



HHS Public Access

Author manuscript

Obs Stud. Author manuscript; available in PMC 2024 July 12.

Published in final edited form as:

Obs Stud. 2023 ; 9(4): 25–48. doi:10.1353/obs.2023.a906627.

Using Pilot Data for Power Analysis of Observational Studies for the Estimation of Dynamic Treatment Regimes

Eric J. Rose,

Department of Epidemiology and Biostatistics, University at Albany, Rensselaer, NY, 12144, USA

Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, H3A 1G1, Canada

Erica E. M. Moodie,

Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, H3A 1G1, Canada

Susan Shortreed

Kaiser Permanente Washington Health Research Institute, Seattle, WA, 98101, USA

Department of Biostatistics, University of Washington, Seattle, WA, 98195, USA

Abstract

Significant attention has been given to developing data-driven methods for tailoring patient care based on individual patient characteristics. Dynamic treatment regimes formalize this approach through a sequence of decision rules that map patient information to a suggested treatment. The data for estimating and evaluating treatment regimes are ideally gathered through the use of Sequential Multiple Assignment Randomized Trials (SMARTs), though longitudinal observational studies are commonly used due to the potentially prohibitive costs of conducting a SMART. Observational studies are typically powered for simple comparisons of fixed treatment sequences; a priori power or sample size calculations for tailored strategies are rarely if ever undertaken. This has led to many studies that fail to find a statistically significant benefit to tailoring treatment. We develop power analyses for the estimation of dynamic treatment regimes from observational studies. Our approach uses pilot data to estimate the power for comparing the value of the optimal regime, i.e., the expected outcome if all patients in the population were treated by following the optimal regime, with a known comparison mean. This allows for calculations that ensure a study has sufficient power to detect the need for tailoring, should it be present. Our approach also ensures the value of the estimated optimal treatment regime has a high probability of being within a range of the value of the true optimal regime, set a priori. We examine the performance of the proposed procedure with a simulation study and use it to size a study for reducing depressive symptoms using data from electronic health records.

ejrose@albany.edu .

Conflict of Interest

No potential conflict of interest was reported by the author(s).

Keywords

Adaptive treatment strategies; Confounding; Precision medicine; Sample size

1. Introduction

Data-driven methods for personalizing treatment assignment have been of great interest to clinicians and researchers. Dynamic treatment regimes (DTRs) operationalize clinical decision-making through a sequence of decision rules that map up-to-date patient information to a recommended treatment (Chakraborty and Moodie, 2013; Tsiatis et al., 2019). An optimal treatment regime is a set of decision rules that maximizes the mean of a measure of positive health outcome when all patients in the population of interest are assigned treatment by following that regime (Murphy, 2003; Robins, 2004). DTRs have been studied to improve decision-making in healthcare across many areas of applications such as cancer (Zhao et al., 2011; Wang et al., 2012), schizophrenia (Shortreed and Moodie, 2012), and depression (Chakraborty et al., 2016).

Many methods for estimating optimal treatment regimes have been proposed, three of which are Q-learning, G-estimation, and dynamic weighted ordinary least squares (dWOLS). Q-learning is a regression-based approach that is straightforward to implement in practice, but is not robust to model misspecification (Watkins and Dayan, 1992). G-estimation uses a contrast function to estimate an optimal regime, and is doubly robust, meaning it is robust to misspecification of either the outcome or propensity score model (Robins, 2004); it has, however, seen little uptake in applications. dWOLS is based on a series of weighted regression models and, like Q-learning, is straightforward to implement in a continuous outcome setting, yet possesses double-robustness (Wallace and Moodie, 2015).

Data for the estimation of optimal treatment regimes are ideally gathered through a Sequential Multiple Assignment Randomized Trial (SMART) (Lavori and Dawson, 2000, 2004; Murphy, 2005). SMARTs are typically sized for comparing specific treatment regimes, as opposed to identifying optimal regimes (Oetting et al., 2011; Lei et al., 2012; Artman et al., 2020; Seewald et al., 2020). However, data sourced from longitudinal observational studies are more commonly used because the resources for conducting SMARTs are often prohibitive and estimation of DTRs is often considered exploratory in nature (Chakraborty and Murphy, 2014). Observational studies are generally powered for simple research questions such as comparisons of fixed strategies. For either study type, power and sample size calculations typically do not take estimation of an optimal DTR into account. This has resulted in many studies of DTRs failing to find a statistically significant benefit to tailoring treatments to individual patient characteristics (e.g. Krakow et al., 2017; Simoneau et al., 2020; Coulombe et al., 2021). Further, there is no guarantee that the performance of an estimated regime will be close to that of the true optimal regime. Therefore, it has been advised that estimated optimal treatment regimes be evaluated with a follow-up study in which patients are randomized to regimes of interest (Murphy, 2005). This approach is costly since it requires conducting two studies. In addition, if the

original study is not powered to guarantee a high-quality estimate of the optimal regime, the follow-up study could focus on a poor-quality treatment regime.

There has been some work on power and sample size calculations for *randomized* trials for estimating optimal DTRs, i.e., estimating treatment strategies that are potentially more highly tailored than the simple, embedded strategies within the trial. These methods provide a sample size that ensures sufficient power to compare the optimal regime with standard of care as well as ensuring that the performance of the estimated optimal regime is close to that of the true optimal regime, however these existing approaches do not account for the potential loss of effective sample size (power) due to the confounding that typically is present in observational studies. Laber et al. (2016) proposed a method for using pilot data to size a two-armed randomized single-stage trial that is based on inverting a projection confidence interval; that is, the approach is suitable for a single, tailored decision but not a multi-stage SMART for DTRs. Rose et al. (2019) proposed two methods for sizing two-stage SMARTs for the estimation of optimal DTRs. The first of the Rose et al. methods imposes strong assumptions on the underlying data-generating model that assumes away the complexities related to nonregularity, leading to a sample size estimator that resembles a comparison of fixed treatment sequences. The second makes minimal assumptions and uses bootstrap oversampling (i.e., resampling with replacement with sample size greater than the original data size) with pilot data to estimate a sample size. No power calculations exist for observational studies for estimating optimal DTRs.

In this paper, we propose a method that uses pilot data to conduct power (or sample size) calculations for a multistage, longitudinal observational study for estimating DTRs. This approach is based on constructing a projection interval and using bootstrap oversampling to estimate the power for a given sample size. Alternatively, this method can also be used to conduct sample size calculations by estimating the sample size that results in the desired power. In observational studies, it is more common to perform power calculations for a given sample size, either prior to conducting an analysis (e.g., for a grant proposal) or following an analysis to determine whether low power could explain the null finding and thus to provide context for the interpretation of results (Morris and van Smeden, 2022; Campbell et al., 2022). In some contexts, however, sample size calculations for observational studies may be very informative. For instance, a sample size calculation inform researchers how many sites to include in a multi-site study or how many years of data are needed in a retrospective cohort study. Alternatively, it may be the case that the information required for the analysis requires additional processing or expense, such as if key biomarkers must be measured from frozen tissue samples or hand-written notes must be extracted from patient files.

This work provides the first procedure for conducting power and sample size calculations for estimating an optimal treatment regime from observational data as well as the first method for power and sample size calculations for any study for estimating an optimal regime that consists of more than two treatment decisions. Our method for power calculations requires a finite-sample (nonasymptotic) estimate of the variance of DTR parameters, which can be accomplished even in nonregular settings via the $m - \text{out} - \text{of} - n$ bootstrap (Chakraborty et al., 2013). This approach, however, has yet to be implemented in more than two stages.

Thus, a further contribution of this work is to demonstrate this resampling method for estimating confidence intervals for the parameters indexing a DTR with more than two stages.

We give an overview of the setup and notation for this work in Section 2. In Section 3, we present the proposed method for power calculations. In Section 4, we provide a simulation study to demonstrate the empirical performance of the proposed approach and the variability in the resulting sample size across different pilot studies for a three-stage observational study. In Section 5, we demonstrate the use of our method for sizing a study to reduce depressive symptoms using electronic health record (EHR) data from Kaiser Permanente Washington (KPWA). In Section 6, we give concluding remarks and a discussion of open problems.

2. Setup and Notation

We consider powering an observational study for estimating an optimal dynamic treatment regime with K sequential treatment decisions. We assume two potential treatment options at each stage. The observed data are of the form $\mathcal{D}_n = \{(\mathbf{X}_{1,i}, A_{1,i}, \dots, \mathbf{X}_{K,i}, A_{K,i}, Y_i)\}_{i=1}^n$, which comprises n i.i.d. replicates of $(\mathbf{X}_1, A_1, \dots, \mathbf{X}_K, A_K, Y)$ where: $\mathbf{X}_1 \in \mathbb{R}^{p_1}$ denotes baseline patient information, $A_k \in \{0, 1\}$ denotes the treatment assigned at the k^{th} stage, $\mathbf{X}_k \in \mathbb{R}^{p_k}$ for $k = 2, \dots, K$ denotes additional patient information recorded during the course of treatment ($k - 1$), and $Y \in \mathbb{R}$ denotes the outcome of interest, coded such that higher values are better. Let \mathbf{H}_k be patient history available to a clinical decision maker at stage k , so $\mathbf{H}_1 = \mathbf{X}_1$ and $\mathbf{H}_k = (\mathbf{X}_1^T, A_1, \dots, A_{k-1}, \mathbf{X}_k^T)^T$ for $k = 2, \dots, K$.

A treatment regime, \mathbf{d} , is defined as a set of decision rules $\mathbf{d} = (d_1, \dots, d_K)$, such that $d_k: \text{dom} \mathbf{H}_k \rightarrow \{0, 1\}$ for $k = 1, \dots, K$ is a function that maps a patient's history to a recommended treatment lying within the domain (dom) of possible treatments. Therefore, a patient with history $\mathbf{H}_k = \mathbf{h}_k$ is recommended to be assigned treatment $d_k(\mathbf{h}_k)$. It is helpful to be able to reference a patient's treatments received and history up to or after a certain stage. To do this, we use an overbar to denote the treatments, covariates, and regimes up to stage k , so we have that $\bar{\mathbf{a}}_k = (a_1, \dots, a_k)$, $\bar{\mathbf{x}}_k = (x_1, \dots, x_k)$, and $\bar{\mathbf{d}}_k = (d_1, \dots, d_k)$. When we are considering the entire sequence of K stages, we will suppress the subscript so that $\bar{\mathbf{a}} = \bar{\mathbf{a}}_K$ and $\bar{\mathbf{x}} = \bar{\mathbf{x}}_K$. An underbar will denote treatments, covariates, and regimes from stage k to K such that $\underline{\mathbf{a}}_k = (a_k, \dots, a_K)$, $\underline{\mathbf{x}}_k = (x_k, \dots, x_K)$, and $\underline{\mathbf{d}}_k = (d_k, \dots, d_K)$.

