment regimen, whereas WRR assumes a predetermined tailoring variable and only defines the treatment regimen. For a TV-SMART design, we do not need to estimate the cut point as it is predefined and embedded in the design. Thus, we use this analytic method with various cutoffs of the continuous tailoring variable to analyze data from TV-SMARTs and compare results with TF-SMARTs and BR-SMARTs using other methods.

For Design I, we estimate the two predefined DTRs that either start with treatment A or start with treatment B. Since patients are randomized to either A or B (and are thus randomized to one of the two predefined DTRs) we compare only the second stage outcomes based on first stage treatment so the model is:

$$Y_2 \sim \beta_1 I(A_1 = A) + \beta_2 I(A_1 = B)$$

For Design II, we must select a second stage treatment for people whose Y_1 values fall below the pre-defined threshold. To do this, we use the weighting and replication method described in Nahum-Shani et al.[17] Here, we examine four possible DTRs: (A, A, A+B), (A, A, C), (B, B, B+A), and (B, B, C). Thus, for participants that receive treatment A in the second stage, they could be following either of the first two DTRs, and for participants that receive treatment B in the second stage, they could be following either of the second two DTRs. Thus we replicate those participants, one replicate for each DTR they could be following, to obtain an estimate of the second stage outcome for each DTR. To account for imbalance

by design, we weight participants using the weights in Equation (3). Let the selected cut off for the tailoring variable be x. Then $f(Y_1)$ in a TV-SMART is defined as:

$$f(Y_1) = \begin{cases} 0 & Y_1 \leq x \\ 1 & Y_1 > x. \end{cases}$$

We then use linear regression with IPTW to model:

$$Y_2 \sim \beta_1 I(DTR = (A, A, A+B)) + \beta_2 I(DTR = (A, A, C)) + \beta_3 I(DTR = (B, B, B+A)) + \beta_4 I(DTR = (B, B, C))$$

The best performing DTR is then selected based on which has the highest estimated β, $\hat{\beta}$.

3.3 | Q-learning

Q-learning is a statistical approach to estimating a DTR that maximizes an outcome based on a sequence of decisions.[11] First, the optimal decision rule at the second stage is determined by identifying A_2 that maximizes $Q_2(A_1, A_2, Y_1) = E(Y_2|A_1, A_2, Y_1)$. Then, the optimal first stage decision rule is constructed by selecting A_1 that maximizes $Q_1(A_1) = E\big(\max_{A_2} Q_2(A_1, A_2, Y_1)\big)$. The quantity $E\big(\max_{A_2} Q_2(A_1, A_2, Y_1)\big)$ is called a pseudo-outcome. This represents the expected second stage outcome if the patient received the best second stage treatment. Q-learning can include covariates measured at baseline before the first stage (X_1) and up to the second stage (X_2) so that $Q_2(A_1, A_2, Y_1, X_1, X_2) = E(Y_2|A_1, A_2, Y_1, X_1, X_2)$ and $Q_1(A_1, X_1) = E\big(\max_{A_2} Q_2(A_1, A_2, X_1, X_2, Y_1)\big)$. Here, we will not use covariates and thus ignore X_1 and X_2 for now. Q_1 and Q_2 are called Q-functions. While Q-learning can be done with a variety of techniques and models, here we focus on linear regression so that Q_2 is modeled as a linear combination of A_1, A_2, and Y_1 and Q_1 is modeled as a linear combination of A_1. In order to account for the randomization scheme, we use IPTW.[5]

To implement Q-learning in our design settings, we use linear models to identify the optimal treatment at each randomization point in the specific design. For Design I, this means we separately identify which is the best second stage treatment for patients that received treatment A in the first stage (add B or switch to C) and the best second stage treatment for patients that received treatment B in the first stage (add A or switch to C). In Design II, we identify the best second stage treatments for those that received A in the first stage (continue A, add B, or switch to C) and those that received B in the first stage (continue B, add A, or switch to C). Lastly, in both designs, we identify which is the best treatment to start with to maximize the second stage outcome (A or B). Let β_{ij} and γ_{ij} be the model parameters for when $A_1 = i$ and $A_2 = j$. The model to identify the best second stage treatment for those starting with treatment A in Design I is:

$$Y_2 \sim \beta_{AA} I(A_2 = A) + \beta_{AB} I(A_2 = A+B) + \gamma_{AA} Y_1 I(A_2 = A) + \gamma_{AB} Y_1 I(A_2 = A+B) \tag{4}$$

Likewise, the model for identifying the best second stage treatment for those starting with treatment B is:

$$Y_2 \sim \beta_{BB} I(A_2 = B) + \beta_{BA} I(A_2 = B+A) + \gamma_{BB} Y_1 I(A_2 = B) + \gamma_{BA} Y_1 I(A_2 = B+A) \tag{5}$$

The equivalent models for Design II are:

$$\begin{aligned} Y_2 \sim\ & \beta_{AA} I(A_2 = A) + \beta_{AB} I(A_2 = A+B) + \beta_{AC} I(A_2 = C) \\ & + \gamma_{AA} Y_1 I(A_2 = A) + \gamma_{AB} Y_1 I(A_2 = A+B) + \gamma_{AC} Y_1 I(A_2 = C) \end{aligned} \tag{6}$$

and

$$\begin{aligned} Y_2 \sim\ & \beta_{BB} I(A_2 = B) + \beta_{BA} I(A_2 = B+A) + \beta_{BC} I(A_2 = C) \\ & + \gamma_{BB} Y_1 I(A_2 = B) + \gamma_{BA} Y_1 I(A_2 = B+A) + \gamma_{BC} Y_1 I(A_2 = C). \end{aligned} \tag{7}$$

We use the linear regression with the above models and IPTW with the appropriate weights from Equations (2) or (3) to obtain estimates for all βs and γs.

Let $\hat{Y}_2$ indicate the predicted second stage outcome based on the parameter estimates from the models above. Then the pseudo outcome to analyze the stage 1 data for Design I is:

$$\tilde{Y}_2 = \max_{A_2 \in A, A+B} \hat{Y}_2 I(A_1 = A) + \max_{A_2 \in B, B+A} \hat{Y}_2 I(A_1 = B) \tag{8}$$

and for Design II is:

$$\tilde{Y}_2 = \max_{A_2 \in A, A+B, C} \hat{Y}_2 I(A_1 = A) + \max_{A_2 \in B, B+A, C} \hat{Y}_2 I(A_1 = B). \tag{9}$$

These equations identify the maximal second stage outcome (Y_2) for a given first stage treatment (A_1) across all possible second stage treatment (A_2). That is, the pseudo outcome is the expected second stage outcome assuming the patient receives the optimal second stage outcome which maximizes the second stage outcome.

We then model to estimate the best first stage treatment:

$$\tilde{Y}_2 \sim \beta_A I(A_1 = A) + \beta_B I(A_1 = B).$$

The first stage treatment with the largest $\hat{\beta}$ is then selected.

3.4 | Tree-based reinforcement learning

We follow methods from Tao et al[6] and Laber and Zhao[14] for TBRL in the TF-SMART setting. As with Q-learning, we start by identifying the optimal second stage treatment and then use pseudo-outcomes to identify the optimal first stage treatment. We do this both with linear regression as in Q-learning and with a random forest using the same predictors.

