To cite this article:
Jeremy Yang, Dean Eckles, Paramveer Dhillon, Sinan Aral (2023) Targeting for Long-Term Outcomes. Management Science

Published online in Articles in Advance 03 Aug 2023

. https://doi.org/10.1287/mnsc.2023.4881 https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–15
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Targeting for Long-Term Outcomes

Jeremy Yang,[a,*] Dean Eckles,[b,*] Paramveer Dhillon,[c] Sinan Aral[b]

[a] Harvard Business School, Boston, Massachusetts 02163; [b] Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [c] University of Michigan, Ann Arbor, Michigan 48109

*Corresponding authors

Contact: jeryang@hbs.edu, https://orcid.org/0000-0001-8639-5493 (JY); eckles@mit.edu, https://orcid.org/0000-0001-8439-442X (DE); dhillonp@umich.edu, https://orcid.org/0000-0002-0994-9488 (PD); sinan@mit.edu, https://orcid.org/0000-0002-2762-058X (SA)

Received: October 7, 2020
Revised: February 12, 2022
Accepted: February 27, 2022
Published Online in Articles in Advance: August 3, 2023

https://doi.org/10.1287/mnsc.2023.4881

Abstract. Decision makers often want to target interventions so as to maximize an outcome that is observed only in the long term. This typically requires delaying decisions until the outcome is observed or relying on simple short-term proxies for the long-term outcome. Here, we build on the statistical surrogacy and policy learning literatures to impute the missing long-term outcomes and then approximate the optimal targeting policy on the imputed outcomes via a doubly robust approach. We first show that conditions for the validity of average treatment effect estimation with imputed outcomes are also sufficient for valid policy evaluation and optimization; furthermore, these conditions can be somewhat relaxed for policy optimization. We apply our approach in two large-scale proactive churn management experiments at *The Boston Globe* by targeting optimal discounts to its digital subscribers with the aim of maximizing long-term revenue. Using the first experiment, we evaluate this approach empirically by comparing the policy learned using imputed outcomes with a policy learned on the ground-truth, long-term outcomes. The performance of these two policies is statistically indistinguishable, and we rule out large losses from relying on surrogates. Our approach also outperforms a policy learned on short-term proxies for the long-term outcome. In a second field experiment, we implement the optimal targeting policy with additional randomized exploration, which allows us to update the optimal policy for future subscribers. Over three years, our approach had a net-positive revenue impact in the range of $4–$5 million compared with the status quo.

History: Accepted by Eric Anderson, marketing.
Funding: This work was supported by Boston Globe Media.
Supplemental Material: The online appendix and data are available at https://doi.org/10.1287/mnsc.2023.4881.

Keywords: long-term effect • statistical surrogate • policy learning • targeting • proactive churn management

1. Introduction

Advertising revenues have been stagnating for newspapers in recent years.[1] As a consequence, newspapers are looking for ways to strengthen their subscription-based business model. Take *The New York Times* as an example: in 2019, its total subscription revenue was twice its total advertising revenue (Online Figures A.1 and A.2). The CEO recently said, " … we still regard advertising as an important revenue stream, but we believe that our focus on establishing close and enduring relationships with paying, deeply engaged subscribers, and the long-range revenues which flow from those relationships, is the best way of building a successful and sustainable news business."[2] Hence, to succeed in a subscription-based business model, news publishers must retain their existing subscribers and maximize their long-term values. A common approach to achieving this goal is to target existing subscribers with marketing interventions, such as price discounts or other personalized offers.

We use news publishers as a motivating example, and it matches our empirical application. But how to optimize long-term customer outcomes by targeting interventions is a problem faced by most firms. Even more generally, decision makers in education, government, and medicine typically care about intervening for long-term outcomes such as employment, income, and survival.

"Long-term" and "short-term" outcomes are fruitfully understood as defined relative to the targeting cycle. For example, if a firm runs a campaign every year, then all outcomes that are observed within a year, such as the one-year revenue, might be considered short term because these outcomes are observed before the firm takes action (decides whom to target with what) in the next campaign. Hence, future policies can be optimized on these observed outcomes. In contrast, long-term outcomes materialize over time horizons longer than the window of opportunity for action, for example, three- or five-year revenue, rendering the firm incapable of optimizing its next campaign based on them. So a natural question arises: how can firms learn and implement an optimal targeting policy when the primary outcome of interest is long-term?

A straightforward solution to this problem is to wait until the long-term outcome materializes and choose a policy based on the realized long-term outcome. But this implies that the firm cannot learn anything in the meantime and, therefore, is unable to implement updated targeting policies until years later. Another solution is to find a short-term proxy (e.g., short-term revenue) for the long-term outcome and optimize for it instead. However, this could be problematic as the proxy and the long-term outcome might not be well-aligned. Hence, a policy that performs well on the proxy might not perform well in the long run.

In this paper, we propose to use surrogates (Prentice 1989, VanderWeele 2013) to impute the missing long-term outcomes and use the imputed long-term outcomes to optimize a targeting policy. We estimate the missing long-term outcome as the expectation of the long-term outcome conditional on surrogates of that outcome in a historical data set in which the long-term outcome is observed. Surrogate index estimators combine multiple surrogates by estimating the conditional expectation of the long-term outcome given the surrogates and using this to impute long-term outcomes (Xu and Zeger 2001, Athey et al. 2019). Once we have the imputed long-term outcomes, we optimize the targeting policy efficiently by using a doubly robust (DR) approach (Dudík et al. 2014, Athey and Wager 2021, Zhou et al. 2023) on the imputed long-term outcomes. We prove that this surrogate index–based approach recovers the optimal policy learned on true long-term outcomes under certain assumptions. We implement the optimal policy via bootstrapped Thompson sampling (Eckles and Kaptein 2014, Osband et al. 2016) to maintain exploration so we can update and reoptimize the policy for future subscribers to allow for potential nonstationarity.

We evaluate the efficacy of our approach empirically by running two large-scale field experiments that target discounts to the digital subscribers of *The Boston Globe*, a regional leader in news media. Boston Globe Media, which operates *The Boston Globe* newspaper and associated websites, is facing a similar problem to many other publishers. Our goal is to learn an optimal targeting policy that treats some subscribers with certain discounts to maximize their retention and long-term revenue. Here, a policy is a mapping from subscriber characteristics to offering a specific discount (or no discount or a distribution over discounts when the policy is stochastic). In this subscriber retention context, this is also known as proactive churn management.[3] To construct the surrogate index, we use the observed revenue and content consumption in the six months after treatment as our surrogates. We compare how well the policies learned using the surrogate index perform against policies optimized directly on short-term proxies (a benchmark) or realized long-term outcomes (the ground truth). We also consider alternative selections of surrogates for the construction of the surrogate index—perhaps most importantly whether we can use less than six months of revenue and consumption data. We estimate that this approach increases the firm's total projected digital subscription revenue by $4–$5 million over a three-year period relative to the status quo in the two experiments.

The rest of the paper is organized as follows. In Section 2, we review related work. The empirical context is described in Section 3. We introduce our method in Section 4: we first explain the imputation of the long-term outcome using the surrogate index and prove sufficient conditions for it to be valid for policy evaluation and optimization, and then, we describe the policy learning framework and how it is implemented. Experimental results and empirical validation of our approach are reported in Section 5. We conclude in Section 6.

2. Related Work

Our paper builds on a large body of literature in biostatistics and medicine on surrogate outcomes (i.e., endpoints, biomarkers); see, for example, Joffe and Greene (2009) and Weir and Walley (2006) for reviews. In clinical trials, the goal is often to study the efficacy of an intervention on outcomes such as the long-term health or survival rate of patients. However, the primary outcome of interest might be very rare, only observed after years of delay, or have high variance compared with the treatment effects (e.g., a 5- or 10-year survival rate). It is common to use the effect of an intervention on surrogate outcomes as a proxy for its effect on long-term outcomes. In a seminal paper, Prentice (1989) argues that, to be a valid surrogate, treatment and outcome have to be independent conditional on the surrogate. One intuitive way for this condition to be satisfied is if the surrogate fully mediates the treatment effect. In practice, it is hard to find a single variable that plausibly satisfies the condition (Freedman et al. 1992), but Xu and Zeger (2001) show that combining multiple surrogates to predict the outcome can be preferable to using a single surrogate because the treatment effect may operate through multiple pathways, and even when there is a single pathway, using multiple surrogates can reduce measurement error. This idea is further developed in a recent paper in econometrics (Athey et al. 2019), in which the combination is referred to as a surrogate index. This literature focuses on using surrogates to identify treatment effects on long-term outcomes, and in this paper, we extend this to policy optimization.

Another popular approach to modeling long-term outcomes is to posit a particular parametric generative model for the long-term outcomes. In the context of marketing, this is typically a model of customer lifetime value (CLV). CLV models are widely used in marketing for customer segmentation and targeting; see, for example, Gupta et al. (2006), Fader et al. (2014), Fader and Hardie (2015), and Ascarza et al. (2017) for surveys. CLV is defined as the sum of discounted future revenues or profits from a customer. To calculate CLV, we typically need to posit a parametric, for example, survival function and extrapolate the survival or retention probability into the future. A recent example in the context of churn management is Godinho de Matos et al. (2018), in which a parametric survival function is used. One advantage of this approach is that we can apply it even when the long-term outcomes are never observed because the prediction is based on functional form assumptions—unlike the surrogate index approach, which needs access to long-term outcomes in a historical data set; on the other hand, standard parametric CLV approaches may suffer from model misspecification. Also, the primary goal of CLV models is typically to predict outcomes, whereas the surrogate index approach focuses on learning treatment effects or optimizing policies: imputing outcomes is just a means to an end. More importantly, outcomes imputed via a surrogate index have provable properties regarding treatment effect estimation (Athey et al. 2019) or policy learning as developed here. Furthermore, building a CLV model may require substantial work to formalize business logic in anything but the simplest subscription businesses. A synthesis of these approaches is also possible in that a CLV prediction, if already available, can also be used as one of the surrogates in the construction of a surrogate index.