To formalize the notion of an optimal regime, we use the potential outcome framework (Rubin, 1978). Let $\mathbf{H}_k^*(\bar{\mathbf{a}}_{k-1})$ be the potential history under the treatment sequence $\bar{\mathbf{a}}_{k-1}$ and $Y^*(\bar{\mathbf{a}})$ denote the potential outcome under the treatment assignment $\bar{\mathbf{a}}$. The set of all potential outcomes is then denoted as $\mathbf{W}^* = \{\mathbf{H}_2^*(a_1), \mathbf{H}_3^*(\bar{\mathbf{a}}_2), \dots, \mathbf{H}_K^*(\bar{\mathbf{a}}_{K-1}), Y^*(\bar{\mathbf{a}}): \bar{\mathbf{a}} \in \{0, 1\}^K\}$. The potential outcome of following a regime, \mathbf{d} , is defined as

$$Y^*(\mathbf{d}) = \sum_{\bar{\mathbf{a}} \in \{0,1\}^K} Y^*(\bar{\mathbf{a}}) \mathbb{1}\{d_1(\mathbf{h}_1) = a_1\} \prod_{k=2}^K \mathbb{1}\{d_k(\mathbf{H}_k(a_{k-1})) = a_k\}$$

where $\mathbb{1}$ is the indicator function. Define the value of any regime by $V(\mathbf{d}) = \mathbb{E}\{Y^*(\mathbf{d})\}$. An optimal regime, \mathbf{d}^{opt} , is then defined as a regime that satisfies $V(\mathbf{d}^{opt}) \geq V(\mathbf{d})$ for all \mathbf{d} .

To be able to express the optimal regime in terms of the observed data, we will need three standard causal assumptions for DTRs (Robins, 2004): (C1) the stable unit treatment value assumption (SUTVA), $Y = Y^*(\bar{\mathbf{A}})$ and $\mathbf{H}_k = \mathbf{H}_k^*(\bar{\mathbf{A}}_{k-1})$ for $k = 2, \dots, K$; (C2) sequential ignorability, $\mathbf{W}^* \perp A_k \mid \mathbf{H}_k$ for $k = 1, \dots, K$; and (C3) positivity, $P(A_k = a_k \mid \mathbf{H}_k = \mathbf{h}_k) > 0$ with probability 1 for each $a_k \in \{0, 1\}$ for $k = 1, \dots, K$.

Many estimation methods for an optimal regime focus on estimating a contrast function which characterizes how the interaction of treatment and patient history affects the outcome of interest. These estimation methods are commonly referred to as A-learning or advantage learning methods (Blatt et al., 2004). The optimal blip-to-zero function, $\gamma_k(\mathbf{h}_k, a_k)$, is defined as the difference in expected outcome between receiving treatment a_k and some reference treatment, which we will take to be treatment 0, for a patient that has history $\mathbf{H}_k = \mathbf{h}_k$ if we assume they are treated optimally after stage k . Therefore, we have that

$$\begin{aligned} \gamma_k(\mathbf{h}_k, a_k) &= \mathbb{E}\{Y^*(\bar{\mathbf{a}}_{k-1}, a_k) - Y^*(\bar{\mathbf{a}}_{k-1}, 0) \mid \mathbf{H}_k = \mathbf{h}_k\}, \\ \gamma_k(\mathbf{h}_k, a_k) &= \mathbb{E}\{Y^*(\bar{\mathbf{a}}_{k-1}, a_k, \underline{\mathbf{d}}_{k+1}^{opt}) - Y^*(\bar{\mathbf{a}}_{k-1}, 0, \underline{\mathbf{d}}_{k+1}^{opt}) \mid \mathbf{H}_k = \mathbf{h}_k\} \\ &\text{for } k = 2, \dots, K-1, \\ \gamma_1(\mathbf{h}_1, a_1) &= \mathbb{E}\{Y^*(a_1, \underline{\mathbf{d}}_2^{opt}) - Y^*(0, \underline{\mathbf{d}}_2^{opt}) \mid \mathbf{H}_1 = \mathbf{h}_1\}. \end{aligned}$$

Note that the treatments assigned by $\underline{\mathbf{d}}_{k+1}^{opt}$ may differ according to its arguments. For example, consider $\gamma_k(\mathbf{h}_k, a_k)$: For the term $Y^*(\bar{\mathbf{a}}_{k-1}, a_k, \underline{\mathbf{d}}_{k+1}^{opt})$ we will have that $\underline{\mathbf{d}}_{k+1}^{opt}$ assigns treatment $\underline{\mathbf{d}}_{k+1}^{opt} \{ \mathbf{H}_{k+1}^*(\bar{\mathbf{a}}_{k-1}, a_k) \}$ while in $Y^*(\bar{\mathbf{a}}_{k-1}, 0, \underline{\mathbf{d}}_{k+1}^{opt})$ we have that $\underline{\mathbf{d}}_{k+1}^{opt}$ assigns treatment $\underline{\mathbf{d}}_{k+1}^{opt} \{ \mathbf{H}_{k+1}^*(\bar{\mathbf{a}}_{k-1}, 0) \}$. Then an optimal treatment can be seen as the treatment that maximizes the blip function at that stage, so we have that $\underline{\mathbf{d}}_k^{opt}(\mathbf{h}_k) = \operatorname{argmax}_{a_k} \gamma_k(\mathbf{h}_k, a_k)$.

A-learning methods alternatively can focus on the regret function. The regret function at stage k , $\mu_k(\mathbf{h}_k, a_k)$, is defined as the decrease in expected outcome from assigning treatment a_k instead of the optimal treatment if we assume that patients were treated optimally in all following stages. Therefore, for the K stages, the regret functions are given by

$$\begin{aligned} \mu_K(\mathbf{h}_K, a_K) &= E\{Y^*(\bar{\mathbf{a}}_{K-1}, \underline{\mathbf{d}}_K^{opt}) - Y^*(\bar{\mathbf{a}}_{K-1}, a_K) \mid \mathbf{H}_K = \mathbf{h}_K\}, \\ \mu_k(\mathbf{h}_k, a_k) &= E\{Y^*(\bar{\mathbf{a}}_{k-1}, \underline{\mathbf{d}}_k^{opt}, \underline{\mathbf{d}}_{k+1}^{opt}) - Y^*(\bar{\mathbf{a}}_{k-1}, a_k, \underline{\mathbf{d}}_{k+1}^{opt}) \mid \mathbf{H}_k = \mathbf{h}_k\} \\ &\text{for } k = 2, \dots, K-1, \\ \mu_1(\mathbf{h}_1, a_1) &= E\{Y^*(\underline{\mathbf{d}}_1^{opt}, \underline{\mathbf{d}}_2^{opt}) - Y^*(a_1, \underline{\mathbf{d}}_2^{opt}) \mid \mathbf{H}_1 = \mathbf{h}_1\}. \end{aligned}$$

The regret function and the optimal blip-to-zero function are then related by $\mu_k(\mathbf{h}_k, a_k) = \gamma_k\{\mathbf{h}_k, d_k^{opt}(\mathbf{h}_k)\} - \gamma_k(\mathbf{h}_k, a_k)$. The value of an optimal regime can then be expressed as

$$V(\mathbf{d}^{opt}) = \mathbb{E}\left[Y^*(\bar{\mathbf{A}}) + \sum_{k=1}^K [\gamma_k\{\mathbf{H}_k, d_k^{opt}(\mathbf{H}_k)\} - \gamma_k(\mathbf{H}_k, A_k)]\right].$$

Our power calculations are based on using dWOLS to estimate the parameters of the blip functions, which then leads to a substitution estimator for the value of the optimal regime. We will let an estimator of \mathbf{d}^{opt} from an observational study of size n be denoted by $\hat{\mathbf{d}}_n$. Let $B_0 \in \mathbb{R}$ be a fixed, known mean value such that we want to test if assigning treatment by following \mathbf{d}^{opt} would lead to a mean outcome greater than B_0 . The choice of B_0 will depend on the research question of interest and could represent the mean outcome under a specific static or dynamic regime, e.g., a fixed treatment sequence or standard of care. Therefore, we will construct an α -level test of the null hypothesis, $H_0: V(\mathbf{d}^{opt}) \leq B_0$, based on the estimator, $\hat{\mathbf{d}}_n$. Let $\eta > 0$ denote a clinically meaningful increase in the expected outcome. We will estimate the power given by the probability of rejecting the null hypothesis conditional on $V(\mathbf{d}^{opt}) \geq B_0 + \eta$.

We will then construct a sample size estimator using our proposed method for power calculations that will satisfy two separate conditions. These conditions match those used in Rose et al. (2019) which were based on those used in Laber et al. (2016). Let $\eta, \epsilon > 0, \phi, \alpha, \zeta \in (0, 1)$ be constants. Our goal is to choose n such that:

(PWR) there exists an α -level test of the null hypothesis, $H_0: V(\mathbf{d}^{opt}) \leq B_0$, based on the estimator, $\hat{\mathbf{d}}_n$, that has a power of at least $(1 - \phi) \times 100 + o(1)$ provided $V(\mathbf{d}^{opt}) \geq B_0 + \eta$; (OPT) $P\{V(\hat{\mathbf{d}}_n) \geq V(\mathbf{d}^{opt}) - \epsilon\} \geq 1 - \zeta + o(1)$.

Each condition guarantees a different aspect of our sample size procedure. The first condition, (PWR), ensures that if tailoring treatments based on patient history provides a *clinically* significant improvement in outcomes, then a study sized with our approach will be sufficiently powered to detect a *statistically* significant difference between the value of the optimal regime and B_0 . The second condition, (OPT), guarantees the true value of a regime estimated from a study based on our power calculations will be within a specified tolerance of the value of the true optimal regime. The quantity $V(\hat{\mathbf{d}}_n)$ represents the marginal mean outcome if the estimated optimal regime $\hat{\mathbf{d}}_n$ is used to assign treatments to the population of interest and can be expressed as $V(\hat{\mathbf{d}}_n) = \mathbb{E}\{Y^*(\hat{\mathbf{d}}_n) \mid \mathcal{D}_n\}$. This will ensure that the performance of our estimated optimal regime will be close to the true optimal regime.

