A RELUCTANT ADDITIVE MODEL FRAMEWORK FOR INTERPRETABLE NONLINEAR INDIVIDUALIZED TREATMENT RULES

BY JACOB M. MARONGE[1,a], JARED D. HULING[2,b] AND GUANHUA CHEN[3,c]

[1]*Department of Biostatistics, University of Texas MD Anderson Cancer Center,* [a]*jmmaronge@gmail.com*

[2]*Division of Biostatistics, University of Minnesota,* [b]*huling@umn.edu*

[3]*Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison,* [c]*gchen25@wisc.edu*

Individualized treatment rules (ITRs) for treatment recommendation is an important topic for precision medicine as not all beneficial treatments work well for all individuals. Interpretability is a desirable property of ITRs, as it helps practitioners make sense of treatment decisions, yet there is a need for ITRs to be flexible to effectively model complex biomedical data for treatment decision making. Many ITR approaches either focus on linear ITRs, which may perform poorly when true optimal ITRs are nonlinear, or black-box nonlinear ITRs, which may be hard to interpret and can be overly complex. This dilemma indicates a tension between interpretability and accuracy of treatment decisions. Here we propose an additive model-based nonlinear ITR learning method that balances interpretability and flexibility of the ITR. Our approach aims to strike this balance by allowing both linear and nonlinear terms of the covariates in the final ITR. Our approach is parsimonious in that the nonlinear term is included in the final ITR only when it substantially improves the ITR performance. To prevent overfitting, we combine crossfitting and a specialized information criterion for model selection. Through extensive simulations we show that our methods are data-adaptive to the degree of nonlinearity and can favorably balance ITR interpretability and flexibility. We further demonstrate the robust performance of our methods with an application to a cancer drug sensitive study.

1. Introduction. Machine learning tools have been developed extensively for healthcare applications and have the opportunity to deeply impact human lives (Rudin (2019)). A major area where machine learning has shown promise is in the development of individualized treatment rules (ITRs), which are used to help personalize and tailor healthcare decisions based on the individual characteristics of patients. Many of these approaches are black boxes in that the model-fitting procedure does not yield results that are readily interpretable to subject-matter experts or patients. The lack of transparency and accountability of black-box models may have severe consequences, such as the misdiagnosis of diseases (Rudin (2019)) or incorrect interpretation of the modeled relationships between factors. This is particularly evident when identifying ITRs. As ITRs (and the methods used to identify them) are directly related to medical treatment decisions (Vayena, Blasimme and Cohen (2018)), there is a need for ITR learning methods that not only yield accurate recommendation but also achieve a level of interpretability. Moreover, intepretable ITRs could be more accessible for stakeholders, including physicians and patients, and easier to implement in medical practice.

The goal of this paper is to develop an approach for estimating ITRs that are both accurate and interpretable. There is often a trade-off between model performance (predictive accuracy) and model complexity (interpretability) (Hastie, Tibshirani and Friedman (2009), Murdoch et al. (2019)), and we aim to strike a balance along this trade-off. Simpler models tend to be easier to interpret but may fail to fully extract signals from rich, heterogeneous data, as

is common in studies of human health. On the other hand, complex black-box models may perform well but can be challenging to interpret and can have issues with overfitting. Most existing methods either result in linear ITRs (Chen et al. (2017), Qian and Murphy (2011), Wallace and Moodie (2015)), which may perform poorly when the underlying true ITR is nonlinear, or black-box ITRs (e.g., the ones based on Random Forests (Wager and Athey (2018), Zhu et al. (2017)), the ones based on deep neural networks (Liang, Ye and Fu (2018), Mi, Zou and Zhu (2019)), or the ones based on supporting vector machines with nonlinear kernel (Zhao et al. (2012), Zhou et al. (2017))), which may be difficult to interpret. In order to achieve a balance between these two problems, in a similar spirit to the reluctant additive model (RAM) of Tay and Tibshirani (2020), we propose a reluctant additive model for identifying ITRs. The key principle of the reluctant additive model is to allow for nonlinearity but to shrink preferentially toward a linear term unless warranted by the data. We refer to our approach as Reluctant Additive Individualized Treatment Rules (RAITR), as it adheres to the reluctant additive principle.