This paper is also related to the literature on targeting policy evaluation and optimization, which has recently further developed within marketing research. Hitsch and Misra (2018) propose an estimation method for conditional average treatment effects (CATEs) based on k-nearest neighbors (kNN) and use it for policy optimization. Simester et al. (2019) show that we can compare targeting policies more efficiently if we only compare the outcome of units on which the policies prescribe different actions. Simester et al. (2020) document nonstationarity, such as covariate and concept shifts between two experiments, and evaluate how robust different machine learning models used to optimize policies are to these changes in the environment. Yoganarasimhan et al. (2023) use different machine learning models to estimate CATEs and evaluate how targeting policies constructed using these models perform against each other. In another recent work, Lemmens and Gupta (2020) examine using a CLV model combined with field experimentation to optimize targeting in the policy learning framework.

Our work complements this literature by developing an approach that is novel in a few ways. First, we focus directly on targeting for long-term outcomes; outcomes used in these other works are short-term (in the sense that they are observable when we optimize and implement the policy) or extrapolation is done using a parametric CLV model.[4] Second, we systematically add randomized exploration around the learned policy, which allows us to evaluate and update the policy for future units in case the environment changes. Hitsch and Misra (2018) and Yoganarasimhan et al. (2023) study the problem in a static setting. Simester et al. (2019) do look at changes in the environment, but they focus on evaluating the robustness of different machine learning models. Third, we use a DR approach (Dudík et al. 2014) for both policy evaluation and learning, in contrast to Hitsch and Misra (2018) and Yoganarasimhan et al. (2023), who used an inverse probability weighting (IPW) estimator for policy evaluation. Lemmens and Gupta (2020) introduce a specialized incremental profit–based loss function that performs well in their empirical evaluation but lacks the asymptotic efficiency results available for doubly robust policy learning; it is also unclear how to combine this with known probabilities of treatment (i.e., design-based propensity scores) that arise in sophisticated experiments. In particular, even when probabilities of treatment are known exactly (as in our setting), DR estimators have advantages in statistical efficiency compared with IPW estimators (Athey and Wager 2021, Zhou et al. 2023).

Substantively, our study adds to the literature on subscriber management and proactive churn management in particular. Earlier work focuses on developing better prediction algorithms to more accurately identify potential churners; Neslin et al. (2006) provides a detailed comparison of different churn prediction models. Recently, the literature has examined causal effects of targeting interventions on churn using field experiments. For example, Ascarza (2018) and Lemmens and Gupta (2020) note that firms should not target customers based on their outcome level (churn risk) but should target based on treatment effects. Ascarza et al. (2016) show evidence from a field experiment with a telecommunication company that proactive churn interventions can backfire and increase the churn rate in practice. They argue that this is because proactive intervention lowers customers' inertia to switch plans and increases the salience of past usage patterns among potential churners. Our paper contributes to this literature by proposing an experimental framework that can be applied to directly optimize targeting policies for long-term customer retention and revenues.

3. Empirical Context

Founded in 1872, *The Boston Globe* is the oldest and largest daily newspaper in the greater Boston area. It has won a total of 27 Pulitzer Prizes and is widely regarded as one of the most prestigious papers in the United States. We ran two targeting experiments on digital only[5] subscribers of *The Boston Globe* in two experiments. Whereas we return to the details of our experiments and analyses in Section 5, we introduce the empirical context here so as to help fix ideas as we describe the methods.

Our analysis is of a random sample of about 45,000 digital subscribers in the first experiment and 95,000 in the second. For each subscriber, we observed the short-term outcome (e.g., monthly churn and revenue) and three sets of features: demographics (e.g., zip code), account activities (e.g., billing address change, credit card expiration date, complaints), and content consumption (e.g., when and what articles they read). There was only one intervention in the first experiment, which lowered the price for treated subscribers from \$6.93 per week to \$4.99 per week for eight weeks. An email (Online Figure B.1a) was sent to all treated subscribers in August 2018 telling them that a discount had been automatically applied to their accounts. We implemented six interventions in the second experiment: a thank you email, a \$20 gift card, a discount to \$5.99 for eight weeks, a discount to \$5.99 for four weeks, a discount to \$4.99 for eight weeks (the same as the intervention in the first experiment), and a discount to \$3.99 for eight weeks. A similar email (Online Figure B.1b) was sent to all treated subscribers in July 2019 with the corresponding message, and a treated subscriber had to click on a button at the bottom of the email to redeem the benefit. There was no overlap of treated subscribers between the two experiments.

4. Methods

In our application, the primary outcome of interest is long-term subscriber retention or revenue,[6] but we do not observe these outcomes in the short-term, that is, after the intervention in the first experiment and before we implemented the learned policy for the second experiment of customers. Hence, we use a surrogate index to address this problem.

Our framework has two components: first, we fit a model for long-term outcomes and use the resulting surrogate index to impute long-term outcomes; second, we learn an optimal policy using the imputed long-term outcomes. In Section 4.1, we explain the imputation and prove sufficient conditions for it to be valid for policy evaluation and optimization. In Section 4.2, we describe the policy evaluation and optimization framework and how it is implemented.

We first introduce the notation that we use throughout the section: let $\pi \in \Pi$ be a targeting policy that maps from the space of unit characteristics $\mathbb{X}$ to a space of distributions (simplexes) over a set of discrete actions $\mathbb{A}$; we index actions by $\{0, 1, 2, .., K-1\}$, where 0 is control and others are different interventions. When the policy is nondeterministic, it defines a nondegenerate probability distribution over possible actions conditional on covariates $\pi(a|\boldsymbol{x}) := \mathbb{P}(A = a|\boldsymbol{X} = \boldsymbol{x}),\ \forall a \in \mathbb{A}, \boldsymbol{x} \in \mathbb{X}$. When it is deterministic, it maps to a fixed action with probability one. Depending on the action chosen, we observe the corresponding potential outcome, that is, $Y_i = Y_i(A_i)$. These potential outcomes may be correlated with unit characteristics $\boldsymbol{X_i}$.

The goal is to learn a policy that maximizes some average outcome Y (if the goal is to minimize some average outcome Y, we can add a negative sign and turn it into a maximization problem):

Definition 1 (A Policy and Its Value).

$$\pi : \mathbb{X} \to \Delta(\mathbb{A}), \tag{1}$$

$$V(\pi) := \mathbb{E}[Y_i(A_i)]. \tag{2}$$

Definition 2 (Optimal Policy).

$$\pi^* := \underset{\pi \in \Pi}{\text{argmax}}\ V(\pi). \tag{3}$$

4.1. Imputing a Long-Term Outcome with a Surrogate Index

We use intermediate outcomes that are observed over the short-term period following the intervention as surrogates. Intuitively, the idea is to select surrogates that capture some of the ways that the actions affect the long-term outcome; in our application, these are subscribers' content consumption and short-term revenue. These surrogate variables are then combined with the long-term outcomes in the historical data set to impute missing long-term outcomes for units in the experiment.

Assume we have two data sets: one from the experiment labeled E and one based on historical (observational) data labeled H. We observe draws of the tuple ($\boldsymbol{X}$, A, $\boldsymbol{S}$) in the experiment, where $\boldsymbol{X} \in \mathbb{X}$ represents units' baseline characteristics, $A \in \mathbb{A}$ is the action (i.e., treatment, intervention), and $\boldsymbol{S} \in \mathbb{S}$ is the potentially vector-valued set of intermediate outcomes or surrogates. Note that the long-term outcome Y is unobserved in the experiment. In the historical data set, we observe draws of the tuple ($\boldsymbol{X}$, $\boldsymbol{S}$, Y); note that there is no known, randomized intervention in this data set (i.e., it is observational), but the long-term outcome Y is observed. We can define a surrogate index $\tilde{Y}$ for the long-term outcome Y as the expectation of the long-term outcome conditional on unit covariates and surrogates in the historical data set H:[7]

Definition 3 (Surrogate Index).

$$\tilde{Y}_i := \mathbb{E}_H[Y_i|\boldsymbol{S_i}, \boldsymbol{X_i}]. \tag{4}$$

Under Assumptions 1–3 below, a central result in Athey et al. (2019) is that the average treatment effect (ATE) on $\tilde{Y}$ recovers the ATE on long-term outcome Y. That is, by constructing the surrogate index, we can identify and feasibly estimate the ATE on some long-term outcomes without having to wait until they are observed.

Assumption 1 (Regular Treatment Assignment Mechanism: Ignorability and Positivity). *The treatment assignment is conditionally independent of potential long-term outcomes (ignorability), and all units have positive probability of being assigned to each action (positivity) in the experimental data set:*

$$A_i \perp\!\!\!\perp (Y_i(a), \boldsymbol{S_i}(a)) | \boldsymbol{X_i} \;\; \forall a \in \mathbb{A}, i \in E, \tag{5}$$

$$0 < \pi(a|\boldsymbol{x}) < 1 \;\; \forall a \in \mathbb{A}, x \in \mathbb{X}. \tag{6}$$

Assumption 1 is satisfied when we have indeed conducted a randomized experiment even if the probability of assignment to actions is conditional on observed covariates, as in our application.