With TBRL, a tree is built to create decision rules for the second stage treatment assignment that maximizes the second stage outcome. Then, a tree is built for the first stage treatment assignment to maximize the expected second stage outcome using pseudo-outcomes.

Decision trees start with all patients and use binary rules to create groups of patients that should all receive the same treatment. An example is shown in Web Figure 2. Nodes are groups of patients that meet specific criteria that will either be split into smaller nodes or all receive the same treatment. Nodes are terminal when no additional split follows. The patients that fall in terminal nodes all receive the same treatment as other patients in that same node.

To build a tree, we first define a node purity measure which represents how well a tree has classified the groups and is a measure of homogeneity within a node. We use the purity measure described in Laber and Zhao:[14]

$$\mathcal{P}(r, \mathcal{R}) = \max_{a,a' \in \mathcal{A}} E_n \left[\frac{[Y - \hat{m}(X)] I(X \in \mathcal{R}) I(A = \pi_{r,a,a'}(X))}{P(\pi_{r,a,a'}(X)|X)} \right] \left(E_n \left[\frac{I(X \in \mathcal{R}) I(A = \pi_{r,a,a'}(X))}{P(\pi_{r,a,a'}(X)|X)} \right] \right)^{-1} \tag{10}$$

where A is the treatment, $\mathcal{R}$ is the node to be split into groups r and r^C, X are patient characteristics for that specific stage, $\pi_{r,a,a'}$ is the decision rule that assigns treatment a to subjects in group r and treatment a' to subjects in r^C, and $\hat{m}(X)$ is $\max_{a \in \mathcal{A}} E(Y|A = a, X)$. E_n is the empirical expectation. Then $P(\pi_{r,a,a'}(X)|X)$ is the propensity score for the patient defined by the inverse of the probability of receiving the treatment that they received. This purity measure is the percent change in the expected outcome when the patients in a given node are split (ie, patients in r receive treatment a and those in r^C receive a') relative to the expected outcome if patients in that node are not split (ie, all patients in the node receive the same treatment).

One distinct difference in this TF-SMART setting versus the observational data setting is that we do not need to estimate propensity scores since the probability of each treatment is a known function of the first stage outcome via the tailoring function and randomization probability.

With the node purity measure, we then define splitting rules. Each split must only depend on one covariate. In the scenario without covariates, this means that there will be at most one split which is solely based on the first stage outcome. If there are other patient covariates that could be considered, then the split is based on one out of all covariates measured up until randomization. The split is chosen to optimize the purity metric defined in Equation (10). Node size is defined as the number of patients in that node. Define n_0 as the minimum node size accepted where n_0 is a positive integer. Define

λ as the minimum improved node purity to warrant a split where λ is a positive constant. Then we define our splitting rules as follows:[14]

Rule 1 Node size is larger than $2n_0$.
Rule 2 At least one possible split results in all child nodes having more than n_0 items.
Rule 3 The tree-depth (or number of splits) has not reached a pre-specified limit.

If these rules are met, then the node is split by selecting r such that $\mathcal{P}(r, \mathcal{R})$ is maximized and the node size of r and r^C are both at least size n_0. Let $|\cdot|$ indicate node size. Then the optimal split is:

$$\hat{r}^{opt} = \arg\max_r [\mathcal{P}(r, \mathcal{R}) \,:\, \min(|r|, |r^C|) \geq n_0].$$

Once this optimal split has been determined, we apply the last rule:

Rule 4 If the purity improvement is less than λ (ie, $\mathcal{P}(r, \mathcal{R}) - \mathcal{P}(\mathcal{R}) < \lambda$), then do not split the node.

This splitting rule forces the expected outcome to improve by some set amount in order to warrant a split. We follow these splitting rules until all nodes are terminal (ie, no split follows all rules defined above). This results in an estimate of the optimal DTR which maximizes the expected second stage outcome. Since we do not include covariates, we only have one possible variable for splitting (Y_1) and thus we do not need to implement Rule 3 regarding maximum tree-depth.

Here, we use two methods for TBRL: TBRL-IPW and tree-based reinforcement learning with random forests TBRL-RF. For the TBRL-IPW method, we start similarly to Q-learning by using the appropriate weighted linear regression models from Equations (4) and (5) or (6) and (7). When using TBRL-RF, we use a random forest instead of a linear regression model, but proceed in the same manner. We calculate the 100-percentiles of the observed outcomes and test each as a potential node split points to determine which split maximizes the purity measure. The split point that maximizes the purity measure based on the parameter estimates is selected if it meets all splitting criteria. For tree building, we follow Tao et al[6] and use $n_0 = 5$ as our minimum node size. We set $\lambda = 0.01$ as our minimum node purity improvement so that there is at least a 1% improvement in the predicted second stage outcome by allowing a split.

As with the Q-learning method, we estimate the best first stage treatment by maximizing psuedo-outcome, $\tilde{Y}_2$ as defined in Equation (8) or (9). We select the first stage treatment with the largest $\hat{\beta}$.

4 | SIMULATION METHODS

Simulated data available on request from the authors. As this is a proposed design and has not been implemented, data from an actual clinical trial is not available.

4.1 | Generation of SMART data

Using the notation established in Section 2, the first stage outcomes are distributed:

$$Y_1|A_1 \sim N\left(\sum_{i=1}^{s} \alpha_i I(A_1 = a_{1i}), \sigma^2\right).$$

Since $s = 2$ (as there are two first stage treatments), $a_{11} = A$ and $a_{12} = B$. Then the distribution of Y_1 is:

$$Y_1|A_1 \sim N\left(\alpha_A I(A_1 = A) + \alpha_B I(A_1 = B), \sigma^2\right).$$

The second stage treatment is randomly assigned using a tailoring function based on the first stage outcome, Y_1 and the first stage treatment, A_1. Let $\boldsymbol{\pi}(Y_1, A_1)$ be the vector of probabilities of each second stage treatment as dictated by the trial design and the tailoring function. Then the second stage treatment assignment is:

$$A_2|Y_1, A_1 \sim Multinomial(\boldsymbol{\pi}(Y_1, A_1)).$$

The distribution of the final outcome Y_2, is dependent on all previous treatments and the first stage outcome. We assume a linear relationship between Y_1 and Y_2 given A_1 and A_2. Thus the distribution of Y_2 is:

$$Y_2|A_2, Y_1, A_1 \sim Normal(\mu(A_2, Y_1, A_1), \sigma^2)$$

where $\mu(A_2, Y_1, A_1)$ is defined in Equation (1).

This model is a general formulation. There are more parsimonious models that can be considered special cases of this model by imposing additional constraints on the α, γ, and/or β parameters. For simplicity here, we use the same σ^2 for all treatments and stages, but this could also vary.

4.2 | Analysis methods

We analyze data from the TV-SMARTs using WRR and we analyze data from the TF-SMARTs and BR-SMARTs using Q-learning, TBRL-IPW, and TBRL-RF.