3. Methodology of Proposed Power Calculations

3.1 Dynamic Weighted Ordinary Least Squares

dWOLS estimates parameters in the blip functions using a sequence of weighted ordinary least squares regressions (Wallace and Moodie, 2015). We posit a model for the blip-to-zero functions given by $\gamma_k(\mathbf{h}_k, a_k; \psi_k)$ for $k = 1, \dots, K$. We will assume that each blip-to-zero function is correctly specified such that $\gamma_k(\mathbf{h}_k, a_k; \psi_k^*) = \gamma_k(\mathbf{h}_k, a_k)$.

The treatment-free outcome is given by

$$G_k(\underline{\psi}_k) = Y - \gamma_k(\mathbf{h}_k, a_k; \psi_k) + \sum_{j=k+1}^K \mu_j(\mathbf{h}_j, a_j; \psi_j)$$

such that $\underline{\psi}_k = (\psi_k, \dots, \psi_K)$. If we assume $\underline{\psi}_k$ is the true value of the parameter in our blip models, then $G_k(\underline{\psi}_k)$ represents the patients' actual outcome adjusted for the expected difference in outcome if they received treatment 0 at stage k and then were treated optimally for the remaining stages. This is referred to as the treatment-free outcome since it does not depend on the treatment received at stage k , though it is "treatment-free" only if the reference treatment is no treatment. When we are considering only active treatments, the treatment-free outcome denotes the expected outcome under the reference treatment at stage k . Under assumptions (C1)-(C3), we have that $\mathbb{E}\{G_k(\underline{\psi}_k^*) \mid \mathbf{H}_k = \mathbf{h}_k\} = \mathbb{E}\{Y^*(\bar{\mathbf{a}}_{k-1}, 0, \underline{\mathbf{d}}_{k+1}^{opt}) \mid \mathbf{H}_k = \mathbf{h}_k\}$. We then specify a model for $\mathbb{E}\{G_k(\underline{\psi}_k^*) \mid \mathbf{H}_k = \mathbf{h}_k\}$ that will be given by $g_k(\mathbf{h}_k; \beta_k)$.

Define the pseudo-outcome for stage k as $\tilde{Y}_k = Y + \sum_{j=k+1}^K \mu_j(\mathbf{h}_j, a_j; \hat{\psi}_j)$. The pseudo-outcome for stage k represents the estimated counterfactual outcome if treatments were assigned via our estimated optimal rules after stage k . We model the pseudo-outcome as the sum of the treatment-free model and the blip-to-zero model

$$\mathbb{E}(\tilde{Y}_k \mid \mathbf{H}_k = \mathbf{h}_k, A_k = a_k; \beta_k, \psi_k) = g_k(\mathbf{h}_k; \beta_k) + \gamma_k(\mathbf{h}_k, a_k; \psi_k).$$

We could then estimate β_k and ψ_k as a standard regression problem, which would lead to an estimated optimal decision rule at stage k , $\hat{d}_k^{opt}(\mathbf{h}_k) = \arg \max_{a_k \in \{0,1\}} \gamma_k(\mathbf{h}_k, a_k; \hat{\psi}_k)$. Note that the estimated regime depends only on ψ_k , so β_k is a nuisance parameter, but a consistent estimator of ψ_k would require correctly specifying our models for both $g_k(\mathbf{h}_k; \beta_k)$ and $\gamma_k(\mathbf{h}_k, a_k; \psi_k)$. Therefore, we posit a model $\pi_k(\mathbf{h}_k; \xi_k)$ for the propensity score $\pi_k(\mathbf{h}_k) = \mathbb{E}(A_k \mid \mathbf{H}_k = \mathbf{h}_k)$. Wallace and Moodie (2015) show that if we perform a weighted ordinary least squares with a weight function that satisfies $\pi_k(\mathbf{h}_k; \xi_k)w(1, \mathbf{h}_k; \xi_k) = \{1 - \pi_k(\mathbf{h}_k; \xi_k)\}w(0, \mathbf{h}_k; \xi_k)$, then the resulting estimate of ψ_k will be consistent as long as the blip model is correctly specified and either the treatment-free model or the propensity score model is correctly specified. Weight functions that satisfy this

equality include $w(a_k, \mathbf{h}_k; \xi_k) = |a_k - \pi(\mathbf{h}_k; \xi_k)|$ and the inverse probability of treatment weights given by $w(a_k, \mathbf{h}_k; \xi_k) = a_k \{\pi(\mathbf{h}_k; \xi_k)\}^{-1} + (1 - a_k) \{1 - \pi(\mathbf{h}_k; \xi_k)\}^{-1}$.

3.2 Inference for the Value

We assume linear models for both the treatment-free and blip models. Therefore the model for the pseudo-outcome will be given by

$$\mathbb{E}(\tilde{Y}_k | \mathbf{H}_k = \mathbf{h}_k, A_k = a_k; \beta_k, \psi_k) = \mathbf{h}_{k,\beta}^T \beta_k + a_k \mathbf{h}_{k,\psi}^T \psi_k$$

where $\mathbf{h}_{k,\beta}$ and $\mathbf{h}_{k,\psi}$ are components of \mathbf{h}_k , each including a leading one. The estimated optimal treatment at stage k is given by $\mathbb{1}(\mathbf{h}_{k,\psi}^T \hat{\psi}_k > 0)$, and the pseudo-outcome for stage k is given by $\tilde{Y}_k = Y + \sum_{j=k+1}^K \mathbf{h}_{j,\psi}^T \hat{\psi}_j \{\mathbb{1}(\mathbf{h}_{j,\psi}^T \hat{\psi}_j > 0) - a_j\}$. Note the pseudo-outcome at stage k is a nonsmooth function of the generative model because of the indicator function. Therefore, the estimator for ψ_k is nonregular when $k < K$ and standard approaches for inference no longer hold because $\sqrt{n}(\hat{\psi}_k - \psi_k)$ is not uniformly normal (Robins, 2004).

Similarly, an estimator for the value is given by

$$\hat{V}_n(\hat{\mathbf{d}}^{opt}) = \mathbb{P}_n \left[Y + \sum_{k=1}^K \mathbf{H}_{k,\psi}^T \hat{\psi}_k \{\mathbb{1}(\mathbf{H}_{k,\psi}^T \hat{\psi}_k > 0) - A_k\} \right]$$

where \mathbb{P}_n denotes the empirical expectation. $V(\mathbf{d}^{opt})$ is a nonsmooth function of the generative model as well, so again, standard approaches for inference do not hold. Therefore, to estimate the power for a given sample size, we invert a projection confidence interval for $V(\mathbf{d}^{opt})$ (Lalor et al., 2014). This interval is valid as long as the blip model is correctly specified, in addition to either the propensity score model or the treatment-free model being correctly specified.

Define $Y(\psi) = Y + \sum_{k=1}^K \mu_k(\mathbf{H}_k, A_k; \psi_k)$, $V(\psi) = \mathbb{E}\{Y(\psi)\}$, and $\hat{V}_n(\psi) = \mathbb{P}_n\{Y(\psi)\}$. Thus $\hat{V}_n(\hat{\psi}) = \hat{V}_n(\hat{\mathbf{d}}^{opt})$. We also have that $V(\mathbf{d}^{opt}) = \mathbb{E}\{Y(\psi^*)\}$. Note that the value function can then be expressed as a function of either a treatment regime, \mathbf{d} , or the parameters indexing a treatment regime, ψ . Define $\zeta^2(\psi) = \mathbb{E}\{Y(\psi) - \mathbb{E}Y(\psi)\}^2$ and $\hat{\zeta}_n^2(\psi) = \mathbb{P}_n\{Y(\psi) - \mathbb{P}_n Y(\psi)\}^2$. Then for a fixed value of ψ , if $\mathbb{E}\{Y^2(\psi)\} < \infty$

$$\sqrt{n}\{\hat{V}_n(\psi) - V(\psi)\} \rightarrow \text{Normal}\{0, \zeta^2(\psi)\}.$$

Let $\Psi_{n,1-\vartheta}$ denote a $(1 - \vartheta) \times 100\%$ confidence region for ψ^* . If we choose ϑ_1 and ϑ_2 such that $\vartheta_1 + \vartheta_2 = \alpha$, then an α -level test for $H_0: V(\mathbf{d}^{opt}) \leq B_0$ rejects when

$$\inf_{\psi \in \Psi_{n,1-\vartheta_1}} \left\{ \hat{V}_n(\psi) - \frac{z_{1-\vartheta_2} \hat{\zeta}_n(\psi)}{\sqrt{n}} \right\} \geq B_0$$

where $z_{1-\vartheta}$ denotes the $(1 - \vartheta)$ quantile of a standard normal distribution. See Section A of the online Supplementary Material for proof that this is an α -level test. The power for this test is given by

$$\begin{aligned} & P \left[\inf_{\psi \in \Psi_{n,1-\vartheta_1}} \left\{ \widehat{V}_n(\psi) - \frac{z_{1-\vartheta_2} \widehat{\zeta}_n(\psi)}{\sqrt{n}} \right\} \geq B_0 \right] \\ & \geq P \left[\inf_{\psi \in \Psi_{n,1-\vartheta_1}} \left[\frac{\sqrt{n} \{ \widehat{V}_n(\psi) - V(\psi) \}}{\widehat{\zeta}_n(\psi)} + \frac{\min[\sqrt{n}\{V(\psi) - B_0\}, \sqrt{n}\eta]}{\widehat{\zeta}_n(\psi)} \right] \geq z_{1-\vartheta_2} \right]. \end{aligned}$$

We replace $V(\psi) - B_0$ with $\min\{V(\psi) - B_0, \eta\}$ as in Rose et al. (2019) so the sample size is based on the minimal effect size of interest instead of the estimated effect size. This will result in our proposed sample size procedure having power $(1 - \phi) \times 100\%$ when the effect size is η , with the power increasing as the true effect size increases to greater than η .