Assumption 2 (Surrogacy). *The treatment assignment is independent of long-term outcomes conditional on the surrogates in the experimental data set:*

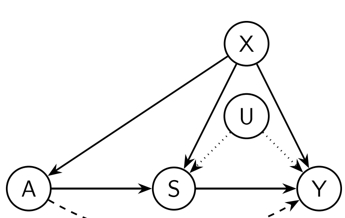
$$A_i \perp\!\!\!\perp Y_i | \boldsymbol{S_i}, \boldsymbol{X_i}, i \in E. \tag{7}$$

Whereas there can be other ways to satisfy this assumption, surrogacy is perhaps most intuitively implied by a generative model in which the set of surrogates fully mediate the causal effects from treatment to the long-term outcome (cf. Lauritzen 2004) as depicted in Figure 1 if the A to Y edge is absent. In our empirical context, it means the effects of price discounts on long-term retention and revenue should occur via some intermediate outcomes we observe, for example, content consumption and short-term revenue. Whereas it may have some testable implications, Assumption 2 is not directly testable.[8] Surrogacy is more plausible if we have a rich set of surrogates; perhaps this is more widely available given the increasing digitization of, for example, commerce and media consumption (as in our application).

Assumption 3 (Comparability). *The distribution of the long-term outcome conditional on the covariates and surrogates is the same across the experimental and historical data sets.*

$$Y_i | \boldsymbol{S_i}, \boldsymbol{X_i}, i \in E \sim Y_i | \boldsymbol{S_i}, \boldsymbol{X_i}, i \in H. \tag{8}$$

Figure 1. Directed Acyclic Graph Representing Causal Relationships Relevant to Satisfying the Assumptions



Notes. A is the treatment, which is randomized (possibly conditional on $\boldsymbol{X}$); S are the surrogates; Y is the long-term outcome, $\boldsymbol{U}$ and $\boldsymbol{X}$ are unobserved and observed covariates, respectively. This graph satisfies the ignorability component of Assumption 1. One way to satisfy Assumption 2 is the absence of causal pathways from A to Y that do not go through S, that is, the dashed edge is absent. One threat to the validity of Assumption 3 is if an unobserved time-varying variable $\boldsymbol{U}$ causes $\boldsymbol{S}$ and/or Y (dotted edges), so the observable relationship between Y and $\boldsymbol{S}$ is changing over time because of $\boldsymbol{U}$.

In our case, this assumption implies that the distribution of long-term retention and revenue (conditional on content consumption and short-term retention and revenue) should be the same between the experimental and historical data sets. Note that, under the comparability assumption, we have

$$\tilde{Y}_i = \mathbb{E}_H[Y_i | \boldsymbol{S_i}, \boldsymbol{X_i}] = \mathbb{E}_E[Y_i | \boldsymbol{S_i}, \boldsymbol{X_i}]. \tag{9}$$

In other words, the conditional expectation of Y_i in the experimental data set is equal to the conditional expectation in the historical data set, which is a quantity we can compute because in the historical data set Y_i is observed. This assumption fails if the distribution of long-term outcome conditional on covariates and surrogates are changing between the experimental and historical data sets. For instance, if the intervention itself modifies the relationship between long-term outcome and surrogates, the two distributions are different. In our empirical setting, it may be that, in the absence of an intervention, only very dedicated (i.e., high retention rate) subscribers read some categories of content; however, some actions might induce other, less dedicated subscribers to read that content. For this reason, having similar (even unobserved) interventions in the historical data could strengthen our confidence in this assumption. More extreme violations of this assumption can occur when measurement of a surrogate is changing (e.g., what counts as reading an article has a different definition in historical data). Note that, whereas not put in potential outcomes notion here or in Athey et al. (2019), one way for comparability to be satisfied involves observational causal inference about the effects of $\boldsymbol{S}$ on Y using the historical data to succeed; thus, we expect that, as in observational causal inference, this is a very strong assumption that is often not exactly true. This motivates our consideration of weaker assumptions and the use of empirical evaluation in our application.

Given these assumptions, we prove that the surrogate index is valid for policy evaluation and optimization. Policy evaluation is the estimation of $V(\pi)$ for a given policy π. Policy optimization is finding a π^* that maximizes $V(\pi)$. See Section 4.2 for more details about doing so in finite samples; here, we simply consider the optimal policy defined on the population. We show that the value of a policy with respect to surrogate index is identical to its value on the long-term outcome; this, in turn, implies that the optimal policy with respect to the surrogate index coincides with that optimal policy with respect to long-term outcomes. We state the main results here, and the proofs are in Online Appendix C. Let $\tilde{V}(\pi)$ denote the value of π with respect to $\tilde{Y}$ rather than Y.

Proposition 1. *Under Assumptions 1–3, policy evaluation conducted on a surrogate index identifies the true policy value defined on long-term outcomes:*

$$\tilde{V}(\pi) = V(\pi) \;\; \forall \pi \in \Pi. \tag{10}$$

Then, because the function being maximized is identical at all points, it is also identical at its maximum.

Proposition 2. *Under Assumptions 1–3, policy optimization conducted on a surrogate index recovers the true optimal policy.*

$$\underset{\pi\in\Pi}{\operatorname{argmax}}\; \tilde{V}(\pi) = \underset{\pi\in\Pi}{\operatorname{argmax}}\; V(\pi). \tag{11}$$

Propositions 1 and 2 could justify the approach developed here and employed in our empirical application. However, somewhat weaker assumptions than those that have been used for results for estimation of the ATE or CATEs are in fact sufficient for Proposition 2.

Define real and surrogate index–imputed CATEs, $\tau_{aa'}(\boldsymbol{x}) = \mathbb{E}_E[Y(a) - Y(a')|\boldsymbol{X} = \boldsymbol{x}]$ and $\tilde{\tau}_{aa'}(\boldsymbol{x}) = \mathbb{E}_E[\tilde{Y}(a) - \tilde{Y}(a')|\boldsymbol{X} = \boldsymbol{x}]$. When, for example, Assumption 2 is violated (perhaps the set of surrogates does not fully mediate the treatment effect on long-term outcomes), the CATE estimated using the surrogate index can be biased (even with infinite data). That is, $\tau_{aa'}(\boldsymbol{x}) \neq \tilde{\tau}_{aa'}(\boldsymbol{x})$ for some $\boldsymbol{x} \in \mathbb{X}$. Here, our aim is not estimating CATEs, but simply optimizing the policy. Bias in CATEs (i.e., nonzero $\tau_{aa'}(\boldsymbol{x}) - \tilde{\tau}_{aa'}(\boldsymbol{x})$) does not result in a loss in the value of the optimized policy unless the bias changes the sign of that CATE.[9]

Thus, we can introduce a somewhat weaker assumption, replacing Assumptions 2 and 3, that is sufficient for policy optimization. The intuition that sign preservation is sufficient is that, for policy optimization purposes, we only care about identifying which is the best action for each unit, not how much better it is (i.e., we just need to correctly order the actions with respect to treatment effects; the magnitude of differences between actions do not matter).

Assumption 4 (Sign Preservation). *The sign of conditional average treatment effects is the same for the surrogate index and the long-term outcome:*

$$sign(\tilde{\tau}_{aa'}(\boldsymbol{x})) = sign(\tau_{aa'}(\boldsymbol{x})) \;\; \forall a, a' \in \mathbb{A}, \boldsymbol{x} \in \mathbb{X}. \tag{12}$$

This is an assumption directly on CATEs and so is not as readily interpretable with respect to the data-generating process. Nonetheless, we can reason about how this assumption may be more plausible in some settings than others. For example, in cases with a binary treatment, if we hypothesize that a treatment "works" (i.e., has a large positive effect) on some groups but not others and this treatment has some small cost (which is incorporated into the definition of Y), then the distribution of CATEs may be bimodal with no density near zero. This could contrast with other cases in which theory might lead us to expect highly heterogeneous benefits and costs of the treatment (both incorporated into the definition of Y). For example, in our empirical application, for subscribers whose behavior is unaffected by a discount, this reduces long-term revenue to varying degrees depending on how long they are retained; similarly, for those affected, this may affect long-run revenue in complex, heterogeneous ways. This highlights the value of empirical validation of surrogate index–based policy optimization in our setting (Section 5.3). Even in the favorable case in which the distribution of CATEs is bimodal with no density near zero, analysis with an impoverished set of covariates may result in loss. Say these available covariates are less informative about treatment effects; then, the distribution of CATEs might have substantial density near zero, raising the concern that any bias in CATE estimation may translate to selecting a suboptimal policy when using a surrogate index.

One can analytically characterize the loss in policy optimization, much as Athey et al. (2019) develop bounds on the bias for the ATE. Here, we state this result with details in Online Appendix C.

Proposition 3. *There is a loss in the value of the optimal policy only when the optimal action estimated on a surrogate index is different than the true optimal action. The total loss, or regret, is*

$$\int_X \tau_{a^*\tilde{a}^*}(\boldsymbol{X}) \cdot \mathbb{1}_{\{a^*(\boldsymbol{X}) \neq \tilde{a}^*(\boldsymbol{X})\}}\, dF(\boldsymbol{X}), \tag{13}$$

$$a^*(\boldsymbol{X}) := \{a \in \mathbb{A} | \tau_{aa'}(\boldsymbol{X}) > 0 \;\; \forall a' \in \mathbb{A}\}, \tag{14}$$

$$\tilde{a}^*(\boldsymbol{X}) := \{a \in \mathbb{A} | \tilde{\tau}_{aa'}(\boldsymbol{X}) > 0 \;\; \forall a' \in \mathbb{A}\}. \tag{15}$$

In summary, assumptions introduced in the surrogacy literature can be used to justify policy evaluation and optimization with a surrogate index. Furthermore, it is possible to relax these assumptions for policy optimization precisely because the optimal policy is only sensitive to the sign of treatment effects.