Classical additive models require the user to choose a priori which covariates require a nonlinear functional form and which should be linear, which is a limitation especially in high-dimensional scenarios. As a result, classical additive models can be overly complex. Several approaches in the literature have explicitly focused on handling high dimensionality for additive models by inducing sparsity with penalties, such as the approaches of Lin and Zhang (2006), Ravikumar et al. (2009), and Meier, van de Geer and Bühlmann (2009). However, these approaches either shrink the entire functional form of the effect of a covariate to zero or otherwise include a fully nonlinear term. The fused LASSO additive model approach of Petersen, Witten and Simon (2016) allows for shrinkage toward piecewise constant effects, enabling more parsimonious selected models than approaches that only focus on selection, which either include or exclude a nonlinear term in its entirety. The sparse partially linear additive model of Petersen and Witten (2019) is able to induce sparsity for GAMs and can shrink toward linear functions; however, we adapted and extend RAM for learning ITRs instead, due to its simple computational nature, which will be explained below.

In the context of ITRs, the work of Park et al. (2022) has extended the notion of sparse additive models to handle sparse additive ITRs, which inherits the limitations of the above sparse additive methods. In contrast, RAMs decompose the effect of a covariate on the outcome into a linear and nonlinear component and prioritize the linear terms in estimation, encouraging more aggressive shrinkage toward a parsimonious model while also allowing selection of individual variables into the model. This is achieved by fitting the linear terms first as well as penalizing the nonlinear terms based on their complexity and only allowing their inclusion when they substantially improve the model performance above and beyond a linear form. As ITR estimation involves treatment-covariate interactions, using a highly flexible model to estimate the interactions could lead to overfitting. As such, the ability of our framework to default to linear contributions to the ITR is critical and beneficial in practice. Moreover, the variable selection ability of our framework makes it more compelling for learning ITRs with high-dimensional covariates, compared to approaches that do not incorporate explicit variable selection such as causal (random) forests (Wager and Athey (2018)) or weighted support vector machines (e.g., with a Gaussian kernel) (Zhou et al. (2017)). We believe that our RAITR method has promise for effective use, in practice, as it offers the possibility for handling complex forms of heterogeneity of effects while encouraging parsimony and interpretable effects through shrinkage toward linear ITRs.

The reluctant additive principle, while advantageous in many ways, also creates unique challenges in the setting of ITR estimation. Since the reluctant additive principle is inherently multistaged, care is required to prevent overfitting. We combine crossfitting, to help prevent double-dipping outcome information, with use of the concordance information criteria (CIC;

Shi, Song and Lu (2021)) to perform the final ITR model selection. As we demonstrate empirically, implementing the CIC performs quite well at identifying complex treatment rules while simultaneously preventing overfitting.

This paper is structured as follows. In Section 2 we define notation and assumptions to estimate ITRs in the context of observational data as well as describe the model-fitting procedure for RAITR. In Section 3 we conduct a series of simulations to compare ITRs estimated from our approach to simple linear ITRs and complicated black-box nonlinear ITRs. In Section 4 we apply RAITR to learn ITRs for cancer treatment using the Genomics of Drug Sensitivity in Cancer (GDSC) study (Iorio et al. (2016)), where the data are high dimensional and of low to moderate signal. We demonstrate that for studies like GDSC it is useful for a method to be flexible enough to detect nonlinear ITRs but also pick simpler models when required. Finally, we conclude with a discussion in Section 5.

2. Methods.

2.1. *Notation and assumptions.* We refer to our treatment of interest as an exposure to emphasize that we are working in an observational context. Each independent observation is composed of a triplet $(Y, A, \mathbf{X})$, where Y is the observed response, A is a binary exposure taking values 1 or -1, and $\mathbf{X}$ is a p-dimensional vector of preexposure covariates of interest. For ease of notation, we make the assumption that a larger value of Y corresponds to a more desirable outcome. Adopting the potential outcomes framework of Rubin (2005), we denote $Y(a)$ as the potential outcome if the exposure were set to level a. We then define

$$(2.1) \qquad \Delta(\mathbf{x}) \equiv E[Y(1)|\mathbf{X} = \mathbf{x}] - E[Y(-1)|\mathbf{X} = \mathbf{x}]$$

as the conditional (causal) average treatment effect (CATE), given $\mathbf{x}$.

In many real-world scenarios, the CATE varies as a function of covariates, which is a form of a "heterogeneous treatment effect" (HTE). The CATE is closely related to individualized treatment rules (ITR) in the sense that

$$d^*(\mathbf{x}) = \text{sign}\{\Delta(\mathbf{x})\}$$

is an optimal ITR. As such, an optimal ITR can be estimated by modeling $\Delta(\cdot)$. Our modeling strategy focuses on estimating a flexible ITR through a flexible parameterization of the CATE, while maintaining the qualitative and visual interpretability of the resulting ITR.