4.3 | Design settings

We simulated 1,000 trials each with a total sample size of 300 patients. The participants are randomized to first stage treatments with equal probabilities, resulting in equal number of participants in each first stage treatment arms. Thus, there are 150 patients on each of the first stage treatments, A and B.

4.3.1 | Parameters for data generation models

For all designs and settings we set $\alpha_A = 50$, $\alpha_B = 30$, and all standard deviations to be $\sigma = 15$. We vary the second stage intercept, β, and slope, γ, parameters to create different scenarios (Web Tables 1 and 2). We then calculated the optimal DTR (Table 1).

TABLE 1 Optimal DTR for each scenario explored for designs I and II in the data generation model for the second stage mean outcome shown in Equation 1.

Scenario	Optimal DTR	Optimal DTR cut off for Y_1
I-i	(A, A, A+B)	60
I-ii	(A, A, A+B)	60
I-iii	(A, A, A+B)	60
I-iv	(A, A)	None
I-v	(A, A, A+B)	60
I-vi	(A, A)	None
II-i	(A, A, A+B)	60
II-ii	(A, A, A+B)	60
II-iii	(A, A, A+B)	60
II-iv	(A, A)	None
II-v	(A, A, A+B)	33.33
II-vi	(A, A, A+B, C)	60, 40

Note: (A, A) indicates that the optimal treatment sequence is A followed by continuing A for all patients. (A, A, A+B) indicates a DTR such that that the optimal first stage treatment is A followed by A for patient's who have an observed Y_1 greater than the cut off and A+B for patient's who have an observed Y_1 lower than the cut off. (A, A, A+B, C) indicates the optimal first stage treatment is A, the optimal second stage treatment if Y_1 is higher than the first cut off is A, the optimal second stage treatment if Y_1 is between the cut offs is A+B, and the optimal second stage treatment if Y_1 is lower than the final cut off is C. A, B, and C are all treatments. DTR = dynamic treatment regimen, Y_1 = first stage outcome. Scenario specifies the Design (I or II) and setting (i – vi).

Plots showing the expected second stage outcome conditional on A_1, A_2, and Y_1 for all scenarios are shown in Web Figure 3. We also briefly describe each scenario here. Scenarios I-i and II-i have the same cut off for the DTRs that start with either treatment A or treatment B. That is, regardless of if the treatment regimen starts with treatment A or B, the DTR has the same cut point based on the paramters in Web Tables 1 and 2. Scenarios I-ii and II-ii have different cut offs for the DTRs depending on which first stage treatment is selected. In Scenarios I-iii and II-iii, $E(Y_2|A_2, A_1 = A) = E(Y_2|A_2, A_1 = B)$ meaning there is no difference between second stage outcomes for treatments $A_1 = A$ and $A_1 = B$ conditional on Y_1. There is still a difference in first stage outcomes so treatment A is the optimal first stage treatment. In scenarios I-iv and II-iv, the cut point is not within a reasonable range of Y_1 (within $\pm 3\sigma$ of the true cut off) so a DTR is not needed. Thus the optimal treatment sequence is A followed by continuing A, denoted (A, A), for all patients. For scenarios I-v and II-v, the optimal DTR is not clear from only examining the second stage expected outcome, $\mu(A_2, Y_1, A_1)$. In Scenario I-vi, Y_1 is prognostic but not predictive, such that Y_1 is yields information about Y_2 but does not yield information about the association of A_2 with Y_2 conditional on A_1. In Scenario II-vi, the optimal DTR has two cut points, so there are three possible second stage treatment options as opposed to the traditional two second stage treatments. We denote this as, (A, A, A+B, C) to indicate the first stage treatment (A), treatment if Y_1 is higher than the first cut off (A), treatment if Y_1 is between the cut offs (A+B), and treatment if Y_1 is lower than the final cut off (C).

4.3.2 | Tailoring functions and tailoring variables

In our simulations, we consider three different tailoring functions. For all, if $Y_1 > 100, f(Y_1) = 1$ and if $Y_1 < 0, f(Y_1) = 0$. Then the three functions used are:

$$\text{TF 1: } f(Y_1) = Y_1/100$$
$$\text{TF 1/2: } f(Y_1) = (Y_1/100)^{1/2}$$
$$\text{TF 2: } f(Y_1) = (Y_1/100)^2$$

These tailoring functions reflect varying numbers of patients staying or switching treatments. We are interested in exploring the effects of having more or less patients in each treatment arm on the estimations. We chose our simulation settings to mimic a patient reported outcome that ranges from 0 to 100 so we chose to use modified linear functions with $Y_{min} = 0$ and $Y_{max} = 100$.

We also consider six different tailoring variables by assigning a cut point of the first stage outcome. If Y_1 is higher than the cut point, then $f(Y_1) = 1$ and if Y_1 is below the cut point then $f(Y_1) = 0$. The cut offs used in simulation are 30, 40, 50, 60, 70, 80 labelled as TV 30, TV 40, TV 50, TV 60, TV 70, and TV 80, respectively. These tailoring variables will also result in differing numbers of patients in each treatment arm.

Lastly, we examine a balanced randomization scheme where each of the second stage outcomes have equal probability. For Design I, this means that each second stage outcome has probability $1/2$, or $f(Y_1) = 1/2$. For Design II, each second stage outcome has probability $1/3$, or $f(Y_1) = 1/3$.

4.4 | Performance metrics

4.4.1 | Definition of the optimal DTR and optimal cut off

In order to compare results from the SMARTs to a defined standard, we use the optimal DTR defined for each scenario by the sets of treatments and cut offs listed in Table 1. The optimal DTR is the DTR with the highest expected second stage outcome. Since we present simulations with a known data generation model, we know the true optimal DTRs, but this is not true in practice. For example, for Scenario I-i, the optimal DTR is to start all patients on treatment A. Then if the patient's Y_1 is greater than 60, they continue to receive treatment A. If the patient's Y_1 is less than or equal to 60, they add treatment B (treatment denoted as A+B). In this scenario, the optimal cut off is 60. For scenarios that have only two treatments listed, (A, A), and "None" as the optimal DTR cut off for Y_1, the optimal sequence or fixed regimen is to have everyone receive treatment A and then continue to receive treatment A regardless of Y_1. There is no optimal cut off. For Scenario II-vi, the optimal DTR is to start everyone on treatment A. Then if $Y_1 > 60$, the patient continues A, if

$40 < Y_1 \leq 60$ the patient adds treatment B (A+B), and if $Y_1 \leq 40$ the patient receives C. This scenario has two optimal cut offs, 60 and 40. All optimal DTRs and associated optimal cut offs were calculated based on the known data generation parameters in Web Tables 1 and 2 as well as the known first stage treatment effects in Section 4.1.

4.4.2 | Definition of an estimated DTR

We estimate an optimal DTR with the highest second stage outcome for each simulated trial data set using Q-learning and TBRL for TF-SMARTs and WRR for TV-SMARTs and BR-SMARTs. For TF-SMARTs, the estimated DTR includes the treatments that make up the DTR and the estimation of the cut off for Y_1 (ie, the definition of the tailoring variable). For TV-SMARTs, the estimated DTR includes the treatments that make up the DTR and the predefined cut off for Y_1 which is the same as the tailoring variable used in the trial.