4.2. Evaluating, Learning, and Implementing Targeting Policies

We describe the off-policy evaluation and learning framework using the imputed long-term outcome $\tilde{Y}$ obtained via the procedure in Section 4.1.[10] Under assumptions articulated in the previous section, this can identify the same optimal policy as using the true long-term outcome Y. We use ~ on variables or functions with parameters constructed with $\tilde{Y}$. We describe each term generically and also make some connections to the quantities in our experiments. Readers familiar with counterfactual policy evaluation and learning may choose to skip to Section 5 in which we discuss the experiments and results.

4.2.1. Off-Policy Evaluation. In off-policy evaluation, we use data collected under the design (or behavior) policy[11] π_D to estimate the value of a counterfactual policy π_P. One popular choice of estimator is based on IPW. The Hájek estimator, a normalized version of the Horvitz–Thompson estimator (Horvitz and Thompson 1952), is typically used to implement IPW (Särndal et al. 2003, section 7.3). The Hájek estimate of the average outcome under an arbitrary targeting policy π_P using data collected under a design or behavior policy π_D is

$$\hat{\hat{V}}_{\text{IPW}}(\pi_P) = \left(\sum_i \frac{\pi_P(A_i|\boldsymbol{X_i})}{\pi_D(A_i|\boldsymbol{X_i})}\right)^{-1} \cdot \sum_i \frac{\pi_P(A_i|\boldsymbol{X_i})}{\pi_D(A_i|\boldsymbol{X_i})} \cdot \tilde{Y}_i, \tag{16}$$

where $\tilde{Y}_i$ is the imputed outcome (e.g., predicted three-year revenue), $A_i \in \{0, 1, 2, \ldots, K-1\}$ is the action (e.g., discount) received by unit i assigned by the design policy π_D, and π_P is the probability of assigning unit i to a given condition under the counterfactual policy that we want to evaluate.[12] We use $A_i = 0$ to denote the control and $A_i = 1$ to denote the treatment when actions are binary.[13] The first term in Equation (16) is simply a normalization term; the ratio between π_P and π_D is also known as the importance weight. As specified by Assumption 1, we need π_D to be strictly positive for all unit–action pairs. Note that we do not require the policy being evaluated π_P to have this property; it can be a deterministic policy. The Horvitz–Thompson estimator is unbiased but typically has higher variance. The Hájek estimator is biased in finite samples but consistent, and it typically has lower variance; it is, therefore, more widely used in practice.[14] The main advantage of IPW is that it is fully nonparametric when the propensity scores are known, and it does not require us to specify a model for the outcome process.

However, the IPW estimator has two main limitations. First, the Hájek estimator can still suffer from high variance. Second, when evaluating a deterministic policy π_P, it only uses observations for which the actions prescribed by the target policy π_P and design policy π_D agree (when they don't agree, $\pi_P(A_i|X_i)$ is always zero). This reduces the effective sample size, especially when π_P and π_D are very different.[15] Following Robins et al. (1994), one way to improve upon IPW is by augmenting it with an outcome model μ to use all observations and further stabilize the estimator. This is known as the augmented IPW or DR estimator (Dudík et al. 2014). Under the DR approach, the value of a policy π_P can be estimated as

$$\hat{\hat{V}}_{\text{DR}}(\pi_P) = \frac{1}{n}\sum_i \left(\hat{\hat{\mu}}(\boldsymbol{X_i}, \pi_P) + \frac{\pi_P(A_i|\boldsymbol{X_i})}{\pi_D(A_i|\boldsymbol{X_i})} \cdot (\tilde{Y}_i - \hat{\hat{\mu}}(\boldsymbol{X_i}, A_i))\right), \tag{17}$$

where

$$\hat{\hat{\mu}}(\boldsymbol{X_i}, \pi_P) = \sum_{a \in \mathbb{A}} \pi_P(a|\boldsymbol{X_i}) \cdot \hat{\hat{\mu}}(\boldsymbol{X_i}, a). \tag{18}$$

The first term in Equation (17), $\hat{\hat{\mu}}(\boldsymbol{X_i}, \pi_P)$, is an outcome model that estimates the expectation of the imputed outcome for a random covariates profile X_i and distribution of actions given by a policy π using data from the experiment. In the most common case of evaluating a deterministic policy, $\hat{\hat{\mu}}(\boldsymbol{X_i}, \pi_P)$ is just $\hat{\hat{\mu}}(\boldsymbol{X_i}, a)$ for the action to which π_P assigns units with covariate profile $\boldsymbol{X_i}$. For example, in our empirical application, it corresponds to the estimated three-year revenue for a subscriber profile $\boldsymbol{X_i}$ under a particular discount a. Note that this outcome model $\hat{\hat{\mu}}$ is different from the one for $\tilde{Y}$ in Equation (9); there, the outcome is estimated as a function of surrogates and covariates using the historical data, whereas $\hat{\hat{\mu}}$ estimates outcome as a function of actions and covariates using the experimental data. The second term is the importance weight multiplied by the prediction error; it corrects the first term toward the direction of the long-term outcome by an amount that is proportional to the prediction error. For a deterministic target policy π_P, it does so whenever the actions prescribed by π_D and π_P agree. Note that the high variance of IPW estimators is from the importance weights (dividing by a small probability when π_D is very unbalanced), and this term vanishes if the prediction error is small. Both IPW and DR estimators are consistent, but DR estimation can achieve semiparametric efficiency (see, e.g., Robins et al. 1994, Hahn 1998, Farrell 2015) and typically has lower variance than IPW estimation. We use the DR estimator for policy evaluation.

4.2.2. Off-Policy Optimization. As shown in the previous section, policy optimization builds on CATE estimation. We focus on using doubly robust estimation.[16] We can first construct a doubly robust score for each unit–action pair (which also has the interpretation of an estimate of an individual potential outcome) (Robins et al. 1994, Dudík et al. 2014, Athey and Wager 2021, Chernozhukov et al. 2022, Zhou et al. 2023):

$$\hat{\hat{\gamma}}_a(\boldsymbol{X_i}) = \hat{\hat{\mu}}(\boldsymbol{X_i}, a) + \frac{\tilde{Y}_i - \hat{\hat{\mu}}(\boldsymbol{X_i}, a)}{\pi_D(a|\boldsymbol{X_i})} \cdot 1_{\{A_i = a\}}. \tag{19}$$

These doubly robust scores are equal to the prediction of an outcome model $\hat{\hat{\mu}}(\boldsymbol{X_i}, a)$ plus a correction term based on IPW; the correction is applied if and only if the action being evaluated is the same as the action taken. This is intuitive because the correction term depends on $\tilde{Y}_i$, which is the outcome under a realized action A_i; it is informative only when the action being evaluated is the same as a; otherwise, the term drops out, and the doubly robust scores reduce to the outcome model. The CATE, relative to the control, given a covariate profile x, can then be estimated as

$$\hat{\hat{\tau}}_{a0}(x) = \frac{1}{|\{i : \boldsymbol{X}_i = x\}|} \sum_{i:X_i=x} \left(\hat{\hat{\gamma}}_a(\boldsymbol{X}_i) - \hat{\hat{\gamma}}_0(\boldsymbol{X}_i)\right). \tag{20}$$

We can use these doubly robust scores for policy optimization (Murphy et al. 2001, Dudík et al. 2014) by solving a cost-sensitive classification problem.[17] That is, the estimated optimal policy is

$$\hat{\hat{\pi}}^* = \underset{\pi\in\Pi}{\operatorname{argmax}} \frac{1}{n}\sum_i \left(\hat{\hat{\gamma}}_1(\boldsymbol{X}_i) - \hat{\hat{\gamma}}_0(\boldsymbol{X}_i)\right) \cdot (2\pi(\boldsymbol{X}_i) - 1), \tag{21}$$

or in a multiaction case

$$\hat{\hat{\pi}}^* = \underset{\pi\in\Pi}{\operatorname{argmax}} \frac{1}{n}\sum_i < \hat{\hat{\boldsymbol{\gamma}}}(\boldsymbol{X}_i), \pi(\boldsymbol{X}_i) >, \tag{22}$$

where $\hat{\hat{\boldsymbol{\gamma}}}(\boldsymbol{X}_i) = (\hat{\hat{\boldsymbol{\gamma}}}_0(\boldsymbol{X}_i), \hat{\hat{\boldsymbol{\gamma}}}_1(\boldsymbol{X}_i), \dots, \hat{\hat{\boldsymbol{\gamma}}}_k(\boldsymbol{X}_i))$ is a vector of doubly robust scores based on Equation (19) and $\pi(\boldsymbol{X}_i)$ is a vector of probabilities with which the policy assigns a unit to each action. $< \cdot >$ is the dot product between vector-valued $\hat{\hat{\boldsymbol{\gamma}}}(\boldsymbol{X}_i)$ and $\pi(\boldsymbol{X}_i)$.