To relate $\Delta(\mathbf{x})$ to observable quantities, we impose the following standard assumptions for causal identification:

(i) Consistency: $Y = \mathbb{1}(A = 1)Y(1) + \mathbb{1}(A = -1)Y(-1)$.
(ii) Strong ignorability: $A \perp\!\!\!\perp (Y(1), Y(-1))|\mathbf{X}$.
(iii) Positivity: $0 < \pi(\mathbf{x}, a) < 1$ for all $\mathbf{x} \in \mathcal{X}$ for $a \in \{-1, 1\}$, where $\pi(\mathbf{x}, a) = \Pr(A = a|\mathbf{X} = \mathbf{x})$.

With these assumptions (2.1) is equivalent to

$$(2.2) \qquad \Delta(\mathbf{x}) = E(Y|A = 1, \mathbf{X} = \mathbf{x}) - E(Y|A = -1, \mathbf{X} = \mathbf{x}).$$

We can highlight the role of the conditional treatment effect by reexpressing the conditional mean function as

$$(2.3) \qquad E(Y|A = a, \mathbf{X} = \mathbf{x}) = \mu(\mathbf{x}) + \frac{a}{2}\Delta(\mathbf{x}),$$

where $\mu(\mathbf{x}) = \frac{1}{2}\{E(Y|A = 1, \mathbf{X} = \mathbf{x}) + E(Y|A = -1, \mathbf{X} = \mathbf{x})\}$, which is a simple average of conditional outcomes under two treatments for a given set of covariate values. We refer to $\mu(\mathbf{X})$ as the main effect of $\mathbf{X}$ for convenience. Please note that this definition differs from the commonly used notion of "main effect" in the ITR literature, which typically refers to the conditional mean outcome under control ($E(Y|A = 0, \mathbf{X} = \mathbf{x})$).

2.2. *Reluctant additive individualized treatment rules.* In this section we describe how we model and estimate ITRs by minimizing a loss function motivated by (2.3). Though linear ITRs are more interpretable than black-box ITRs, they may be insufficient in accurately describing complex HTEs. We aim to strike a balance between interpretability and complexity by extending linear ITRs via sparse generalized additive models (GAMs, Hastie and Tibshirani (1990)), which allows each variable to impact the outcome in a nonlinear and additive fashion but omits interactions between components in **X**. We adopt the "reluctant GAM" idea (Tay and Tibshirani (2020)) of fitting sparse GAMs, wherein linear approximations of $\Delta(\mathbf{x})$ are favored over nonlinear ones unless otherwise warranted by the data. As a guiding principle, we prefer an ITR to only contain effects that are linear in the original set of variables: nonlinearities are only included thereafter if they significantly improve performance (Tay and Tibshirani (2020)). Since we focus on estimating treatment rules, we coin our approach as Reluctant Additive Individualized Treatment Rules (RAITR). An outline of the RAITR procedure is provided in Algorithm 1. Our approach differs from the approach of Tay and Tibshirani (2020) in several key ways, some of which are due to the particularities in estimating ITRs, and others are general modifications that we have found to improve performance and mitigate the effects of overfitting arising from the procedure's sequential nature.

We propose to approximate $\Delta(\mathbf{X})$ with an additive model, that is,

$$(2.4) \qquad \Delta(\mathbf{X}) \approx f(\mathbf{X}) \equiv \sum_{j=1}^{p} f_j(X_j),$$

where f_j is a function of only the jth entry, X_j, of **X**. In the following we will introduce a computationally efficient and effective procedure for estimating the additive components f_j that will comprise our estimated ITR.

Many frameworks have been proposed to model ITRs with the primary focus on linear ITRs, including outcome-weighted learning (Zhao et al. (2012)), Q-learning (Qian and Murphy (2011)), weighted learning (Chen et al. (2017)), and dynamic weighted ordinary least

---

Algorithm 1 Reluctant GAMs for Interpretable Nonlinear ITR

Input: $(\mathbf{X}_i, Y_i, A_i)$, $i = 1, \ldots, n$: covariate **X** (with dimension p), outcome Y, treatment A, and penalty terms $\lambda_1, \lambda_2 \geq 0$.

Process: Fit the weighted learning with least square loss and linear rule:

$$\hat{\boldsymbol{\beta}}_1 = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \hat{\pi}(\mathbf{X}_i, A_i)^{-1} \left\{ Y_i - \hat{m}(\mathbf{X}_i) - \frac{A_i}{2} \boldsymbol{\beta}^T \mathbf{X}_i \right\}^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j|,$$