4.4.3 | Comparisons of SMART designs

For each design, scenario, and tailoring function/variable, we simulated 1,000 trials. We estimated the DTR from one trial using Q-learning, TBRL-IPW, and TBRL-RF for TF-SMARTs and BR-SMARTs and WRR for TV-SMARTs. We apply the resulting estimated DTR and apply it in a new population of 10,000 patients. We then average the second stage outcomes in the new population. We compare the average second stage outcomes, $\overline{Y}_2$, between TF-SMARTs, BR-SMARTs, and TV-SMARTs with higher being better. We also compare $\overline{Y}_2$ within TF-SMARTs for TF 1, TF 2, and TF 1/2 and within TV-SMARTs for TV 30, TV 40, TV 50, TV 60, TV 70, and TV 80 defined in Section 4.3.2. Additionally, we compare $\overline{Y}_2$ resulting from applying the estimated DTRs in the new population to the $\overline{Y}_2$ resulting from applying the optimal DTRs in the new population. We also examine the proportion of patients in a new population of 10,000 that are assigned the same treatments under the estimated DTRs as they would be under the optimal DTR. In addition to comparing $\overline{Y}_2$ in a new population, we also examine outcomes for patients within the trials. Lastly, we examine the proportion of trials that correctly selected treatment A as the optimal first stage treatment. In all scenarios, A is the optimal first stage treatment (Table 1).

4.4.4 | Comparison of analytic methods for TF-SMARTs

We compare Q-learning, TBRL-IPW, and TBRL-RF as analytic methods for TF-SMARTs. To do this we examine the same metrics as described in Section 4.4.3: (1) $\overline{Y}_2$, (2) proportion of patients on the same treatment under the estimated DTR as they would be under the optimal DTR, and (3) the proportion of trials that correctly selected treatment A as the optimal first stage treatment. For scenarios that had a single optimal cut off for Y_1 (Scenarios I-i, I-ii, I-iii, I-v, II-i, II-ii, II-iii, and II-v; Table 1), we also examined how well the methods estimated the optimal DTR cut off for Y_1. To do this we looked at bias defined as the difference between the estimated cut off and the optimal DTR cut off. For Scenarios I-i, I-ii, I-iii, I-v, II-i, II-ii, and II-iii, the optimal DTR cut off is 60. For Scenario II-v, the optimal DTR cut off is 33.33.

5 | RESULTS

Results for scenarios I-i, I-ii, and I-iii as well as results for scenarios II-i, II-ii, and II-iii are very similar. Thus, we present only results for I-i and II-i here (Figures 2 and 3) and results for all scenarios in the supplemental material (Web Figures 4 and 5).

5.1 | Comparison between randomization designs

The average second stage outcome, $\overline{Y}_2$, when applying the estimated DTRs from the TF-SMARTs to a population of 10,000 was fairly consistent between the tailoring functions and the balanced randomization (Figure 2). However, there was significant variation in $\overline{Y}_2$ when using a pre-selected cut off tailoring variable in the TV-SMARTs. For most scenarios

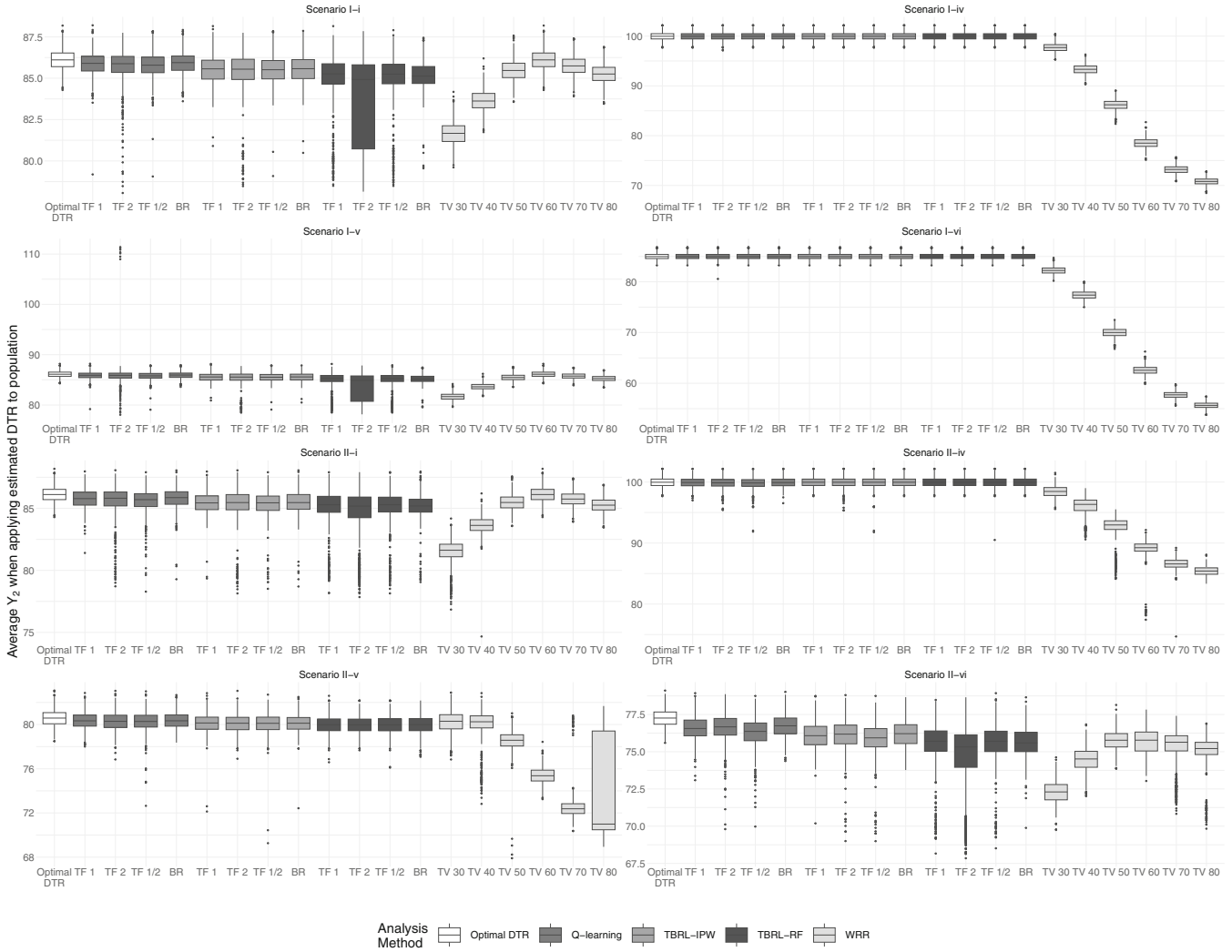


FIGURE 2 The average second stage outcome when applying the estimated DTR to a new sample of 10,000 patients from the target population. TBRL is done using a minimum purity improvement parameter of $\lambda = 0.01$ DTR = dynamic treatment regimen, Y_2 = second stage outcome, TBRL-IPW = tree-based reinforcement learning with inverse propensity weight, TBRL-RF = tree-based reinforcement learning with a random forest, WRR = weighted and replicated regression, TF = tailoring function, TV = tailoring variable.

there was at least one cut off that performed similarly to the tailoring functions. As expected, the TV-SMARTs using the cut off of 60 as a tailoring variable (TV 60) performs well in most settings since that is the optimal cut off in scenarios I-i, I-ii, I-iii, I-v, II-i, II-ii, and II-iii. However, the optimal cut off of 60 is not known at the time of trial design. In these scenarios with the TV-SMART design using TV 60, 100% of patients are recommended to receive the same treatment as they would with the optimal DTR (Figure 3). Thus, these DTRs estimated from these trials yield estimated optimal DTRs nearly equivalent performance to the true optimal DTR. However, there were also cut offs that did not perform as well. Overall, there was a wider variation in $\overline{Y}_2$ when conducting a TV-SMART compared to a TF-SMART or a BR-SMART.