In the cost-sensitive classification problem, for each unit, the correct label is the action that corresponds to the highest doubly robust score, and the loss for classifying a unit to action a, when the correct label is a^*, is $\hat{\hat{\gamma}}_{a^*}(\boldsymbol{X}_i) - \hat{\hat{\gamma}}_a(\boldsymbol{X}_i)$, which is the loss of the imputed outcome (e.g., predicted three-year revenue) when a unit is assigned to a suboptimal action. In multiaction cases, a cost-sensitive binary classification is done on every pair of actions, and the final action is chosen by a majority vote. In practice, the policy class Π is often restricted by the choice of a specific type of classifier (e.g., logistic regression or decision trees for interpretation or transparency reasons) or by using only a subset of covariates in the classifier (still using all information to construct the doubly robust scores). A practical advantage of this approach is that, once the doubly robust scores or labels are constructed, we can plug them into off-the-shelf classifiers to optimize the policy.

4.2.3. Policy Implementation and Exploration. Whereas we have estimated the optimal policy, it is typically desirable to account for remaining statistical uncertainty and continue randomized exploration, which can be particularly important if there is nonstationarity, that is, changes in the environment that make a policy that is optimal today no longer optimal in the future. Whereas other approaches can be suitable, we find particularly suitable a variant of Thompson sampling, bootstrap Thompson sampling (BTS) (Eckles and Kaptein 2014, Osband et al. 2016, Lu and Van Roy 2017), that is readily implemented with models for which Thompson sampling might be cumbersome to implement; see Eckles and Kaptein (2019) and Osband et al. (2019) for reviews. We use BTS as a heuristic approach to adding randomized uncertainty-based exploration to the estimated optimal targeting policy in which a unit i is assigned to action a with probability proportional to the fraction of times an action is estimated to be optimal across all bootstrap replicates of the data. That is,

$$\hat{\hat{\pi}}_{\mathrm{BTS}}(a|X_i) = \frac{1}{R}\sum_{r=1}^{R} 1_{\{\hat{\hat{\pi}}_r^*(X_i)=a\}}, \tag{23}$$

where $\hat{\hat{\pi}}_r^*$ is the policy estimated according to Equation (21) or (22) on the rth bootstrap replicate.[18]

4.3. Summary of the Methods

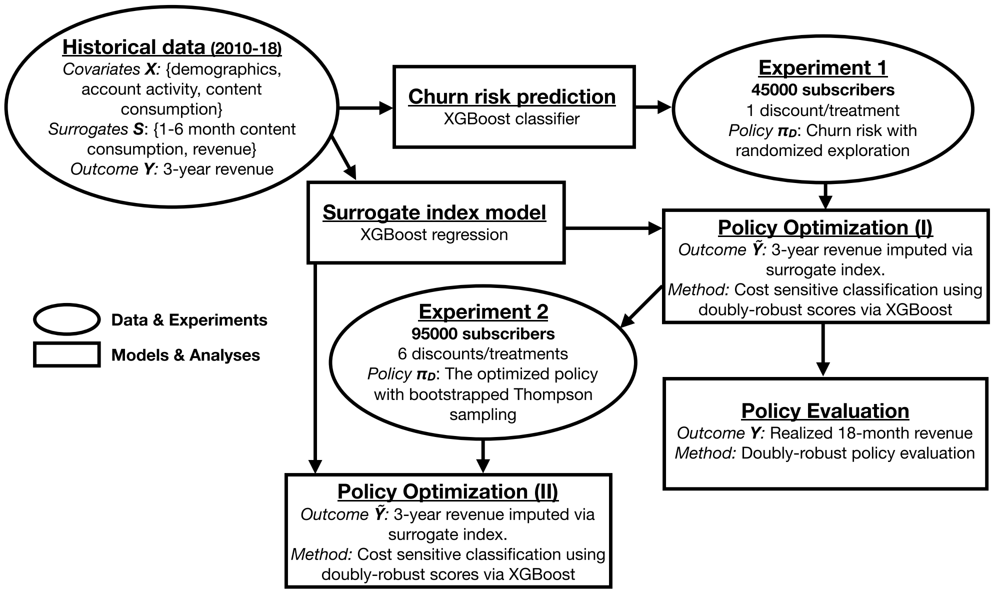
We summarize the key steps in combining these methods as follows:

0. Identify the long-term outcome of interest (Y), intervention (A), covariates ($\boldsymbol{X}$), and surrogates ($\boldsymbol{S}$).
1. Run a randomized experiment through a design policy π_D to generate experimental data ($\boldsymbol{X}, A, \boldsymbol{S}$). Gather historical data ($\boldsymbol{X}, \boldsymbol{S}, Y$).
2. Impute the missing long-term outcomes in the experiment using the surrogate index $\tilde{Y}$ through Equation (9).
3. Do policy optimization using imputed long-term outcomes $\tilde{Y}$ to get an estimated optimal policy $\hat{\hat{\pi}}^*$ through Equations (19) and (21) or (22).
4. Implement the estimated optimal policy $\hat{\hat{\pi}}^*$, potentially with added randomization as in $\hat{\hat{\pi}}_{\mathrm{BTS}}$ through Equation (23).
5. Consider step 4 as running a new randomized experiment with $\hat{\hat{\pi}}_{\mathrm{BTS}}$ being the new π_D, and repeat steps 1–4 as desired.

5. Experiments and Results

We now turn to applying and evaluating this approach in the context of reducing churn at *The Boston Globe*, at which we offer discounts to existing subscribers. Figure 2 gives an overview of how the historical observational data and two field experiments relate to each other and the main analyses. Experiment 1 randomized subscribers to receive a discount or not. We then optimized the targeting policy using results from Experiment 1 and a surrogate index constructed from historical data; the surrogates we use are content consumption (number of articles read in each of the 20 most visited sections[19] on *The Boston Globe* website) and revenue over the first six months. We selected these surrogates based on the following reasoning. First, revenue captures whether a subscriber has already churned as well as whether the subscriber has perhaps received other discounts (e.g., via reactive churn management). Second, subscribers get value from their subscription primarily by consuming articles and other content on *The Boston Globe* website. We expected that some of this content is more differentiated from that otherwise available (e.g., local sports coverage).

Figure 2. Summary of Observational and Experimental Data and Analyses



Notes. Historical observational data are used to train a churn prediction model and a model for long-term outcomes (producing a surrogate index). Experiment 1 uses the churn predictions in randomly assigning subscribers to treatments. Using the data from Experiment 1, we learn a policy using the surrogate index, which is then used in (a) the design of Experiment 2 and (b) evaluations compared with actual 18-month revenue. Similarly, we learn a policy from the Experiment 2 data and the surrogate index.

These surrogates could be measured over shorter or longer periods. Intuitively, the longer we wait, the better we can estimate the long-term revenue, but firms also want to learn the optimal policy quickly so we can implement it. In particular, we should expect that it will be important to observe revenue and consumption for some time after the discounts expire. Given these considerations, we used surrogates computed over six months of data.

We implemented the policy with additional randomized exploration in Experiment 2. Once 18 months had passed since the start of Experiment 1, we were able to compare the performance of the policy we learned using the surrogate index to the policy we would have learned using the longer-term, 18-month outcomes.[20] All treatment effects are from intent-to-treat analyses that do not condition on potentially endogenous posttreatment behaviors, such as opening the email or redeeming the benefit. We report the survival curves and treatment effects estimated from the resulting data in both experiments in Online Appendix D.3. In this section, we focus on the experiment design, policy learning, and surrogate index validation results.

5.1. Experiment 1

As is typical of a new effort in proactive churn management, we lacked prior experimental data in which subscribers were assigned to discounts. However, we anticipated that the discount treatment would not have substantial beneficial effects on subscribers with a low probability of churning. Thus, we assigned subscribers to treatment using a design policy π_D in the first experiment that balances exploration and exploitation; we do so by assigning subscribers with higher predicted churn probability into treatment with higher probability, ensuring that all subscribers $0 < \pi_D(X_i) < 1$, thus satisfying Assumption 1; see Online Appendices D.1 and D.2 for a more detailed discussion. This assigned 806 subscribers to receive a discounted subscription rate ($4.99 per week) for eight weeks.

We estimate the optimal policy via the binary cost-sensitive classification (Equation (21)) on imputed long-term revenue, defined as either 18-month or 3-year revenue. In this section, we focus on the policy using imputed 3-year revenue; we return to the policy using imputed 18-month revenue in our validation in Section 5.3. We first construct doubly robust scores for each subscriber using Equation (19), in which $\hat{\tilde{\mu}}$ is estimated using XGBoost via cross-fitting.[21] We then split the data into training (80%) and test sets (20%) and use XGBoost as the classifier with hyperparameters tuned via cross-validation. The policy learned using the surrogate index, $\hat{\tilde{\pi}}$, would treat 21% of subscribers in the experimental data. We evaluate policy performance on the test data using the doubly robust estimator as in Equation (17). According to the surrogate index, it would generate a $40 revenue increase per subscriber (95% confidence interval [$10, $75]) over three years compared with the current policy that treats no one, which is $1.7 million in total for subscribers in the first experiment.

We use tools in interpretable machine learning to look at what variables are most important in determining the optimal policy and how the optimal policy depends on these variables (see Online Appendix D.4). The top three variables are risk score (predicted risk of churn), tenure, and number of sports articles read in the last six months. The optimized policy treats subscribers with shorter tenure (more recently registered subscribers) at a higher rate. The relationship between number of sports articles read and treatment is not monotone: the fraction treated is low for very inactive and very active subscribers and higher for subscribers in between. The relationship with risk score is interestingly also not monotone; for subscribers with the highest risk scores, the treatment fractions are higher, and this is consistent with our prior. But, for some subscribers with very low risk scores, the treatment probabilities are even higher. This also highlights potential blind spots of targeting solely based on risk scores.