3.3 Confidence Region for ψ

The proposed hypothesis test requires constructing a confidence region for ψ . When $K = 1$, constructing a confidence region for ψ can be done using standard theory for m-estimators (Van Der Vaart, 1998). Let $\mathbf{H}_{k,\beta}$ be the components of the history in the treatment-free model at stage k and let $\mathbf{H}_{k,\psi}$ be the components of the history in the blip-to-zero model at stage k . For $K = 1$, the joint estimating equations are given by

$$\begin{aligned} & \sum_{i=1}^n \left(\begin{array}{c} \mathbf{H}_{1,\beta} \\ A_1 \mathbf{H}_{1,\psi} \end{array} \right) w_1(\mathbf{H}_1, A_1; \xi_i) (Y - \mathbf{H}_{1,\beta}^T \beta_1 - A_1 \mathbf{H}_{1,\psi}^T \psi_1) = 0 \\ & \sum_{i=1}^n \left(\begin{array}{c} 1 \\ \mathbf{H}_1 \end{array} \right) \left\{ A_1 - \frac{\exp(\xi_{i1} + \mathbf{H}_1^T \xi_{i2})}{1 + \exp(\xi_{i1} + \mathbf{H}_1^T \xi_{i2})} \right\} = 0. \end{aligned}$$

The standard sandwich variance estimator that does not adjust for the propensity score estimation performs well in practice (Wallace et al., 2017). Denoting the variance estimator by $\Sigma_{\widehat{\psi}_1}$, $\mathfrak{Z}_\epsilon = \{\psi_1 : n(\psi_1 - \widehat{\psi}_1)^T \Sigma_{\widehat{\psi}_1}^{-1} (\psi_1 - \widehat{\psi}_1) \leq \chi_{1-\epsilon, p_1}^2\}$ is a Wald-type asymptotic $(1 - \epsilon) \times 100\%$ confidence region for ψ_1 .

When $K > 1$, as previously mentioned, the estimator for ψ_k when $k < K$ is nonregular due to the nonsmoothness of the pseudo-outcome. One potential solution to constructing a valid confidence set for ψ is to use a projection region. Define $\widetilde{Y}_k(\underline{\psi}_{k+1}) = Y + \sum_{j=k+1}^K \mathbf{h}_{j,\psi}^T \psi_j \{ \mathbb{I}(\mathbf{h}_{j,\psi}^T \psi_j > 0) - a_j \}$ so $\widetilde{Y}_k(\underline{\psi}_{k+1})$ is equivalent to the pseudo-outcome at stage k if $\underline{\psi}_{k+1} = \widehat{\psi}_{k+1}$. For $k = 1, \dots, K - 1$ define

$$\psi_k^*(\underline{\psi}_{k+1}) = \arg \min_{\psi_k} \mathbb{E} \left[w_k(\mathbf{H}_k, A_k) \{ \widetilde{Y}_k(\underline{\psi}_{k+1}) - \mathbf{H}_{k,\beta}^T \beta_k - A_k \mathbf{H}_{k,\psi}^T \psi_k \}^2 \right]$$

so that $\psi_k^*(\underline{\psi}_{k+1})$ denotes the population-level parameter for the blip-to-zero model if we know that $\underline{\psi}_{k+1}^* = \underline{\psi}_{k+1}$. Therefore, we also have that $\psi_k^* = \psi_k^*(\underline{\psi}_{k+1}^*)$. Define an estimator for $\psi_k^*(\underline{\psi}_{k+1})$ to be given by

$$\hat{\psi}_k(\underline{\psi}_{k+1}) = \arg \min_{\psi_k} \mathbb{P}_n \left[w_k(\mathbf{H}_k, A_k; \hat{\xi}) \left\{ \tilde{Y}_k(\underline{\psi}_{k+1}) - \mathbf{H}_{k,\beta}^T \beta_k - A_k \mathbf{H}_{k,\psi}^T \psi_k \right\}^2 \right].$$

This estimator is the weighted least squares estimator used in dWOLS with the pseudo-outcome replaced with $\tilde{Y}_k(\underline{\psi}_{k+1})$. Let $\Sigma_{\hat{\psi}_k}(\underline{\psi}_{k+1})$ denote the variance of $\hat{\psi}_k(\underline{\psi}_{k+1})$. Then

$$\mathfrak{Z}_{k,n,\epsilon}(\underline{\psi}_{k+1}) = \left\{ \psi_k : n(\psi_k - \hat{\psi}_k)^T \Sigma_{\hat{\psi}_k}^{-1}(\underline{\psi}_{k+1})(\psi_k - \hat{\psi}_k) \leq \chi_{1-\epsilon, pk}^2 \right\}$$

gives a $(1 - \epsilon) \times 100\%$ Wald-type asymptotic confidence region for $\psi_k^*(\underline{\psi}_{k+1})$. Then, given $\epsilon_1, \dots, \epsilon_K \in (0, 1)$ such that $\vartheta = \sum_{k=1}^K \epsilon_k \leq 1$,

$$\Psi_{n,\vartheta} = \left\{ \psi : \psi_K \in \mathfrak{Z}_{K,n,\epsilon_K} \text{ and } \psi_k \in \mathfrak{Z}_{k,n,\epsilon_k}(\underline{\psi}_{k+1}) \text{ for } k = 1, \dots, K-1 \right\}$$

represents a $(1 - \vartheta) \times 100\%$ confidence region for ψ^* . This approach to creating a confidence region for ψ will be increasingly conservative as K increases. Also, recall that our hypothesis test involves finding $\inf_{\psi \in \Psi_{n,1-\vartheta_1}} \left\{ \widehat{V}_n(\psi) - \frac{z_{1-\vartheta_2} \hat{\sigma}_n(\psi)}{\sqrt{n}} \right\}$. Note that this is a constrained optimization problem with ψ constrained by $\Psi_{n,1-\vartheta_1}$. For the projection region, $\Psi_{n,1-\vartheta_1}$, we have that ψ_K is constrained by $\mathfrak{Z}_{K,n,\epsilon_K}$, but for values of $k < K$ we have that the constraint depends on the value of $\underline{\psi}_{k+1}$. This makes performing the optimization very computationally difficult and infeasible for large values of K . We will instead focus on a bootstrap-based method for forming a confidence region that does not have the theoretical guarantees of the projection region, but has been found to perform well in practice (refer to Section B of the online Supplementary Material).

We can construct a valid confidence set using the m -out-of- n bootstrap, a tool for producing valid confidence intervals for nonsmooth functionals (Swanepoel, 1986; Dumbgen, 1993; Shao, 1994; Bickel et al., 1997). The m -out-of- n bootstrap uses a resampling size m that is smaller than the sample size n . Chakraborty et al. (2013) used the m -out-of- n bootstrap to create valid confidence intervals for the parameters indexing a DTR when estimating an optimal regime using Q-learning as well as an adaptive method to select m . Simoneau et al. (2018) examined using this procedure to create valid confidence intervals when using dWOLS to estimate the optimal regime. Both papers focused on two-stage DTRs so the estimator for only the first stage has a nonregular limiting distribution. We propose a method for generalizing this procedure to a K -stage DTR in which the estimators in all $k = 1, \dots, K-1$ stages suffer from nonregularity.

We first discuss how Chakraborty et al. (2013) proposed using the m -out-of- n bootstrap to construct confidence intervals for stage-one parameters for a two-stage DTR. Define

$\rho \triangleq P(\mathbf{H}_{2,\psi}^T \boldsymbol{\psi}_2 = 0)$, so that ρ is a measure of the degree of nonregularity in the data. When $\rho = 0$, the distribution of $\sqrt{n}(\hat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1)$ is asymptotically normal and the standard bootstrap will produce valid confidence intervals. Chakraborty et al. (2013) proposed using a resample size of $\hat{m} \triangleq n \frac{1 + \kappa(1 - \hat{\rho})}{1 + \kappa}$, where $\kappa > 0$ is a tuning parameter and $\hat{\rho}$ is an estimate of ρ . When $\rho = 0$, we have that $m = n$ and as ρ increases, our resample size will decrease, while κ determines the smallest acceptable resample size with m taking values within the interval $[n \frac{1}{1 + \kappa}, n]$. Chakraborty et al. proposed using a plug-in estimator for ρ given by $\rho = \mathbb{P}_n \mathbb{I}\{n(\mathbf{H}_{2,\psi}^T \hat{\boldsymbol{\psi}}_2)^2 \leq \tau_n(\mathbf{H}_{2,\psi})\}$, where $\tau_n(\mathbf{h}_{2,\psi})$ is given by $\mathbf{h}_{2,\psi}^T \hat{\Sigma}_{\hat{\boldsymbol{\psi}}_2} \mathbf{h}_{2,\psi} \chi_{1,1-\nu}^2$ such that $\hat{\Sigma}_{\hat{\boldsymbol{\psi}}_2}$ is the plug-in estimator of $n\text{Cov}(\hat{\boldsymbol{\psi}}_2, \hat{\boldsymbol{\psi}}_2)$. Let $\hat{\boldsymbol{\psi}}_{1,\hat{m}}^{(b)}$ denote the bootstrap estimate for $\boldsymbol{\psi}_1$ from using a resample size of \hat{m} . To construct a $(1 - \vartheta) \times 100\%$ confidence interval for $\boldsymbol{\psi}_1$, calculate the $\vartheta/2 \times 100$ and $(1 - \vartheta/2) \times 100$ percentiles of $\sqrt{\hat{m}}(\hat{\boldsymbol{\psi}}_{1,\hat{m}}^{(b)} - \hat{\boldsymbol{\psi}}_1)$, which we denote by \hat{l} and \hat{u} , respectively. Then a $(1 - \vartheta) \times 100\%$ confidence interval for $\boldsymbol{\psi}_1$ is given by $(\hat{\boldsymbol{\psi}}_1 - \hat{u}/\sqrt{\hat{m}}, \hat{\boldsymbol{\psi}}_1 - \hat{l}/\sqrt{\hat{m}})$.