Interestingly, in Scenario II-v where there was not a clear DTR based on second stage models, there is extreme variance of the average second stage outcome based on which tailoring variable was used. This was due to the WRR method described in Section 3.2 selecting a DTR with the incorrect first stage treatment. In this scenario, the TV-SMART using TV 80 selected treatment B for the first stage treatment over treatment A 29.2% of the time. In Scenario I-v, the TF-SMART using TF 2 selected treatment B in 0.7% of simulations when using Q-learning, 0.8% when using TBRL-IPW, and 3.5% when using TBRL-RF. These results explain the outliers for the TF-SMART using TF 2 with Q-learning in Figure 2. In Scenario II-v, all tailoring function methods selected treatment B at least 0.3% of the time except TBRL-RF which selected correctly 100% of the time. However, in Scenario II-v the TBRL-RF method failed to identify the DTR (Web Table 3). In Scenario I-vi, when using TF 2, Q-learning and TBRL-IPW selected treatment B 0.1% of the time. All other design, settings,

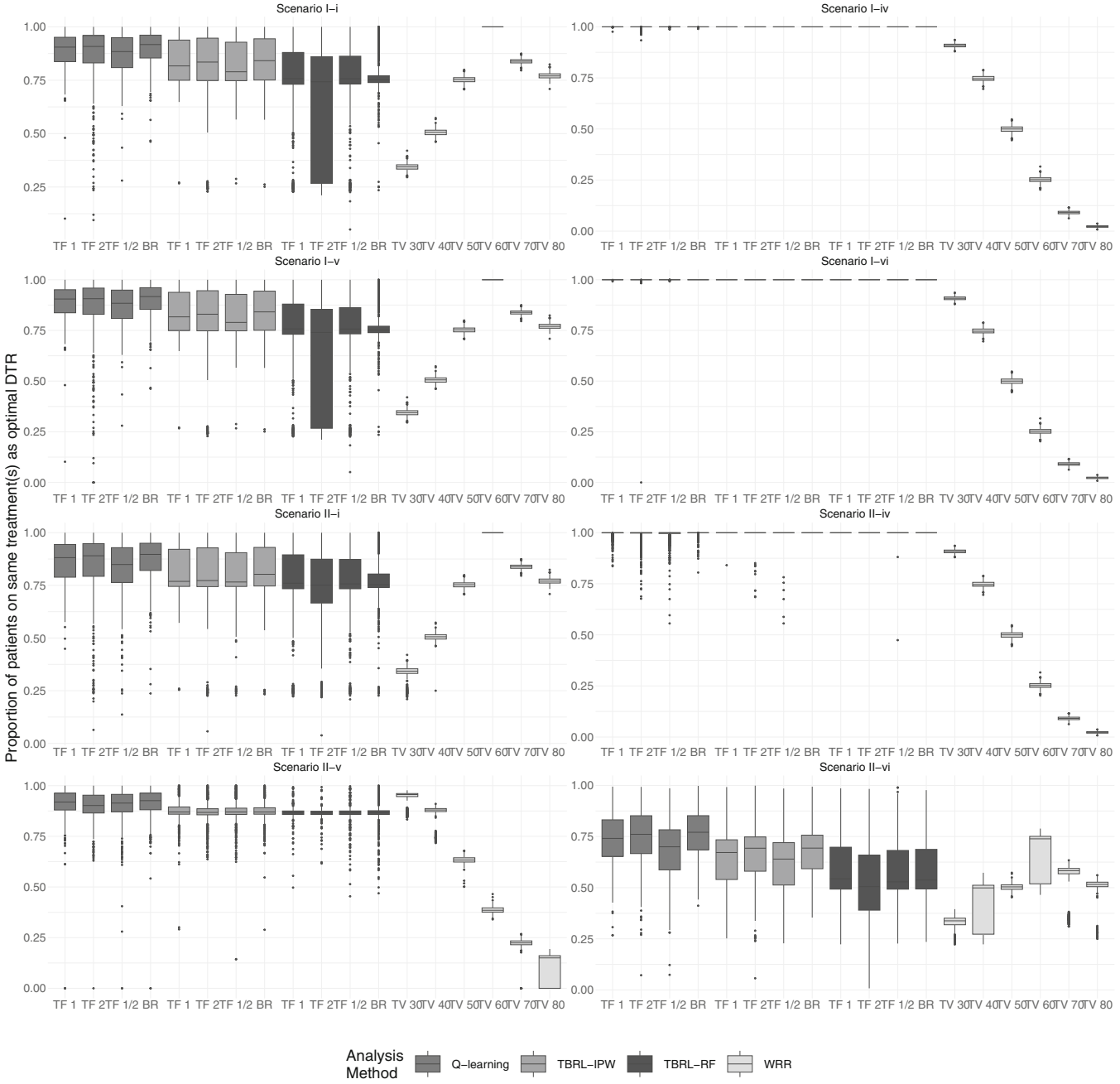


FIGURE 3 Overall proportion of patients in sample of 10,000 from the target population that end up on the same treatments as they would be assigned with the optimal DTR as defined in Section 4.4.1 and Table 1. DTR = dynamic treatment regimen, TBRL-IPW = tree-based reinforcement learning with inverse propensity weight, TBRL-RF = tree-based reinforcement learning with a random forest, WRR = weighted and replicated regression, TF = tailoring function, TV = tailoring variable.

and methods selected treatment A 100% of the time. In trials that selected the incorrect first stage treatment, none of the patients then received the treatments recommended by the optimal DTR (Figure 3).

In scenarios where no DTR was useful (I-iv and II-iv), the TV-SMARTs resulted in worse average second stage outcomes in the new population than the TF-SMARTs. TF-SMARTs have the flexibility to select no DTR, whereas TV-SMARTs do not. Assigning any cut point for a tailoring variable had poor performance. Performance decreased as the cut off increased since more people were assigned to a different treatment. In Scenario II-iv where the optimal DTR was composed of three second stage treatment options as opposed to two, the TV-SMARTs never reached the same performance in $\overline{Y}_2$ as the TF-SMARTs which allow for the flexibility to identify three second stage treatment options.

However, the percent of patients that receive the same treatments as would be assigned based on the optimal DTR is nearly equivalent between trials using TV 60 and the trials using tailoring functions.