5.2. Experiment 2

Having learned a policy using the first experiment, we turned to exploiting this knowledge and further learning through experimentation in a second experiment. Furthermore, the success of the first experiment prompted creating and trying a larger set of six treatments: a thank you email, a $20 gift card, a discount to $5.99 for eight weeks, a discount to $5.99 for four weeks, a discount to $4.99 for eight weeks (the same as the intervention in the first experiment), and a discount to $3.99 for eight weeks.

We use the learned policy based on imputed three-year revenue—with two modifications—to allocate subscribers to treatments. First, as discussed in Section 4.2.3, adding randomization to an estimated optimal policy is a desirable practice especially in a potentially nonstationary environment. We added randomization to the optimized policy through bootstrap Thompson sampling as in Equation (23). This assigned 5,688 subscribers to treatments. Second, because all but one of the treatments were new, the learned policy was not directly informative about which noncontrol actions to take; therefore, conditional on a subscriber being assigned to treatment, we assigned subscribers to the six noncontrol conditions uniformly at random. For future subscribers, we can learn and implement an optimal policy over all interventions based on the results from Experiment 2.

We optimize the policy via multiclass cost-sensitive classification (Equation (22)) using data from Experiment 2 following a similar procedure as in Experiment 1. The optimized policy using the surrogate index, $\hat{\hat{\pi}}$, allocates around a quarter of subscribers each to control, the thank you email, and the two smallest discounts; a few subscribers are allocated to other actions (Table 1). This optimized policy improves three-year revenue by $30 per subscriber (95% confidence interval [$12, $50]) relative to the status quo that treats no one such that it would have generated $2.8 million for subscribers in Experiment 2.

We further compare the two experiments to see whether there are significant changes in the environment in terms of covariate and concept shift (Online Appendix D.5). When the environment is stationary, it is more efficient to pool data from the two experiments together to estimate the optimal policy for future subscribers, and when the environment is substantially changing, it is better to downweight observations from the first experiment using a time-decaying case weight (e.g., Russac et al. 2019). We only use data from the second experiment to estimate the optimal policy because there is some evidence for concept shift, and there is only one common treatment condition between the two experiments.

Table 1. Distribution of Optimal Actions Estimated from Experiment 2

Action	Percentage
Control	23
Thank you email only	25
Gift card	<1
$5.99/8 weeks	25
$5.99/4 weeks	27
$4.99/8 weeks	<1
$3.99/8 weeks	<1

Note. Percentage is the percentage of subscribers in Experiment 2 that are assigned to this action according to the policy optimized using the surrogate index, $\hat{\hat{\pi}}$.

5.3. Surrogate Index Validation and Comparison

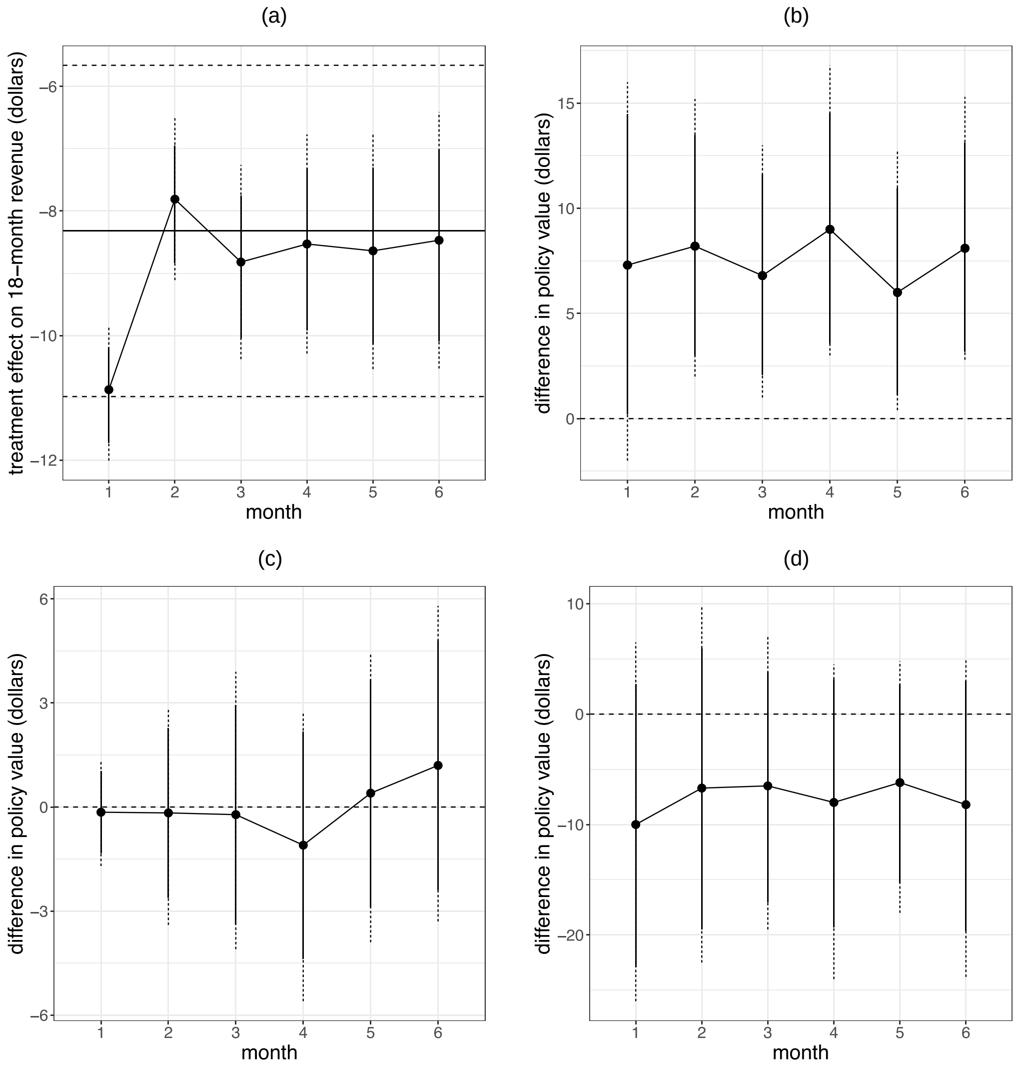
The assumptions underlying surrogate index–based policy learning are strong, and it is often implausible that they are strictly true; this is similar to, for example, doubts about conditional ignorability in observational causal inference or the exclusion restriction in instrumental variables analyses. Thus, as in those settings (e.g., Dehejia and Wahba 2002, Gordon et al. 2019, Eckles and Bakshy 2021), it is often valuable to empirically evaluate the results of our approach when that is possible (i.e., when we do observe long-term outcomes). Researchers can wait until the true long-term outcomes are observed and then compare the effect estimates and policies based on the surrogate index with those based on the true long-term outcomes. Here, it takes three years to observe the long-term outcome for which *The Boston Globe* is targeting; instead, we use 18-month revenue (from August 2018 to February 2020), which is already realized at the time this is written, as the long-term outcome and repeat the analysis. Policy values are estimated using the doubly robust approach as in Equation (17) except that the outcomes we use are observed Y_i, not imputed $\tilde{Y}_i$.

We first look at how well the surrogate index recovers the treatment effect estimated on the true long-term outcome. We then evaluate it by looking at how it performs against a benchmark policy that is learned on some short-term proxies of the long-term outcomes (e.g., one- to six-month revenue) and a policy learned on the true long-term outcome (e.g., realized 18-month revenue). We also look at how the performance changes if we chose a different subset of surrogates. All policy values here are defined relative to the status quo of treating no one. We report confidence intervals from 1,000 bootstrap draws on the test data.

First, we look at how the average treatment effect on the treated (ATT) calculated using the surrogate index compares with ATT calculated using the true outcome (Figure 3(a)). After the first month, the surrogate index–based ATT estimates are indistinguishable from the estimates using realized 18-month revenue. That the one-month surrogate index–based ATT is distinguishable from those using surrogates computed on longer periods may indicate that one month is too short a period; this is intuitively consistent as the treatment is an eight-week discount, so no reaction to the subsequent price increase is yet observed. Note that the confidence intervals of ATT estimated on true outcomes are wider than the ones estimated on the surrogate index. When the surrogacy assumption holds, it is more efficient to estimate the treatment effect on the surrogate index because it discards irrelevant variation in the long-term outcome.

Next, we look at the value of the surrogate index–based policy (Figure 3(b)). All results are significantly better than the status quo except when we only use

Figure 3. Empirical Validation Using Experiment 1 of Using the Surrogate Index for Treatment Effect Estimation and Policy Learning



Notes. (a) ATT on revenue using surrogate indices estimated with data from the first one to six months. The horizontal lines are the ATT estimated with true 18-month revenues and its 95% confidence interval. The solid and dashed vertical lines are 75% and 95% confidence intervals, respectively. (b) The value difference between policies optimized on surrogate indices constructed with surrogates from the first one to six months and the current policy. Except for a single month, they outperform the status quo. The solid and dashed vertical lines are 75% and 95% confidence intervals, respectively. (c) The value difference between policies optimized with a single short-term proxy (revenues from the first one to six months) and the current policy. The value is indistinguishable from the status quo. The solid and dashed vertical lines are 75% and 95% confidence intervals, respectively. (d) The value difference between policies optimized on surrogate indices constructed with surrogates from the first one to six months and true outcomes. They are statistically indistinguishable. The solid and dashed vertical lines are 75% and 95% confidence intervals, respectively.

information from the first month; recall that the discount ends after eight weeks. By contrast, optimizing the policy directly on short-term proxies (one- to six-month revenue) does not detectably outperform the status quo (Figure 3(c)). We also compare the surrogate index–based policy with a policy learned on the true long-term outcome (Figure 3(d)). Although all the point estimates of the value difference are negative, none of them is distinguishable from zero; the difference between the value of policy learned on surrogate indices using the first six-month and true outcomes is −\$8 per subscriber (95% confidence interval [−\$24, \$5]). This comparison does not take into account the gain in time and opportunity cost by implementing an optimized policy at 6 versus 18 months. These two policies also agree on 72% of subscribers; that is, they assign them to the same treatment condition. This is encouraging, but it also contributes to imprecision in estimating differences between them as the estimates are determined by the long-term revenue of a smaller number of subscribers.