To generalize to a K stage DTR, we start by defining $\rho_k \triangleq P(\mathbf{H}_{k+1,\psi}^T \boldsymbol{\psi}_{k+1} = 0)$. Therefore, ρ_k indicates the degree of nonregularity in the estimation of $\boldsymbol{\psi}_k$ at stage k . The plug-in estimator for ρ_{K-1} can be calculated using $\hat{\Sigma}_{\hat{\boldsymbol{\psi}}_K}$. For ρ_k for $k = 1, \dots, K - 2$, the nonregularity will cause the usual plug-in estimator to no longer be valid. Instead, we use the $m - \text{out} - \text{of} - n$ bootstrap to construct a valid confidence interval for $\mathbf{h}_{k,\psi}^T \boldsymbol{\psi}_k$. The estimator $\hat{\rho}_k$ for ρ_k is then given by the proportion of individuals in the sample for which the confidence interval for $\mathbf{h}_{k,\psi}^T \boldsymbol{\psi}_k$ contains zero. We then move backwards through the stages, getting an estimate of $\hat{\rho}_k$ at each stage using the $m - \text{out} - \text{of} - n$ bootstrap with a resample size of $\hat{m}_k = n \frac{1 + \kappa(1 - \hat{\rho}_k)}{1 + \kappa}$ at each stage. Let $\hat{\rho} = \max_k \hat{\rho}_k$. We calculate \hat{m} using the same formula as before and use this as our resample size. We calculate the $\epsilon_k/2 \times 100$ and $(1 - \epsilon_k/2) \times 100$ percentiles of $\sqrt{\hat{m}_k}(\hat{\boldsymbol{\psi}}_{k,\hat{m}_k}^{(b)} - \hat{\boldsymbol{\psi}}_k)$, which we denote by \hat{l}_k and \hat{u}_k for each value of k . Then, given $\epsilon_1, \dots, \epsilon_K \in (0, 1)$, such that $\vartheta = \sum_{k=1}^K \epsilon_k \leq 1$, a $(1 - \vartheta) \times 100\%$ confidence region for $\boldsymbol{\psi}^*$ is given by

$$\Psi_{n,\vartheta} = \{\boldsymbol{\psi}: \boldsymbol{\psi}_k \in (\hat{\boldsymbol{\psi}}_k - \hat{u}_k/\sqrt{\hat{m}_k}, \hat{\boldsymbol{\psi}}_k - \hat{l}_k/\sqrt{\hat{m}_k}) \text{ for } k = 1, \dots, K - 1 \text{ and } \boldsymbol{\psi}_K \in \mathfrak{Z}_{K,n,\epsilon_K}\}.$$

Section B of the online Supplementary Material contains simulations demonstrating the coverage of confidence intervals generated using this procedure when applied to $K = 3$ stage DTRs and a data-driven approach to selecting κ using the double-bootstrap.

3.4 Bootstrap Power Calculations

We estimate the power for a given sample size using a bootstrap of pilot data, i.e., resampling from the pilot data samples of size n and assessing power with that sample size over a grid of candidate sample sizes $n \in \mathbb{N}$. We assume pilot data $\mathcal{D}_{n_0} = \{(\mathbf{X}_{1,i}, A_{1,i}, \dots, \mathbf{X}_{K,i}, A_{K,i}, Y_i)\}_{i=1}^{n_0}$ that comprises n_0 i.i.d. replicates from the same population of interest as the full study. To conduct sample size calculations, we search for the smallest sample size n for which the estimated power exceeds the threshold given in condition (PWR).

Let $\mathbb{P}_{n_0, n}^{(b)}$ denote the empirical bootstrap distribution for a resample size n from a pilot study of size n_0 . For any functional $Z_n = f(\mathbb{E}, \mathbb{P}_{n_0})$, define the bootstrap equivalent by $Z_{n_0, n}^{(b)} = f(\mathbb{P}_{n_0}, \mathbb{P}_{n, n_0}^{(b)})$. Let P_B denote probabilities computed with respect to the bootstrap distribution conditional on the pilot data. An estimate of the power for a given sample size, n , is then given by

$$P_B \left\{ \inf_{\psi \in \Psi_{n_0, n, 1 - \vartheta_1}^{(b)}} \left(\frac{\sqrt{n} \{ \widehat{V}_{n_0, n}^{(b)}(\psi) - \widehat{V}_{n_0}(\psi) \}}{\widehat{\zeta}_{n_0, n}^{(b)}(\psi)} + \frac{\min[\sqrt{n} \{ \widehat{V}_{n_0}(\psi) - B_0 \}, \sqrt{n\eta}]}{\widehat{\zeta}_{n_0, n}^{(b)}(\psi)} \right) \geq z_{1 - \vartheta_2} \right\}$$

such that $\vartheta_1 + \vartheta_2 = \alpha$. It is recommended to set ϑ_1 to be relatively small compared to ϑ_2 (Berger and Boos, 1994). The bootstrap estimator of the minimum sample size required to satisfy condition (PWR) is given by the smallest n that satisfies

$$P_B \left\{ \inf_{\psi \in \Psi_{n_0, n, 1 - \vartheta_1}^{(b)}} \left(\frac{\sqrt{n} \{ \widehat{V}_{n_0, n}^{(b)}(\psi) - \widehat{V}_{n_0}(\psi) \}}{\widehat{\zeta}_{n_0, n}^{(b)}(\psi)} + \frac{\min[\sqrt{n} \{ \widehat{V}_{n_0}(\psi) - B_0 \}, \sqrt{n\eta}]}{\widehat{\zeta}_{n_0, n}^{(b)}(\psi)} \right) \geq z_{1 - \vartheta_2} \right\} \geq 1 - \phi.$$

Rose et al. (2019) proved that a bootstrap oversampling estimator of this form is consistent as n_0 and n diverge under mild assumptions. Here, the form of $\widehat{V}_{n_0}(\psi)$ is different, requiring slightly different assumptions. If we assume:

$$(A1) \inf_{\psi} \mathbb{E}[Y(\psi) - \mathbb{E}\{Y(\psi)\}]^2 > 0 \text{ and } \sup_{\psi} \mathbb{E}[Y(\psi) - \mathbb{E}\{Y(\psi)\}]^2 < \infty;$$

$$(A2) \text{ the classes } \mathcal{F}_1\{Y(\psi): \psi \in \Theta\} \text{ and } \mathcal{F}_2\{Y^2(\psi): \psi \in \Theta\} \text{ are Donsker;}$$

$$(A3) \mathbb{E}\{Y(\psi)\} \text{ is uniformly continuous in a neighborhood of } \psi^*;$$

we have that consistency holds and the proof then follows that of Rose et al. (2019).

Now we focus on determining sample sizes for the (OPT) condition. Recall that this condition states that $P\{V(\widehat{\mathbf{d}}_n) \geq V(\mathbf{d}^{OPT}) - \epsilon\} \geq 1 - \zeta + o(1)$. Note that for any sequence $\tilde{\psi}_n \in \Psi_{n, 1 - \vartheta}$ such that $\widehat{V}_n(\psi^*) \leq \widehat{V}_n(\tilde{\psi}_n) + o_P(1/\sqrt{n})$ we have that

$$P \left[V(\tilde{\psi}_n) \geq V(\mathbf{d}^{OPT}) + \inf_{\psi \in \Psi_{n, 1 - \vartheta_1}} \{ \widehat{V}_n(\psi) - V(\psi) \} - \sup_{\psi \in \Psi_{n, 1 - \vartheta_1}} \{ \widehat{V}_n(\psi) - V(\psi) \} \right] \geq 1 - \vartheta_1 + o(1).$$

Then if $\mathfrak{Q}_{n, 1 - \vartheta_2, 1 - \vartheta_1}$ is the $(1 - \vartheta_2)$ th quantile of

$$\sup_{\psi \in \Psi_{n, 1 - \vartheta_1}} \{ \widehat{V}_n(\psi) - V(\psi) \} - \inf_{\psi \in \Psi_{n, 1 - \vartheta_1}} \{ \widehat{V}_n(\psi) - V(\psi) \},$$

(OPT) holds asymptotically if $\vartheta_1 + \vartheta_2 \leq \zeta$ and $\mathfrak{Q}_{n, 1 - \vartheta_1, 1 - \vartheta_2} \leq \epsilon$. We again use bootstrap oversampling to estimate the smallest n such that this holds. Let $\mathfrak{Q}_{n_0, n, 1 - \vartheta_2, 1 - \vartheta_1}^{(b)}$ be the bootstrap estimate of $\mathfrak{Q}_{n, 1 - \vartheta_2, 1 - \vartheta_1}$ from a pilot study of size n_0 with a resample size of n . Then the

estimate of our sample size is given by the smallest n such that $\mathfrak{D}_{n_0, n, 1 - \theta_2, 1 - \theta_1}^{(b)} \leq \epsilon$. To calculate a sample size that satisfies both conditions simultaneously, we recommend calculating a sample size for each condition individually and using the max of the two.

4. Simulation Study

We examined the finite sample performance of our proposed method for conducting power and sample size calculations with a simulation study. We considered sizing and conducting power calculations for a three-stage study with two treatment options at each stage. To evaluate the performance of our method for sizing a study, we conducted simulations for each of the two conditions (PWR) and (OPT) individually. Section C of the online Supplementary Material contains additional simulations for power calculations and sizing a two-stage study. The data generating model for our simulations was:

$$\begin{aligned} X_1 &\sim N(0, 1), & P(A_k = 1 \mid \mathbf{H}_k = \mathbf{h}_k) &= \left\{ 1 + e^{-(\varpi_{k,0} + \varpi_{k,1}X_k)} \right\}^{-1} \text{ for } k = 1, 2, 3, \\ \tau_1 &\sim N(0, 1), & X_2 &= \mu_{20} + \mu_{21}X_1 + \tau_1, \\ \tau_2 &\sim N(0, 1), & X_3 &= \mu_{30} + \mu_{31}X_1 + \mu_{32}X_2 + \tau_2, \\ \mathbf{H}_{3,1}^T &= (1, X_1, X_2, X_3), & \mathbf{H}_{3,0}^T &= (1, X_1, A_1, A_1X_1, X_2, A_2, A_2X_1, A_2X_2, X_3, X_1^2) \\ v &\sim N(0, 1), & Y &= \mathbf{H}_{3,0}^T \lambda_{3,0} + A_3 \mathbf{H}_{3,1}^T \lambda_{3,1} + v. \end{aligned}$$

The parameters of the data generating model were given by:

$$\begin{aligned} \varpi_1 &= (0.25, 1), & \varpi_2 &= (0.25, 1, -1, -1), \\ \varpi_3 &= (0.25, 0.5, 0.5, -0.5, 1, -0.5), \\ \mu_2 &= (0, 0.5), & \mu_3 &= (0, -0.5, 0.5), \\ \lambda_{3,0} &= (1, 1, 0.5, -0.75, 0.5, -0.5, -0.5, 0.5, 0.5, 0.25), & \lambda_{3,1} &= (0.25, 0.5, 0.5, -0.5). \end{aligned}$$

We posited models such that the blip model at each stage was correctly specified, but both treatment-free models were misspecified by leaving out X_1^2 . We modeled the propensity score with a correctly specified logistic regression model so that dWOLS produced consistent estimates of the blip parameters.