For both Designs I and II, Scenarios i, ii, and iii had very similar results for the expected second stage outcome. This is because the optimal DTR including the optimal cut off point for Y_1 are identical between these settings.

When comparing the proportion of patients that end up on the optimal treatment pathway within the trial, we see a distinction between the TF-SMARTs and the BR-SMARTs (Figure 4). In a BR-SMART, the proportion of patients on the optimal DTR, given that the patients start on treatment A, is inversely proportional to the number of second stage treatments. In Design I the proportion is 1/2, and in Design II the is 1/3. With a TF-SMART design, there is the possibility of having more or fewer patients within the trial on the optimal treatment pathway. In general, if the tailoring function matches the data well and assuming a positive correlation between Y_1 and $E(Y_2)$, then there will be more patients receiving the optimal treatment within the trial.

5.2 | Comparison of analysis methods for TF-SMART designs

Q-learning tended to perform better than both forms of TBRL in terms of maximizing $\overline{Y}_2$ (Figure 2). The average second stage outcomes are higher with Q-learning, but there is also wider range of average second stage outcomes when using Q-learning. The performance of Q-learning and TBRL-RF vary greatly, primarily when using TF 2. This is likely due to having very few patients that are randomized to certain treatment arms. In Q-learning, the low number of patients in turn results in poor slope and intercept parameter estimates ($\hat{\beta}$s and $\hat{\gamma}$s, respectively) when modeling Equations (4)–(7). Since the slope and intercept parameter estimates are directly used to estimate the DTR and cut off in Q-learning, more accurate estimation of the slope and intercept parameters yields better estimation of the cut off. The wide variation in results for TBRL-RF when using TF 2 are also likely due to having a low number of patients in that arm which results in very little information for the random forest to make splits. Meanwhile, TBRL-IPW relies on the estimated parameters, splitting rules, and observed outcomes to define a cut off so it may be more robust to having fewer patients on certain treatment arms.

Next we compare the ability to estimate the correct optimal DTR cut off in a TF-SMART between Q-learning and TBRL by examining the bias as defined in Section 4.4.4. Cases where no cut off was defined were excluded. We considered

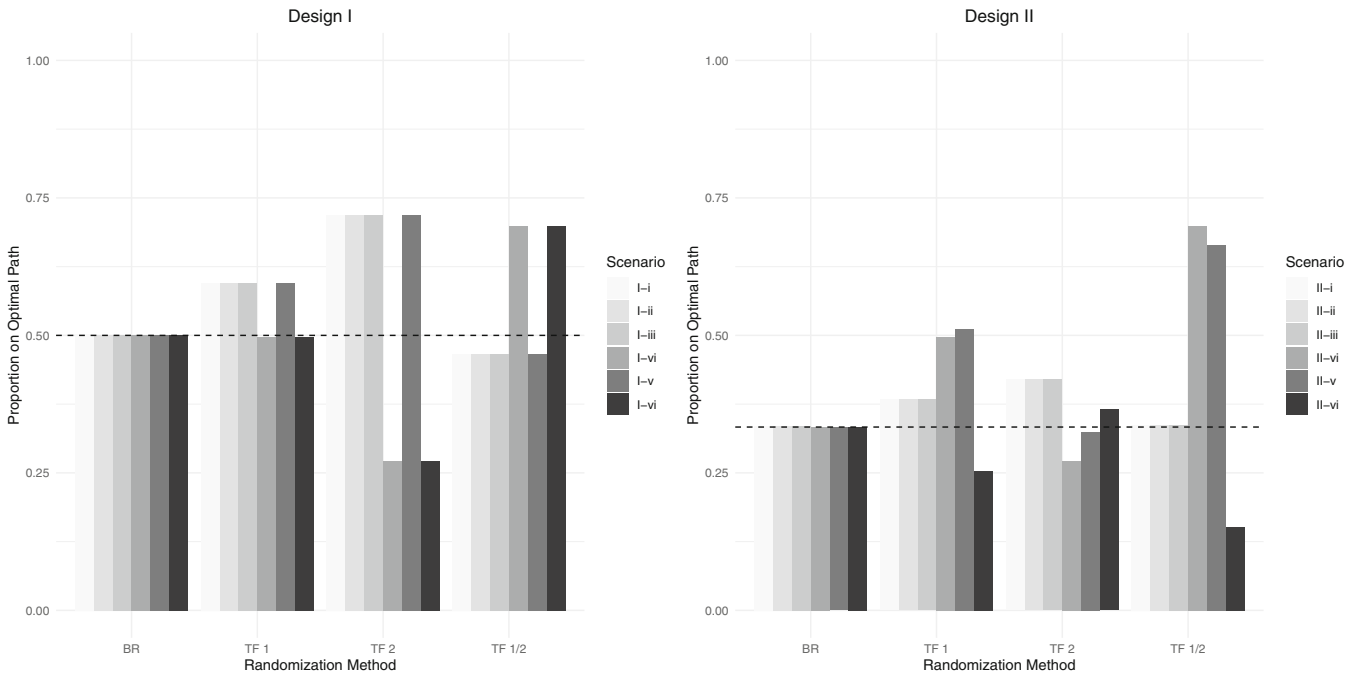


FIGURE 4 The proportion of patients in the trial receiving treatment A for the first stage treatment who end up receiving the same treatment as they would receive if following the optimal DTR as defined in Section 4.4.1 and Table 1. BR = balance randomization, TF = tailoring function.

simulations in which the estimated cut off was outside $\pm 3\sigma$ of the optimal cut as not defining a cut off as the majority of patients' first stage outcomes would lie within these bounds based on the known distribution of Y_1. That is, nearly all patients will receive the same treatment sequence in this case. We also excluded cases when the TBRL did not make any splits based on the splitting rules defined in Section 3.4. The TBRL method did not select a cut off in over 40% of trials where the optimal DTR does include a cut off (Web Table 3). We found that analyzing the data using Q-learning resulted in less bias in the estimation of the cut off, but wider variance than trials using TBRL that selected a cut off (Figure 5). However, since the TBRL method did not select a cut off in many of the trials, this result should be interpreted with caution. For Q-learning, the median bias of the estimated cut off ranged between −0.75 to 0.86 across all scenarios. Meanwhile, for TBRL-IPW, the median bias of the estimated cut off ranged between −6.31 and 4.63 and for TBRL-RF the median bias ranged from −9.05 to 6.39 across all scenarios. We then focused on TBRL-IPW to identify the source of this bias. The pattern where TBRL-IPW has larger bias than Q-learning also held in simulations where the trial size was 1000 patients with 500 on each initial arm (Web Figure 6) indicating that this bias is not due to insufficient patients on the trial to estimate the optimal split and cut off. When we changed the minimum purity requirement to allow a split from $\lambda = 0.01$ to $\lambda = 0.00001$, we saw improvement in the bias for this method (Web Figure 7). This also resulted in a smaller