Finally, we compare the performance of policies learned on surrogate indices constructed using only content consumption information, only short-term revenue, and both; the three approaches are not detectably different though there is substantial uncertainty, so this does not rule out relevant differences (Online Appendix D.6).[22]

6. Conclusion

Many applied problems, ranging from the subscriber management problem studied here to others in business, medicine, public policy, and social sciences in which there is a need to personalize interventions to optimize some long-term outcomes, can be fruitfully characterized as learning a targeting policy for some long-term outcomes. Here, we advance the practice of policy learning by incorporating the use of a learned surrogate index to impute the long-term outcomes. We first show analytically when a surrogate index is valid for policy evaluation and optimization in place of true unobserved long-term outcomes. Then, to validate our approach empirically, we run two large-scale experiments that prescribe who should be targeted with what incentives in order to maximize long-term subscription revenue for *The Boston Globe*. Combining data from the first experiment and the passage of time, we show that the policy optimized on long-term outcomes imputed by a surrogate index outperforms a policy optimized on a short-term proxy of the long-term outcomes and that it performs similarly to the policy optimized on true long-term outcomes. We then implement the optimized policy with additional randomized exploration so that we can respond to potential nonstationarity and update the optimized policy after each experiment. The total three-year revenue impact of implementing policies optimized using the surrogate index relative to the status quo in the two experiments sums to \$4–\$5 million. Our paper adds to and complements a recent and growing literature in marketing on policy evaluation and learning (e.g., Hitsch and Misra 2018; Simester et al. 2019, 2020; Yoganarasimhan et al. 2023) and empirical work in proactive churn management (e.g., Ascarza 2018) by focusing on optimizing targeting policies for long-term retention and revenue.

A natural question to ask is how to choose surrogates when imputing long-term outcomes. If we have the generative model in Figure 1 in mind, we want to choose variables that lie on the causal path from treatment to long-term outcomes as suggested by domain knowledge or theory. We also want to choose surrogates that are observable shortly after the intervention so that the policy can be learned quickly. These two considerations may be in tension. If relevant experiments have been conducted in the past, then the quality of surrogates can be evaluated on the realized long-term outcomes as we have done here. Surrogates that are highly predictive of the outcome are potential candidates, but there is no guarantee that they will produce high policy values as predicting the outcome level is a different task than predicting the treatment effect or learning the policy. Future research may further examine selection of potential surrogates. In practice, we may only have noisy measurements of such surrogates; thus, a fruitful direction for future work may be incorporating recent developments from mediation analysis with multiple noisy measurements (Ghassami et al. 2021). Finally, because surrogacy is fundamentally a question about the underlying causal mechanism, once some surrogates are shown to be valid for a given problem, they may be likely to remain valid for similar problems in the future. For example, we show that short-term revenues and content consumption are suitable surrogates for the effect of price discounts on long-term retention and subscription revenues, so the firm can tentatively rely on this assumption as they continue to iterate on targeting policies. We can imagine building such a knowledge base for different sets of problems and long-term outcomes as more empirical researchers work in this general framework.

The present work is not without important limitations. Some of these are limitations of the approach as developed here. For example, it is directly applicable when there is essentially no constraint on how many units can be treated as in our case. When there is a budget constraint and heterogeneous treatment costs, a policy can be optimized based on the ratio between individual-level treatment effect and the cost of treatment as in Sun et al. (2021). There are also important limitations to the strength of the conclusions from our empirical application. For example, though we were not able to detect differences in performance between the surrogate index–based policy and one based on true long-term outcomes, this may reflect remaining statistical uncertainty in estimating this contrast; similar considerations apply to other comparisons, such as between the value of policies using different sets of surrogates. More generally, the quite promising results observed here may not be indicative of what practitioners can expect in other, even somewhat similar subscriber management settings, perhaps especially if a very different variety of actions are used. Thus, we hope that subsequent work offers both further methodological development and empirical validation.

Acknowledgments

The authors thank Boston Globe Media and particularly Jessica Bielkiewicz, Thomas Brown, Ryan McVeigh, and Shannon Rose for their partnership in conducting the field experiments. This work benefited from comments by Susan Athey, John Hauser, Günter Hitsch, and Duncan Simester as well as participants at Yale University Junior Quantitative Marketing Conference, Summer Institute in Competitive Strategy, Marketplace Innovations Workshop, Management, Analytics, and Data Conference, Harvard University Marketing Research Camp, Theory + Practice in Marketing Conference, Meta Core Data Science, Big Data in Mobile Analytics Conference, Quantitative Marketing and Economics Conference, RAND Corporation Causal Inference Symposium, American Economic Association Annual Meeting, Hebrew University of Jerusalem School of Business Seminar, Workshop on Information Systems and Economics, Lyft Inference and Statistics Reading Group, Institute of Operations Research and the Management Sciences Annual Meeting (2020/11), Conference on Computational Social Science, Massachusetts Institute of Technology Initiative on the Digital Economy Annual Conference, Massachusetts Institute of Technology Marketing Seminar, Conference on Neural Information Processing Systems CausalML Workshop, Harvard Business School Digital Doctoral Workshop, Conference on Digital Experimentation, Advances with Field Experiments Conference, Marketing Science Conference.

Endnotes

[1] The print advertising revenue is declining with a compound annual growth rate (CAGR) of −12.6% from 2016 to 2021; whereas digital ad revenue is still growing at a CAGR of 2.2%, it's not enough to compensate for the loss in print. (Source: U.S. Online and Traditional Media Advertising Outlook).

[2] See https://www.nytimes.com/2018/02/08/business/new-york-times-company-earnings.html.

[3] Here, proactive simply means that the intervention (discount) happens before a churn intention is observed; by contrast, reactive churn management means that the company first waits for customers to request to cancel their subscription and then offers some discount or other benefits in reaction to this in the hope of retaining them. One analogy is that the proactive approach is similar to diagnosing and preventing illness before a patient shows clear symptoms, and the reactive approach is similar to treating patients who are already ill.

[4] Yoganarasimhan et al. (2023) show that, in their particular case, the policy learned on short-term outcomes also does well on long-term outcomes, but the policy is not directly optimized on long-term outcomes.

[5] The *Globe* also has a combined print and digital subscription. All subscribers are paying customers.

[6] Being a digital service, marginal costs are negligible compared with subscription revenue.

[7] One advantage of this approach is that the estimation of the conditional expectation can be treated as a supervised learning problem and can be performed using flexible nonparametric machine learning methods such as XGBoost (Chen et al. 2015, Chen and Guestrin 2016).

[8] This can also be described as an exclusion restriction as in instrumental variables. As in that case, this assumption has both testable and untestable implications. It might be tempting to regress the outcome on surrogate and treatment and test if the coefficient of treatment is zero. This naive test is not valid when there are unobserved confounders for the surrogate and outcome: conditioning on the surrogate or a "collider" in such a case generates spurious correlation between treatment and confounder and, hence, between treatment and outcome. See Joffe and Greene (2009) for a more detailed discussion.

[9] Concern with getting the sign of the treatment effect correct using surrogates features prominently in the literature on the "surrogate paradox" in which various surrogacy definitions are satisfied by the effect on the surrogate and outcome have opposite signs; see, for example, Chen et al. (2007), VanderWeele (2013), and Jiang et al. (2016).

[10] In an abuse of notation, we now use $\tilde{Y}$ (rather than, e.g., $\hat{\tilde{Y}}$) to denote the actually imputed long-term outcome, which is estimated, whereas in Definition 3, it denotes the true conditional expectation as otherwise this makes some further expressions cumbersome.

[11] In the reinforcement learning literature (e.g., Sutton and Barto 2018, section 5.5, p. 103), the policy used to collect training data is called a behavior policy. We call it a design policy in our experimental setting.

[12] The corresponding unnormalized Horvitz–Thompson estimator is $\frac{1}{n}\sum_i \frac{\pi_P(A_i|X_i)}{\pi_D(A_i|X_i)} \cdot \tilde{Y}_i$.

[13] For example, when $A_i = 1$, it means unit i was in treatment and was assigned to treatment with probability $\pi_D(1|X_i)$, and $\pi_P(1|X_i)$ is the probability that i receives treatment under counterfactual policy π_P. Similarly, when $A_i = 0$, it means unit i was in the control and was assigned to control with probability $\pi_D(0|X_i)$, and $\pi_P(0|X_i)$ is the probability that i is in control (or not be treated) under counterfactual policy π_P.

[14] For more discussion about normalization in IPW estimation, see Owen (2019, chapter 9) and Khan and Ugander (2023).