For the simulation study examining our proposed method for estimating the power for a given sample size, we let $\alpha = 0.05$ and $\eta = 1.4$. Therefore, we calculated the power for a 0.05 level test of $H_0: V(\mathbf{d}^{\text{opt}}) \leq B_0$. We evaluated the performance of the procedure when the effect size of tailoring was equal to $\eta = 1.4$ and examined how it changed as the effect size increased. An effect size of $\eta = 1.4$ corresponded to a standardized effect size of 0.72, which is relatively moderate (Cohen, 1992). We let $V(\mathbf{d}^{\text{opt}}) = B_0 + \eta + \Delta\eta$ and varied $\Delta \in \{0, 0.25, 0.5\}$. The data generating model was fixed across all settings, which caused $V(\mathbf{d}^{\text{opt}})$ to be fixed, so we let B_0 vary with Δ such that $B_0 = V(\mathbf{d}^{\text{opt}}) - \Delta\eta - \eta$. We let the size of the pilot study vary such that $n_0 \in \{200, 400\}$ and estimated the power for a set of different sample sizes given by $n \in \{250, 500, 750\}$. For each sample size, we also estimated the true power via simulation. Each of the 500 repetitions of the simulation study involved simulating a pilot study of size n_0 and using the proposed method to estimate the power for

each of the sample sizes n . We also repeatedly simulated studies of size n and conducted the proposed hypothesis test to calculate the true power.

Table 1 contains the mean, median, and standard deviation of the estimated power across the 500 repetitions for each combination of the sample and effect sizes. The average estimated power was close to the true power for most settings of the simulation. The settings in which the mean differed from the true power were due to the distribution of the estimated power truncating at 1 and 0. This caused the mean to underestimate the power when the true power was close to 1, with the median remaining close to the true power. This was observed in the case of $n = 750$ and $\Delta = 0$, which had a true power of 0.98, to have a mean and median for the estimated power of 0.84 and 0.99, respectively. As Δ increased, the variability in the estimated power decreased. Increasing the size of the pilot study from 200 to 400 also caused the standard deviation in the estimated powers to decrease. As the variability decreased, the mean of the estimated powers moved closer to the true power, even when the true power was close to 1.

For the simulations using our proposed procedure for sample size calculations, we assumed $\alpha = 0.05$, $\phi = 0.1$, and $\eta = 1.4$. Therefore, the first condition (PWR) held if we had a 0.05 level test of $H_0: V(\mathbf{d}^{\text{opt}}) \leq B_0$ that had power 90%, provided $V(\mathbf{d}^{\text{opt}}) \geq B_0 + 1.4$. We again evaluated the performance of the sample size procedure when the effect size of tailoring was equal to $\eta = 1.4$ and examined how it changed as the effect size increased by varying $\Delta \in \{0, 0.25, 0.5, 1\}$. We let $\zeta = 0.1$ and varied $\epsilon \in \{0.3, 0.5, 0.7\}$. Therefore, the second condition (OPT) held if $P\{V(\hat{\mathbf{d}}_n) \geq V(\mathbf{d}^{\text{opt}}) - \epsilon\} \geq 0.9$.

Each repetition of the simulation study consisted of the following steps. First, we generated a pilot study of size $n_0 \in \{200, 400\}$. Second, we estimated the power on a grid of potential sample sizes using 500 bootstrap repetitions for each sample size considered. To construct a confidence set for ψ , we used the $m - \text{out} - \text{of} - n$ bootstrap with $\kappa = 0.2$. Third, we used least squares to regress the estimated power on the potential sample sizes and used the fitted model to estimate the smallest sample size, $\hat{n}(\mathcal{D}_{n_0})$, that achieved the desired power of 90%. We fit this model using only the tested sample sizes that resulted in an estimated power in a small neighborhood of the targeted power, as the power curve will be approximately linear only within a small region. The possibility existed that the estimated value from the pilot study would be less than the comparison mean such that $\hat{V}_{n_0}(\hat{\psi}) \leq B_0$. When that occurred, we would not be able to find a sample size using that pilot that would be powered for a comparison with B_0 , and we define $\hat{n}(\mathcal{D}_{n_0}) = \infty$. Fourth, for each $\hat{n}(\mathcal{D}_{n_0}) < \infty$, we generated a study of size $\hat{n}(\mathcal{D}_{n_0})$ and performed a hypothesis test for condition (PWR) calculating the empirical power over the 500 simulations using this process. For condition (OPT), we estimated the optimal regime using the study of size $\hat{n}(\mathcal{D}_{n_0})$ and calculated the true value of the estimated regime using the known data generating functions. Last, we checked whether this value was within ϵ of the value of the true optimal regime. Section D of the online Supplementary Material contains high-level pseudocode for the simulation study.

Tables 2 and 3 contain the sizing results for (PWR) and (OPT), respectively. The comparison to B_0 was slightly underpowered when $\Delta = 0$ and $n_0 = 200$ with a power of 74.39%. This is partly due to the size of the pilot study, as we can see that as the pilot size increased to $n_0 = 400$, the power increased to 82.68%. Increasing the size of the pilot also decreased the variance in the estimated sample size. As Δ increased, the power increased as expected and converged to 100%. For $\Delta = 0$, the distribution of $\hat{n}(\mathcal{D}_{n_0})$ was right skewed as expected; generally, pilot studies in which the estimated benefit to tailoring is very small will result in a large estimated sample size. This also caused the variance of the estimated sample size to increase. In general, this will occur when using the proposed method when η is small and $V(d^{\text{opt}})$ is close to $B_0 + \eta$. As Δ increased, the degree of skewness and variance in the estimated sample size declined. Table 3 shows that the nominal concentration of 90% was achieved for all values of ϵ and n_0 . This procedure is conservative, as we can see that the concentration was 100% for all simulation settings. As ϵ decreased, the estimated sample size increased as expected. We also see that as we increased the size of the pilot study, the variance in the sample size estimate decreased. The simulation results from the two-stage study in Section C of the online Supplementary Material were very similar.

5. Illustration Using Data Gathered from EHRs

Kaiser Permanente Washington is a health system providing both clinical care and health insurance to members. This study used data extracted from electronic health records and health insurance claims from KPWA clients. The data for this study consisted of records of 82,691 patients who began antidepressant treatment for depression from 2008 through 2018 and included information on demographics, prior diagnoses of mental health conditions, prescription fills, and depressive symptoms as measured by patient report with the Patient Health Questionnaire (PHQ) (Kroenke et al., 2001). To be included in the study, patients had to be 13 years or older; be enrolled in KPWA for the past year; have a diagnosis of a depressive disorder in the 365 days before or 15 days after treatment initiation; have no antidepressant prescription fills in the prior year (excluding trazodone, doxepin, and amitriptyline, which are primarily used to treat conditions other than depression); and have no diagnoses of a personality, bipolar, or psychotic disorder in the past year. We used a subset of the data to represent a pilot study to conduct power and sample size calculations for estimating an optimal DTR to minimize depression symptoms in this population.

The PHQ-9 score is a measure of the severity of depressive symptoms that is used in the KPWA health system for diagnosing depression and monitoring depressive systems (Kroenke et al., 2001). The first 8 questions yield a score that ranges from 0 to 24, with higher values indicating more severe symptoms. The outcome of interest for this study was the negative of the PHQ score after 1 year of treatment, to be consistent with our framework that higher values correspond with better patient outcomes. The PHQ score after 1 year was defined as the PHQ score recorded closest to exactly one year after starting treatment, but was required to be recorded between 305 and 425 days after beginning treatment. Initially, patients received one of 17 different antidepressants. The first-stage treatment was classified as an antidepressant from the selective serotonin reuptake inhibitor (SSRI) class or an antidepressant from a different class. Section E of the online Supplementary

Material contains a list of all the antidepressants assigned. A total of 63,060 patients received an SSRI at treatment initiation while 19,657 received an antidepressant from an alternative class. As this study followed people receiving their regular healthcare, patients were observed to switch to a different antidepressant or augment their initial treatment with an additional antidepressant or an antipsychotic. Our goal was to estimate an optimal DTR that tailored treatment based on age, gender, baseline PHQ score, and diagnosis of an anxiety disorder in the past year.

The data set had a significant amount of missingness and censoring. The baseline PHQ score, defined as a score recorded 15 days or fewer before treatment initiation or up to 3 days after, was observed for 34,541(41.8%) of patients in the sample. Of those, 8,757(25.35%) were censored due to disenrolling from the health system during the first year or were administratively censored because the study ended less than a year after the patient started treatment. Of the remaining patients, 8,511 (24.6%) had an observed PHQ at one year after treatment initiation. We also artificially censored follow-up if patients discontinued treatment or changed treatment more than once during the first year. Our final sample size for potential pilot studies was 2,008. We constructed a pilot study by taking a simple random sample from these 2,008 remaining patients. Because follow-up data were missing and not everyone followed a regime of interest, our sample pilot study might not have a representative sample of the population. This could lead to bias in the estimated optimal regime from the pilot. Since the purpose of this analysis was to demonstrate how to conduct power calculations and size a study, we assumed our pilot was representative and did not use any methods to adjust for the potential bias. Due to the potential for selection bias, the estimated regime from this pilot should not be used for any clinical interpretation in practice. We will consider this point further in the discussion.

Define the outcome, Y , as the negative PHQ score 1 year after initiating treatment. Let A_1 denote the first-stage treatment, such that $A_1 = 1$ if prescribed an SSRI and $A_1 = 0$ if assigned a non-SSRI. Let A_2 denote the second-stage treatment, with $A_2 = 1$ if the patient switched treatment and $A_2 = 0$ if the patient augmented treatment with an additional antidepressant or antipsychotic while staying on the initial medication. We use \mathbf{X}_1^p to denote a vector of age, gender, indicator variables for race, baseline PHQ score, and an indicator for the diagnosis of an anxiety disorder in the past 365 days. Let \mathbf{X}_1^w denote the same vector of patient characteristics with race removed, as we did not consider tailoring treatment based on race. We posited the following models for the treatment-free and blip functions:

$$\begin{aligned} \gamma_1(\mathbf{h}_1, a_1; \boldsymbol{\psi}_1) &= a_1(\psi_{1,0} + \boldsymbol{\psi}_{1,1}^T \mathbf{x}_1^w), & \gamma_2(\mathbf{h}_2, a_2; \boldsymbol{\psi}_2) &= a_2(\psi_{2,0} + \boldsymbol{\psi}_{2,1}^T \mathbf{x}_1^w + \psi_{2,2} a_1), \\ g_1(\mathbf{h}_1; \boldsymbol{\beta}_1) &= \beta_{1,0} + \boldsymbol{\beta}_{1,1}^T \mathbf{x}_1^p, & g_2(\mathbf{h}_2; \boldsymbol{\beta}_2) &= \beta_{2,0} + \boldsymbol{\beta}_{2,1}^T \mathbf{x}_1^p + \beta_{2,2} a_1. \end{aligned}$$

The propensity score models, $\pi_k(\mathbf{h}_k; \boldsymbol{\xi}_k)$ for $k = 1, 2$, were estimated with logistic regression using the same set of variables as the treatment-free models.