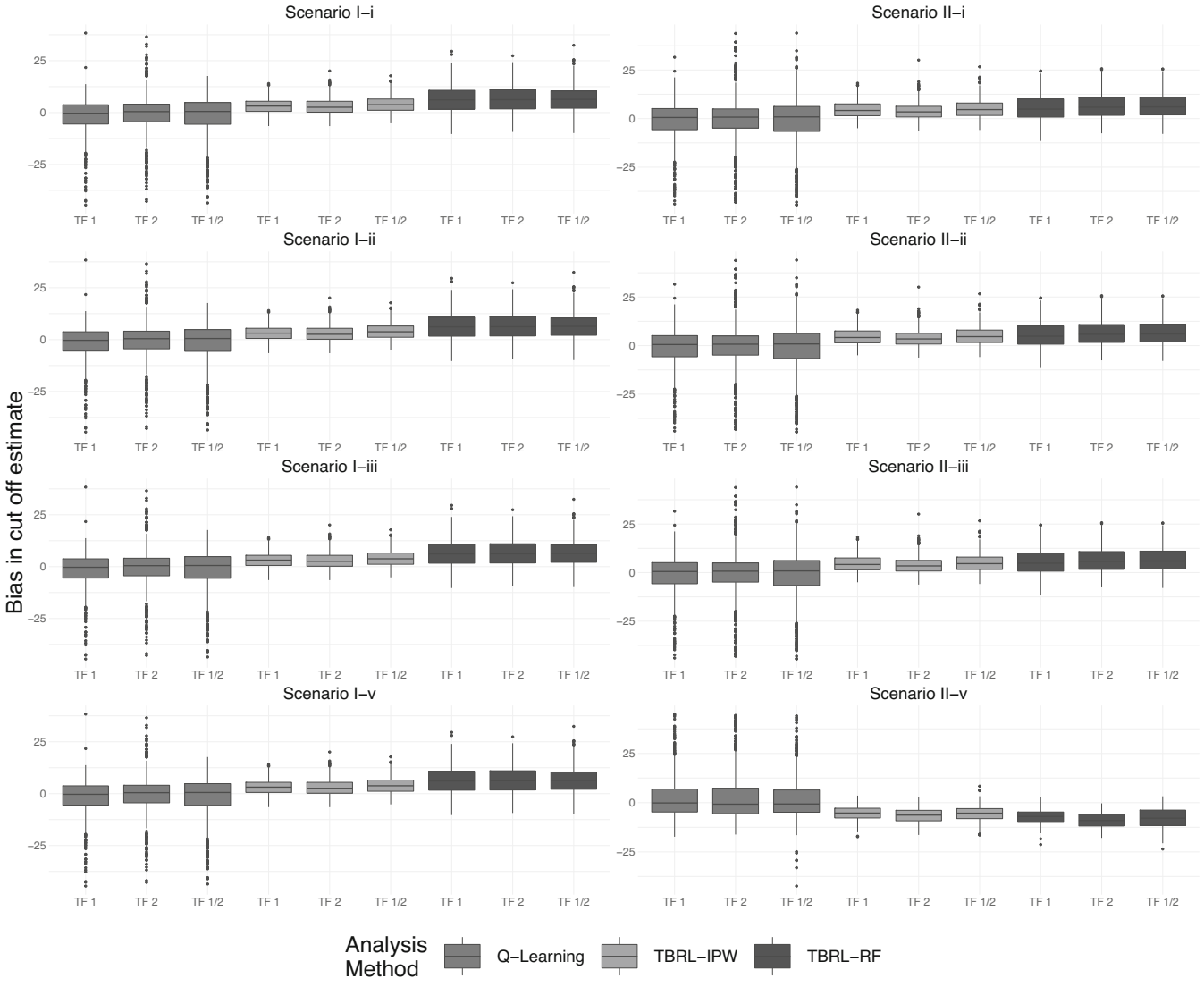


FIGURE 5 The bias in the estimated cut offs relative to the optimal cut off with total sample size of 300 and the minimum purity measure $\lambda = 0.01$. The optimal cut off is defined in Section 4.4.1 and Table 1. The bias is defined in Section 4.4.4. DTR = dynamic treatment regimen, TBRL-IPW = tree-based reinforcement learning with inverse propensity weight, TBRL-RF = tree-based reinforcement learning with a random forest, TF = tailoring function.

proportion of trials that incorrectly recommended no DTR (Web Table 4). This suggests that if the minimum improvement is set too high, the TBRL algorithm will have bias when estimating the optimal cut off or will incorrectly recommend no DTR and recommend just one fixed treatment. Additionally, when the minimum improvement parameter is set lower, the performance in terms of average second stage outcomes of the TBRL method becomes comparable to the performance of Q-learning (Web Figure 8). However, in the literature, it is recommended that this parameter not be treated as a tuning parameter.[14] Despite the bias in the estimated DTR cut off when using TBRL-IPW with sample size of 300 and $\lambda = 0.01$, the majority of patients still ended up on the same treatment as they would receive when using the optimal DTR (Figure 3).

6 | DISCUSSION

Here we present a TF-SMART design and analytic methods that estimate the optimal DTR, including defining the tailoring variable, using a continuous intermediate outcome. DTRs estimated from trials with our proposed flexible TF-SMART design lead to improved patient outcomes compared to the current standard TV-SMART designs while circumventing the need to define a binary or categorical tailoring variable a priori. When the tailoring variable is a dichotomization of a continuous variable, there is wide variation in average second stage outcomes based on what cut off is used. When the cut off selected is not the true optimal cut off, there is a decrease in performance. Tailoring functions circumvent the need to select a cut off before the trial has been conducted. This allows researchers greater flexibility and shorter time to treatment effect estimates while maintaining good statistical properties.

While the TF-SMART and BR-SMART designs have similar performance in terms of estimating the correct DTR, TF-SMARTs can benefit patients enrolled in the trials by increasing the probability of the patient being placed on the optimal treatment pathway. This potential benefit could make a clinical trial more appealing to patients. However, in a situation where the first stage and second stage treatment are not positively correlated (ie, patients do well in the first stage, but worse in second stage on some treatment pathway), patients would likely not benefit from a TF-SMART. Additionally, if a good candidate for a tailoring function is unknown, a BR-SMART may be preferable so that the data can be mined at the end of the trial to estimate the optimal DTR.

In our analytic methods, we focused on Q-learning and TBRL. However, there are many other methods that could be used as well as extensions to the methods presented here. There are methods that allow TBRL to estimate the minimum purity parameter improvement parameter in order to accurately predict when a split is warranted.[6,14] We used 0.01 as this was the default setting and a differences smaller than 1% in second stage outcome is likely not clinically significant. Additionally, the literature states that this parameter should not be viewed as a tuning parameter and rather set to be the minimum clinically significant difference. However, since we saw a change in performance based on the minimum purity parameter, using an estimated purity measure may result in improved performance for this method. There are also other methods that have been used to estimate DTRs with observational data that could be implemented in the TF-SMART setting such as OWL,[16] sequential outcome weighted multicategory learning,[18] and Bayesian machine learning.[15]

When randomization is appropriate, one major advantage to developing DTRs in a TF-SMART as opposed to using observational data is that propensity weights are known by design. We do not require a model to estimate the propensity weights. This is especially beneficial as estimates can be biased when the propensity model is misspecified.[19-21]

Although we presented models without covariates to illustrate the analysis of our trial design, it is easy to incorporate covariates. Both Q-learning and TBRL have been used to analyze observational data with a multitude of covariates to define DTRs and this can easily be translated to a randomized trial setting. The final product would be a DTR that could include covariates and the first stage outcome for a more personalized treatment regimen. While estimating sample size is outside the scope of this paper, incorporating covariates may require a larger sample size than a design that does not incorporate covariates. Simulations could be run to estimate the required sample size. In addition to including covariates in the tailoring variable for the estimated DTR, the tailoring function could also include covariates in the design stage. The tailoring function is flexible and could be defined with more arguments than just the first stage outcome (such as baseline measures) and still output a probability of receiving each second stage treatment.