[15] Two policies are similar if they tend to prescribe the same action for a given unit profile; the more often they prescribe different actions for a given unit, the more different they are.

[16] Estimation of CATE can also be implemented in different ways. Hitsch and Misra (2018) distinguish between what they label "indirect" approaches (which first estimate the outcome model as a function of covariates and actions and then take the difference between actions as treatment effects) and "direct" methods that estimate the CATE directly without first estimating an outcome function (e.g., causal trees, Athey and Imbens 2016; causal forest Wager and Athey 2018; and causal kNN, Hitsch and Misra 2018). This typology may be confusing to readers familiar with contextual bandit and policy learning literatures in which, at least since Dudík et al. (2014), "direct methods" are those using outcome regressions without IPW (i.e., what Hitsch and Misra 2018 label "indirect").

[17] When $\pi_D(x)$ must be estimated, this approach comes with guarantees on asymptotic regret compared with the true optimal policy (Athey and Wager 2021, Zhou et al. 2023).

[18] In cases in which a unit is always or never assigned to some conditions, we may want to impose a probability floor and ceiling to ensure that all units have positive probability being assigned to all conditions, thereby satisfying the assumption.

[19] The sections are metro, sports, news, lifestyle, business, opinion, arts, Sunday magazine, ideas, search, member center, south, spotlight, page not found, nation, north, magazine, circulars, and politics.

[20] We use the most recent historical data to do the imputation; that is, for Experiment 1, run in 2018, we used the observed revenue data from 2015–2018 to estimate the three-year revenue for subscribers in the experiment.

[21] Cross-fitting means that $\hat{\hat{\mu}}$ for individual i is estimated without using i's own data in the training process. We can split data randomly into n folds, and then $\hat{\mu}$ for individuals in a given fold is trained only using data from the other $n-1$ folds; it reduces overfitting and improves efficiency (Athey and Wager 2021, Zhou et al. 2023). We use $n=3$ in our estimation.

[22] Athey et al. (2019) suggest that, when the surrogacy condition holds, the smallest set of surrogates has the highest precision in estimating the treatment effect.

References

Ascarza E (2018) Retention futility: Targeting high-risk customers might be ineffective. *J. Marketing Res.* 55(1):80–98.

Ascarza E, Fader PS, Hardie BGS (2017) Marketing models for the customer-centric firm, chapter 10. Wierenga B, van Der Lans R, eds. *Handbook of Marketing Decision Models* (Springer, Berlin), 297–329.

Ascarza E, Iyengar R, Schleicher M (2016) The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *J. Marketing Res.* 53(1):46–60.

Athey S, Imbens G (2016) Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* 113(27):7353–7360.

Athey S, Wager S (2021) Policy learning with observational data. *Econometrica* 89(1):133–161.

Athey S, Chetty R, Imbens GW, Kang H (2019) The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Technical report, National Bureau of Economic Research, Cambridge, MA.

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. Krishnapuram B, Shah M, General Chairs. *Proc. 22nd ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM, New York), 785–794.

Chen H, Geng Z, Jia J (2007) Criteria for surrogate end points. *J. Roy. Statist. Soc. Ser. B Statist. Methodology* 69(5):919–932.

Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T (2015) XGBoost: Extreme gradient boosting. R package version 0.4-2. 1(4):1–4.

Chernozhukov V, Escanciano JC, Ichimura H, Newey WK, Robins JM (2022) Locally robust semiparametric estimation. *Econometrica* 90(4):1501–1535.

Dehejia RH, Wahba S (2002) Propensity score-matching methods for nonexperimental causal studies. *Rev. Econom. Statist.* 84(1): 151–161.

Dudík M, Erhan D, Langford J, Li L (2014) Doubly robust policy evaluation and optimization. *Statist. Sci.* 29(4):485–511.

Eckles D, Bakshy E (2021) Bias and high-dimensional adjustment in observational studies of peer effects. *J. Amer. Statist. Assoc.* 116(534):507–517.

Eckles D, Kaptein M (2014) Thompson sampling with the online bootstrap. Preprint, submitted October 15, https://arxiv.org/abs/1410.4009.

Eckles D, Kaptein M (2019) Bootstrap Thompson sampling and sequential decision problems in the behavioral sciences. *SAGE Open* 9(2).

Fader PS, Hardie BGS (2015) Simple probability models for computing CLV and CE. Kumar V, Shah D, eds. *Handbook of Research on Customer Equity in Marketing* (Edward Elgar Publishing, Cheltenham, UK), 77–100.

Fader PS, Hardie BGS, Sen S (2014) Stochastic models of buyer behavior. Winer RS, Neslin SA, eds. *The History of Marketing Science* (World Scientific Publishing Co. Pvt. ltd., Singapore), 165–205.

Farrell MH (2015) Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* 189(1):1–23.

Freedman LS, Graubard BI, Schatzkin A (1992) Statistical validation of intermediate endpoints for chronic diseases. *Statist. Medicine* 11(2):167–178.

Ghassami A, Shpitser I, Tchetgen E (2021) Proximal causal inference with hidden mediators: Front-door and related mediation problems. Preprint, submitted November 4, https://arxiv.org/abs/2111.02927.

Godinho de Matos M, Pedro F, Rodrigo B (2018) Target the ego or target the group: Evidence from a randomized experiment in proactive churn management. *Marketing Sci.* 37(5):793–811.

Gordon BR, Zettelmeyer F, Bhargava N, Chapsky D (2019) A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Marketing Sci.* 38(2):193–225.

Gupta S, Hanssens D, Hardie B, Kahn W, Kumar V, Lin N, Ravishanker N, Sriram S (2006) Modeling customer lifetime value. *J. Service Res.* 9(2):139–155.

Hahn J (1998) On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 66(2):315–331.

Hitsch GJ, Misra S (2018) Heterogeneous treatment effects and optimal targeting policy evaluation. Preprint, submitted February 6, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3111957.

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* 47(260):663–685.

Jiang Z, Ding P, Geng Z (2016) Principal causal effect identification and surrogate end point evaluation by multiple trials. *J. Roy. Statist. Soc. Ser. B Statist. Methodology* 78(4):829–848.

Joffe MM, Greene T (2009) Related causal frameworks for surrogate outcomes. *Biometrics* 65(2):530–538.

Khan S, Ugander J (2023) Adaptive normalization for IPW estimation. *J. Causal Inference* 11(1):20220019.

Lauritzen SL (2004) Discussion on causality. *Scandinavian J. Statist.* 31(2):189–201.

Lemmens A, Gupta S (2020) Managing churn to maximize profits. *Marketing Sci.* 39(5):956–973.

Lu X, Van Roy B (2017) Ensemble sampling. Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. *Adv. Neural Inform. Processing Systems* (Neural Information Processing Systems Foundation, Inc., San Diego), 3258–3266.

Murphy SA, van der Laan MJ, Robins JM, and Conduct Problems Prevention Research Group (2001) Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* 96(456):1410–1423.

Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH (2006) Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *J. Marketing Res.* 43(2):204–211.

Osband I, Blundell C, Pritzel A, Van Roy B (2016) Deep exploration via bootstrapped DQN. Lee DD, von Luxburg U, Garnett R, Sugiyama M, Guyon I, eds. *Adv. Neural Inform. Processing Systems* (Neural Information Processing Systems Foundation, Inc., San Diego), 4026–4034.

Osband I, Van Roy B, Russo D, Wen Z (2019) Deep exploration via randomized value functions. *J. Machine Learn. Res.* 20:1–62.

Owen AB (2019) Monte Carlo theory, methods and examples. Accessed October 2020, https://artowen.su.domains/mc/.

Prentice RL (2006) Surrogate endpoints in clinical trials: Definition and operational criteria. *Statist. Medicine* 8(4):431–440.

Robins JM, Rotnitzky A, Zhao LP (1994) Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* 89(427):846–866.

Russac Y, Vernade C, Cappé O (2019) Weighted linear bandits for non-stationary environments. Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, eds. *Adv. Neural Inform. Processing Systems* (Neural Information Processing Systems Foundation, Inc., San Diego), 12017–12026.

Särndal CE, Swensson B, Wretman JH (2003) *Model Assisted Survey Sampling* (Springer-Verlag, New York).

Simester D, Timoshenko A, Zoumpoulis SI (2019) Targeting prospective customers: Robustness of machine learning methods to typical data challenges. *Management Sci.* 66(6):2495–2522.

Simester D, Timoshenko A, Zoumpoulis SI (2020) Efficiently evaluating targeting policies: Improving on champion vs. challenger experiments. *Management Sci.* 66(8):3412–3424.

Sun H, Du S, Wager S (2021) Treatment allocation under uncertain costs. Preprint, submitted March 20, https://arxiv.org/abs/2103.11066.

Sutton RS, Barto AG (2018) *Reinforcement Learning: An Introduction*, 2nd ed. (MIT Press, Cambridge, MA).

VanderWeele TJ (2013) Surrogate measures and consistent surrogates. *Biometrics* 69(3):561–565.

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113(523):1228–1242.

Weir CJ, Walley RJ (2006) Statistical evaluation of biomarkers as surrogate endpoints: A literature review. *Statist. Medicine* 25(2): 183–203.

Xu J, Zeger SL (2001) The evaluation of multiple surrogate endpoints. *Biometrics* 57(1):81–87.

Yoganarasimhan H, Barzegary E, Pani A (2023) Design and evaluation of personalized free trials. *Management Sci.* 69(6):3220–3240.

Zhou Z, Athey S, Wager S (2023) Offline multi-action policy learning: Generalization and optimization. *Oper. Res.* 71(1):148–183.