We included a random sample of 400 patients in the pilot study. We assumed that the comparison mean was given by $B_0 = -10$, as a PHQ score of greater than or equal to 10 is used to identify moderate depression. We let $\alpha = 0.05$, $\phi = 0.1$, and $\eta = 3$ so that the

first condition (PWR) held if we had a 0.05 level test of $H_0: V(\mathbf{a}^{\text{opt}}) \leq B_0$ that had 90% power if $V(\mathbf{a}^{\text{opt}}) \geq B_0 + 3$. We let $\epsilon = 1$ and $\zeta = 0.1$ so that the second condition (OPT) held if $P\{V(\hat{\mathbf{d}}_n) \geq V(\mathbf{a}^{\text{opt}}) - 1\} \geq 0.9$. A confidence set for ψ was constructed using the m – out – of – n bootstrap with $\kappa = 0.2$.

Table 4 shows the estimated coefficients for the second-stage and first-stage blip models from the pilot data. The estimated value of the optimal regime in the pilot was given by $\hat{V}_{n_0}(\hat{\mathbf{d}}_{n_0}) = -6.67$. Therefore, we found some evidence that an adaptive treatment strategy could be effective in reducing depressive symptoms, but this difference was not statistically significant in the pilot data. Note the wide confidence intervals for the parameters of the blip models. We first applied our method to estimate the power for comparing the value of the optimal regime to B_0 , if all 2,008 patients were included in the study. This resulted in an estimated power of 43.8%. Therefore, the full study was underpowered for detecting any benefit to tailoring treatment. To be able to detect a benefit to tailoring treatment, we would need to increase the study size by including additional study sites or increasing the range of years under consideration. We applied our proposed sample size method to power the comparison of the value of the optimal regime to B_0 (PWR), resulting in a sample size of $\hat{n}(\mathcal{D}_{n_0}) = 5,230$. When sizing to guarantee the value of the estimated regime was close to the value of the true optimal regime (OPT), we calculated a sample size of $\hat{n}(\mathcal{D}_{n_0}) = 4,276$. Therefore, for both conditions to hold, we recommend a study of size at least $\hat{n}(\mathcal{D}_{n_0}) = 5,230$.

6. Discussion

We propose a method for conducting power calculations for a K -stage longitudinal observational study for estimating DTRs using a pilot study. The method is based on bootstrapping a projection interval for the value of the optimal regime. We implemented the bootstrap with oversampling to estimate the power for a given sample size and used our method to size a study by calculating the smallest sample size that achieved the desired power. We demonstrated this method attains the desired power in finite samples in a simulation study. We also propose a method for extending the m – out – of – n bootstrap to multistage DTRs to obtain valid confidence intervals for the parameters indexing the treatment regime.

As in any realistic planning of an analysis, additional sources of variability or loss of power must be considered. Examples include the possibility of missing information, measurement error, additional unmeasured confounding, and more. In a randomized trial setting, such additional concerns are sometimes addressed in sensitivity analyses. For the case of missing data, for example, the researchers may assume a specific rate of missingness or withdrawal of consent, and inflate the sample size (equivalently, in our setting, decrease the power) accordingly (Hsieh et al., 2003).

We implicitly assumed throughout this work that the pilot data are drawn from the same population as the data in which the full analysis will be conducted and, more generally, to which the results will be applied. This is quite a realistic assumption, as observational

analyses tend to be done within a given ‘system’, e.g., a particular healthcare system such as a national or provincial health service or within a given health care management organization. It may be the case, as noted in the introduction, that only a small dataset (the “pilot data”) are initially available due to the cost of extracting certain information, such as free-text box fields or analyzing stored blood specimens. Alternatively, it may be that the power/sample size calculations encourage a wider collaboration with other similar health systems (in Canada, this might be collaboration across provinces; in the context of a healthcare management organization, this might mean pooling of data from different centers within the same organization).

However, should the pilot be drawn from a population that differs from the target population, it may be possible to leverage the resampling used in the power calculations to our advantage. The weighted bootstrap has been used as a method to alter the bootstrap empirical distribution. Hall et al. (2008) proposed using a weighted bootstrap to align the bootstrap-weighted empirical distributions of covariates between treatment groups to more effectively compare the treatment response between the groups. To account for differences in the distribution of the data between the pilot and full study, we could use a weighted bootstrap to induce a shift in the distribution of the bootstrap samples to align with the postulated distribution of the data in the full study. This could also be conducted as a sensitivity analysis to examine how sensitive the power is to changes in the data distribution by repeating this procedure with multiple different bootstrap weights. The performance of this approach and the impact of such population changes or “distribution shifts” warrant additional research.

This paper focused on DTRs estimated via dWOLS and assumed the blip models were correctly specified. This method could be easily adapted to other regression-based estimation methods. Value-search or direct-search estimators are an alternative class of estimators for identifying optimal treatment regimes that are frequently used (Orellana et al., 2010; Zhao et al., 2012; Laber and Zhao, 2015). We leave extensions to this class of estimators for future work. In this paper, we focused on discrete treatments with continuous outcomes. Extensions to other outcome and/or treatment types, such as survival outcomes and continuous doses, are possible, but will require careful thought.

A pilot study may include potential sources of bias, in addition to confounding, due to censoring or missing data. The resulting sample size calculations could be adjusted to account for bias by using some form of sample size inflation factor or decrease in effective sample size based on the level of censoring/missingness and how informative it is. Similarly, a sample size inflation factor can be used to deflate the sample size of interest when estimating the power for a given sample size. Multiple imputation with bootstrapping has also been used for inference (Schomaker and Heumann, 2018; Bartlett and Hughes, 2020), and could be adjusted to use bootstrap oversampling to conduct sample size calculations in the presence of missing data.

Our proposed method relies on having access to pilot data. Unfortunately, such data are not always available. Sizing a study without pilot data would require much stronger assumptions about the underlying generative model, which we leave as future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R01 MH114873. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. EEMM is a Canada Research Chair (Tier 1) in Statistical Methods for Precision Medicine and acknowledges the support of a chercheur de mérite career award from the Fonds de Recherche du Québec, Santé.

References

- Artman William J, Nahum-Shani Inbal, Wu Tianshuang, Mckay James R, and Ertefaie Ashkan. Power analysis in a SMART design: sample size estimation for determining the best embedded dynamic treatment regime. *Biostatistics*, 21(3):432 – 448, 2020. [PubMed: 30380020]
- Bartlett Jonathan and Hughes Rachel. Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Statistical Methods in Medical Research*, 29(12): 3533–3546, 2020. [PubMed: 32605503]
- Berger Roger L and Boos Dennis D. P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, 89(427):1012–1016, 1994.
- Bickel PJ, Gotze F, and Zwet W R van. Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7:1–31, 1997.
- Blatt D, Murphy Susan A., and Zhu J. A-learning for approximate planning. Technical Report 04–63, The Methodology Center, Pennsylvania State University, 2004.
- Campbell Harlan, Jong Valentijn M.T. de, Debray Thomas P.A., and Gustafson Paul. A few things to consider when deciding whether or not to conduct underpowered research. *Journal of Clinical Epidemiology*, 144:194–197, 2022. [PubMed: 34875377]
- Chakraborty Bibhas and Moodie Erica E. M.. *Statistical Methods for Dynamic Treatment Regimes*. New York: Springer, 2013.
- Chakraborty Bibhas and Murphy Susan A. Dynamic treatment regimes. *Annual Review of Statistics and Its Application*, 1:447 – 464, 2014.
- Chakraborty Bibhas, Laber Eric B, and Zhao Yingqi. Inference for optimal dynamic treatment regimes using an adaptive m – out – of – n bootstrap scheme. *Biometrics*, 69(3):714–723, 2013. [PubMed: 23845276]
- Chakraborty Bibhas, Ghosh Palash, Moodie Erica E. M., and John Rush A. Estimating optimal shared-parameter dynamic regimens with application to a multistage depression clinical trial. *Biometrics*, 72(3):865–876, 2016. [PubMed: 26890628]
- Cohen Jacob. A power primer. *Psychological Bulletin*, 112(1):155–159, 1992. [PubMed: 19565683]
- Coulombe Janie, Moodie Erica E. M., Shortreed Susan, and Renoux Christel. Can the risk of severe depression-related outcomes be reduced by tailoring the antidepressant therapy to patient characteristics? *American Journal of Epidemiology*, 2021. 10.1093/aje/kwaa260.
- Dumbgen Lutz. On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, 95:125–140, 1993.
- Hall Peter, Leng Xiaoyan, and Müller Hans-Georg. Weighted-bootstrap alignment of explanatory variables. *Journal of Statistical Planning and Inference*, 138(6):1817–1827, 2008.
- Hsieh FY, Lavori PW, Cohen HJ, and Feussner JR. An overview of variance inflation factors for sample-size calculation. *Evaluation & the Health Professions*, 26:239–257, 2003. [PubMed: 12971199]
- Krakow Elizabeth F, Hemmer Michael, Wang Tao, Logan Brent, Arora Mukta, Spellman Stephen, et al. Tools for the precision medicine era: how to develop highly personalized treatment recommendations from cohort and registry data using Q-learning. *American Journal of Epidemiology*, 186(2):160–172, 2017. [PubMed: 28472335]