In some trials, the difference in outcomes between treatments and DTRs may be much smaller than what we presented here. In this case, the sample size will need to be larger to be able to effectively identify the optimal DTR.

Additional variations of the TF-SMART design beyond Figure 1 are also feasible. The presented analytic methods can incorporate as many treatments and stages as practical. The methods here can also handle if there are uneven treatment arms where there are a different number of second stage treatment options based on first stage treatment. The weights in Section 3.1 and the models in Section 3.3 would need to be modified to match the design.

While the TF-SMART design presented here has the advantage that researchers do not need extensive knowledge about the optimal cut off a priori, if there is a wealth of knowledge, then a traditional TV-SMART may be more appropriate. We view this design as being most useful as a Phase II trial to estimate the optimal DTR to use in a larger Phase III trial. If there is prior knowledge about the treatment effects and a TF-SMART design is still desired, data from the TF-SMART could be analyzed using Bayesian methods for defining a DTR (such as Bayesian Machine Learning as proposed by Murray et al[15]) with a strong prior to leverage that knowledge.

We recommend a TV-SMART when the optimal tailoring variable cut off is fully known, BR-SMART when nothing is known about the distribution of the outcome, and TF-SMARTs when there is some prior knowledge about the outcome distribution but not of the tailoring variable. SMART designs are often used when comparing multiple available and approved treatments and this scenario would lend itself well to a TF-SMART.

Overall, we demonstrated the use of a tailoring function in a SMART which offers a new and flexible way to develop DTRs in a clinical trial setting. By using a tailoring function to assign probabilities of receiving each treatment and applying methods for observational studies to our trial data, we expect to expand the scope of SMARTs and enable the development of effective DTRs across a wide variety of diseases and disorders.

ACKNOWLEDGEMENTS

This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University. This work was partially supported by National Institutes of Health Grant P30-CA046592 and Patient-Centered Outcomes Research Institute (PCORI) Award ME-1507-31108.

DATA AVAILABILITY STATEMENT

Simulated data available on request from the authors.

ORCID

Holly Hartman https://orcid.org/0000-0002-9101-4381
Matthew Schipper https://orcid.org/0000-0001-6728-4720
Kelley Kidwell https://orcid.org/0000-0002-1717-4483

REFERENCES

1. Chakraborty B, Murphy SA. Dynamic treatment regimes. *Ann Rev Stat Appl*. 2014;1:447-464. doi:10.1146/annurev-statistics-022513-115553
2. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy S. A 'Smart' design for building individualized treatment sequences. *Annu Rev Clin Psychol*. 2012;8(1):21-48.
3. Almirall D, Compton SN, Gunlicks-Stoessel M, Duan N, Murphy SA. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Stat Med*. 2012;31(17):1887-1902. doi:10.1002/sim.4512
4. Murphy SA, Laan V, Robins JM. CPPRG. Marginal mean models for dynamic regimes. *J Am Stat Assoc*. 2001;96(456):1410-1423. doi:10.1021/nl061786n.Core-Shell
5. Moodie EEM, Chakraborty B, Kramer MS. Q-learning for estimating optimal dynamic treatment rules from observational data. *Can J Stat*. 2012;40(4):629-645.
6. Tao Y, Wang L, Almirall D. Tree-based reinforcement learning for estimating optimal dynamic treatment regimes. *Ann Appl Stat*. 2018;12(3):1914-1938.
7. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat Med*. 2005;24(10):1455-1481. doi:10.1002/sim.2022
8. Seewald NJ, Hackworth O, Almirall D. *Sequential, Multiple Assignment, Randomized Trials (SMART)*. Cham: Springer International Publishing; 2020:1-19.
9. Hartman HE, Tamura RN, Schipper MJ, Kidwell KM. Design and analysis considerations for utilizing a mapping function in a small sample, sequential, multiple assignment, randomized trials with continuous outcomes. *Stat Med*. 2021;40(2):312-326. doi:10.1002/sim.8776
10. Kidwell KM, Seewald NJ, Tran Q, Kasari C, Almirall D. Design and analysis considerations for comparing dynamic treatment regimens with binary outcomes from sequential multiple assignment randomized trials. *J Appl Stat*. 2018;45(9):1628-1651. doi:10.1080/02664763.2017.1386773
11. Nahum-Shani I, Qian M, Almirall D, et al. Q-learning: a data analysis method for constructing adaptive interventions. *Psychol Methods*. 2012;17(4):478-494. doi:10.1037/a0029373
12. Ky L, Zeger SL. Longitudinal data analysis using generalized linear models kung-Yee. *Biometrika*. 1986;73(1):13-22.
13. Wu F, Laber EB, Lipkovich IA, Severus E. Who will benefit from antidepressants in the acute treatment of bipolar depression? A reanalysis of the STEP-BD study by Sachs et al. 2007, using Q-learning. *Int J Bipolar Disorders*. 2015;3(1):7. doi:10.1186/s40345-014-0018-5
14. Laber EB, Zhao YQ. Tree-based methods for individualized treatment regimes. *Biometrika*. 2015;102(3):501-514. doi:10.1093/biomet/asv028

15. Murray TA, Yuan Y, Thall PF. A Bayesian machine learning approach for optimizing dynamic treatment regimes. *J Am Stat Assoc*. 2018;113(523):1255-1267. doi:10.1080/01621459.2017.1340887
16. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *J Am Stat Assoc*. 2012;107(449):1106-1118. doi:10.1080/01621459.2012.695674
17. Nahum-Shani I, Qian M, Almirall D, et al. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychol Methods*. 2012;17(4):457-477. doi:10.1037/a0029372
18. Zhou X, Wang Y, Zeng D. Outcome-weighted learning for personalized medicine with multiple treatment options. *Proc Int Conf Data Sci Adv Anal*. 2018;2018:565-574. doi:10.1109/DSAA.2018.00072
19. Drake C. Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*. 1993;49(4):1231-1236.
20. Lee BK, Lessler J, Elizabeth AS. Improving propensity score weighting using machine learning. *Stat Med*. 2010;29(3):337-346. doi:10.1002/sim.3782
21. Rosenbaum PR, Rubin D. The central role of the porpensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. doi:10.1111/wbm.12134

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hartman H, Schipper M, Kidwell K. A sequential, multiple assignment, randomized trial design with a tailoring function. *Statistics in Medicine*. 2024;43(21):4055-4072. doi: 10.1002/sim.10161