- Kroenke Kurt, Spitzer Robert L., and Williams Janet B. W.. The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9):606 – 613, 2001. [PubMed: 11556941]
- Laber EB and Zhao YQ. Tree-based methods for estimating individualized treatment regimes. *Biometrika*, 102(3):501–514, 2015. [PubMed: 26893526]
- Laber Eric B, Lizotte Daniel J, Qian Min, Pelham William E, and Murphy Susan A. Dynamic treatment regimes: technical challenges and applications. *Electronic Journal of Statistics*, 8(1):1225 – 1272, 2014. [PubMed: 25356091]
- Laber Eric B, Zhao Ying-Qi, Regh Todd, Davidian Marie, Tsiatis Anastasios, Stanford Joseph B, Zeng Donglin, Song Rui, and Kosorok Michael R. Using pilot data to size a two-arm randomized trial to find a nearly optimal personalized treatment strategy. *Statistics in Medicine*, 35(8):1245–1256, 2016. [PubMed: 26506890]
- Lavori Philip W and Dawson Ree. Dynamic treatment regimes: practical design considerations. *Clinical trials*, 1(1):9–20, 2004. [PubMed: 16281458]
- Lavori PW and Dawson R. A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163(1):29–38, 2000.
- Lei H, Nahum-Shani I, Lynch K, Oslin D, and Murphy SA. A “SMART” design for building individualized treatment sequences. *Annual Review of Clinical Psychology*, 8: 21–48, 2012.
- Morris Tim P. and Smeden Maarten van. Causal analyses of existing databases: the importance of understanding what can be achieved with your data before analysis (commentary on hernán). *Journal of Clinical Epidemiology*, 142:261–263, 2022. [PubMed: 34560253]
- Murphy SA. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24(10):1455–1481, 2005. [PubMed: 15586395]
- Murphy Susan A. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.
- Oetting Alena I, Levy Janet A, Weiss Roger D, and Murphy Susan A. Statistical methodology for a SMART design in the development of adaptive treatment strategies. In Shrout Patrick E, Keyes Katherine M, and Ornstein Katharine, editors, *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, chapter 8, pages 179–205. Oxford University Press, New York, 2011.
- Orellana L, Rotnitzky A, and Robins J. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: Main content. *International Journal of Biostatistics*, 6(2):1–49, 2010.
- Robins JM. Optimal structural nested models for optimal sequential decisions. In Lin DY and Heagerty P, editors, *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, pages 189 – 326. New York: Springer, 2004.
- Rose Eric J., Laber Eric B., Davidian Marie, Tsiatis Anastasios A., Zhao Ying-Qi, and Kosorok Michael R.. Sample size calculations for SMARTs. arXiv: 1906.06646, 2019.
- Rubin DB. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, 6(1):34–58, 1978.
- Schomaker Micheal and Heumann Christian. Bootstrap inference when using multiple imputation. *Statistics in Medicine*, 37(14):2252–2266, 2018. [PubMed: 29682776]
- Seewald Nicholas J, Kidwell Kelley M, Nahum-Shani Inbal, Wu Tianshuang, McKay James R, and Almirall Daniel. Sample size considerations for comparing dynamic treatment regimens in a sequential multiple-assignment randomized trial with a continuous longitudinal outcome. *Statistical Methods in Medical Research*, 29(7):1891 – 1912, 2020. [PubMed: 31571526]
- Shao Jun. Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society*, 122(4):1251–1262, 1994.
- Shortreed Susan M. and Moodie Erica E. M.. Estimating the optimal dynamic antipsychotic treatment regime: evidence from the sequential multiple-assignment randomized clinical antipsychotic trials of intervention and effectiveness schizophrenia study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4):577 – 599, 2012. [PubMed: 23087488]

- Simoneau Gabrielle, Moodie Erica E M, Platt Robert, and Chakraborty Bibhas. Nonregular inference for dynamic weighted ordinary least squares: understanding the impact of solid food intake in infancy on childhood weight. *Biostatistics*, 19(2):233–246, 2018. [PubMed: 28968752]
- Simoneau Gabrielle, Moodie Erica E M, Azoulay Laurent, and Platt Robert W. Adaptive treatment strategies with survival outcomes: an application to the treatment of type 2 diabetes using a large observational database. *American Journal of Epidemiology*, 189 (5):461 – 469, 2020. [PubMed: 31903490]
- Jan WH Swanepoel. A note on proving that the (modified) bootstrap works. *Communications in Statistics-Theory and Methods*, 15:3193–3203, 1986.
- Tsiatis Anastasios A., Davidian Marie, Holloway Shannon T., and Laber Eric B.. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. CRC Press, Boca Raton, FL, 2019.
- Vaart Aad Van Der. *Asymptotic Statistics*. Cambridge University Press, 1998.
- Wallace Michael P and Moodie Erica E M. Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3):636–644, 2015. [PubMed: 25854539]
- Wallace Michael P, Moodie Erica E M, and Stephens David A. Model validation and selection for personalized medicine using dynamic-weighted ordinary least squares. *Statistical Methods in Medical Research*, 26(4):1641 – 1653, 2017. [PubMed: 28486872]
- Wang Lu, Rotnitzky Andrea, Lin Xihong, Millikan Randall E, and Thall Peter F. Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer. *Journal of the American Statistical Association*, 107(498):493–508, 2012. [PubMed: 22956855]
- Watkins CJCH and Dayan P. Q-learning. *Machine Learning*, 8(3):279–292, 1992.
- Zhao Yingqi, Zeng Donglin, Rush A John, and Kosorok Michael R. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012. [PubMed: 23630406]
- Zhao Yufan, Zeng Donglin, Socinski Mark A, and Kosorok Michael R. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics*, 67(4):1422–1433, 2011. [PubMed: 21385164]

Table 1:

Estimated power from the proposed power calculations using a pilot study of size n_0 for varying sample sizes n . We assume the effect size under the alternative hypothesis is given by $\eta = 1.4$. Δ denotes the difference between the true value of the optimal regime and B_0 which is given by $\eta(1 + \Delta)$, so $\Delta = 0$ corresponds to the true effect size being equal to η . The remaining columns display the mean, median, and standard deviation of the estimated powers across 500 simulated pilot studies as well as the true power which is calculated via simulation. We do not estimate the power for $n = 250$ when the pilot is of size $n_0 = 400$, since it is implausible for the pilot sample size to exceed the full study sample size.

Δ	n	n_0	True PWR	Mean PWR	Med PWR	SD PWR
0	250	200	0.42	0.54	0.54	0.32
0.25	250	200	0.99	0.96	0.99	0.09
0.5	250	200	1.00	1.00	1.00	0.01
0	500	200	0.83	0.77	0.92	0.30
0.25	500	200	1.00	0.99	1.00	0.04
0.5	500	200	1.00	1.00	1.00	0.00
0	750	200	0.98	0.84	0.99	0.27
0.25	750	200	1.00	1.00	1.00	0.02
0.5	750	200	1.00	1.00	1.00	0.00
0	500	400	0.82	0.77	0.85	0.24
0.25	500	400	1.00	1.00	1.00	0.01
0.5	500	400	1.00	1.00	1.00	0.00
0	750	400	0.98	0.89	0.98	0.19
0.25	750	400	1.00	1.00	1.00	0.00
0.5	750	400	1.00	1.00	1.00	0.00

Table 2:

Empirical power (PWR) using the projection-based sample size procedure at a nominal level of 90 using a pilot study of size $n_0 = 200$ and $n_0 = 400$; Δ denotes the difference between the true value of the optimal regime and B_0 which is given by $\eta(1 + \Delta)$. $P(\hat{n} = \infty)$ represents the proportion of pilot studies for which $\hat{n}(\mathcal{D}_{n_0}) = \infty$. The remaining columns give the mean, median, quartiles, and standard deviation of the estimated sample sizes across the 500 simulation repetitions.

Δ	n_0	$\mathbb{E}(\hat{n})$	Q1(\hat{n})	Med(\hat{n})	Q3(\hat{n})	SD(\hat{n})	$P(\hat{n} = \infty)$	PWR
0	200	640.09	307.00	490.00	780.00	511.39	0.04	74.39
0.25	200	180.54	140.00	154.50	180.00	82.45	0.00	90.20
0.5	200	142.23	131.00	141.00	151.00	18.21	0.00	99.80
1	200	142.73	132.00	141.00	152.00	18.32	0.00	100.00
0	400	678.77	410.00	571.00	801.75	408.25	0.01	82.68
0.25	400	160.35	125.00	141.00	179.25	51.96	0.00	88.20
0.5	400	116.67	109.00	116.00	123.00	11.38	0.00	100.00
1	400	116.39	109.00	115.50	123.00	11.11	0.00	100.00

Table 3:

Empirical concentration (OPT) using the projection-based sample size procedure at a nominal level of 90 using a pilot study of size $n_0 = 200$ and $n_0 = 400$. We test whether the true value of the estimated regime is within ϵ of the true value of the true optimal regime. The remaining columns give the mean, median, quartiles, and standard deviation of the estimated sample sizes across the 500 simulation repetitions.

ϵ	n_0	$E(\hat{n})$	$Q1(\hat{n})$	$Med(\hat{n})$	$Q3(\hat{n})$	$SD(\hat{n})$	OPT
0.30	200	1844.93	1561.50	1796.00	2060.00	365.42	100.00
0.50	200	754.27	663.50	748.00	836.00	128.73	100.00
0.70	200	431.14	387.00	426.00	473.00	65.13	100.00
0.30	400	1319.70	1191.50	1294.50	1451.75	193.43	100.00
0.50	400	543.24	500.00	526.50	586.00	69.92	100.00
0.70	400	320.63	293.25	319.00	345.75	37.88	100.00

Table 4:

Parameter estimates and 95% confidence intervals for the first- and second-stage blip model from the Kaiser Permanente Washington pilot data

Covariate	Estimate	Confidence Interval
Second Stage Blip Model		
A_2	0.06	(-2.27, 2.39)
$A_2 \times \text{Gender}$	4.86	(2.99, 6.73)
$A_2 \times \text{Age}$	-0.04	(-0.10, 0.02)
$A_2 \times \text{Baseline PHQ}$	-0.32	(-0.55, -0.09)
$A_2 \times \text{Anxiety}$	5.84	(3.93, 7.75)
$A_2 \times A_1$	5.10	(3.11, 7.09)
First Stage Blip Model		
A_1	2.20	(-15.53, 11.32)
$A_1 \times \text{Gender}$	0.73	(-3.53, 7.83)
$A_1 \times \text{Age}$	-0.07	(-0.12, 2.31)
$A_1 \times \text{Baseline PHQ}$	0.12	(-0.39, 2.53)
$A_1 \times \text{Anxiety}$	2.19	(-3.80, 7.67)