and compute residuals $r_i = Y_i - \hat{m}(\mathbf{X}_i) - (A_i/2)\hat{\boldsymbol{\beta}}_1^T \mathbf{X}_i$, $i = 1, \ldots n$. For each covariate $j \in \{1, \ldots, p\}$, fit a weighted ($\hat{\pi}(\mathbf{X}_i, A_i)^{-1}$ as the weight) penalized smoothing spline (with data-dependent degrees of freedom) of r_i on $g_j(X_{ij})$. Fit penalized regression with LASSO penalty of A on **X** and **G**:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \hat{\pi}(\mathbf{X}_i, A_i)^{-1} \left\{ Y_i - \hat{m}(\mathbf{X}_i) - \frac{A_i}{2}(\boldsymbol{\beta}_{\text{lin}}^T \mathbf{X}_i + \boldsymbol{\beta}_{\text{non}}^T \mathbf{G}_i) \right\}^2$$

$$+ \lambda_2 \sum_{j=1}^{p} (|\boldsymbol{\beta}_{\text{lin}, j}| + \gamma_j |\boldsymbol{\beta}_{\text{non}, j}|),$$

with penalty $\gamma_j = \min(\sqrt{p}, 1 + s_j^{-1})$.

Output: $\hat{f}_{\lambda_2}(\mathbf{X}_i) = \widehat{\boldsymbol{\beta}}_{\text{lin}} \mathbf{X}_i + \widehat{\boldsymbol{\beta}}_{\text{non}} \mathbf{G}_i$ for $i = 1, \ldots, n$.

---

squares (Wallace and Moodie (2015)). We propose a general framework that can enhance any least square loss-based linear ITR learning method (Chen et al. (2017), Qian and Murphy (2011), Wallace and Moodie (2015)) to learn interpretable nonlinear ITRs. As an illustrative example, we focus on extending the weighted learning framework of Chen et al. (2017) to learn interpretable nonlinear ITRs. The loss function for learning optimal ITR ($f_{opt}(\mathbf{X})$) under weighted learning is defined as follows:

$$(2.5) \qquad f_{opt}(\mathbf{X}) = \underset{f}{\arg\min}\, E\left[\frac{\{Y - m(\mathbf{X}) - \frac{A}{2} f(\mathbf{X})\}^2}{\pi(\mathbf{X}, A)}\right],$$

where $m(\mathbf{X})$ is any function of $\mathbf{X}$ only and $\pi(\mathbf{X}, A)$ is the propensity score. The validity of (2.5) is not impacted by the choice of $m(\mathbf{X})$ when $\pi(\mathbf{X}, A)$ is correctly specified. However, using $m(\mathbf{X}) = \mu(\mathbf{X})$ can protect against the misspecification of $\pi(\mathbf{X}, A)$. For this reason we use this choice throughout our paper and generally advocate for use of flexible models to estimate this function. The inclusion of the inverse of propensity score is to eliminate measured confounding in observational studies. In particular, this weighting framework can be considered as the extension of inverse propensity score weighted regression for average treatment effect estimation to ITRs estimation.

Since we are not interested in the interpretation of the main effects and the propensity score, we can use flexible machine learning techniques to estimate them so that the model misspecification issues can be avoided. However, using flexible models for $m(\mathbf{X})$ and $\pi(\mathbf{X}, A)$ can also lead to overfitting issue and bias the estimation of ITR. To mitigate this issue, we use the crossfitting procedure (Chernozhukov et al. (2018)) together with flexible models to estimate these nuisance parameters. For the rest of the section, we assume that we have access to an estimate of $m(\mathbf{X})$ and $\pi(\mathbf{X}, A)$ that can be plugged into (2.5), and we defer from describing crossfitting version of our algorithm to Section 2.3.

Consider the case with n independent observations $(Y_i, A_i, \mathbf{X}_i)_{i=1}^n$. The first step of our proposed procedure is to construct an initial sparse linear approximation of $\Delta(\mathbf{X})$ using weighted learning (Chen et al. (2017), Tian et al. (2014)). This involves fitting a linear model for $\Delta(\mathbf{X})$ with a regularization term. For simplicity of presentation, we consider the LASSO (Tibshirani (1996)), but other penalties can also be used,

$$(2.6) \qquad \hat{\boldsymbol{\beta}}_1 = \underset{\boldsymbol{\beta}}{\arg\min} \sum_{i=1}^n \hat{\pi}(\mathbf{X}_i, A_i)^{-1}\left\{Y_i - \hat{m}(\mathbf{X}_i) - \frac{A_i}{2}\boldsymbol{\beta}^T \mathbf{X}_i\right\}^2 + \lambda_1 \sum_{j=1}^p |\beta_j|.$$

This results in coefficients $\hat{\boldsymbol{\beta}}_1$, which yield linear ITRs. We use K-fold cross-validation to choose the tuning parameter in (2.6) with the weighted mean-squared error as the criterion. In particular, we use the corresponding inverse propensity score to weight the error contribution of each subject. Thus, the tuning parameter is selected to optimize accuracy in estimating the linear portion of the interaction function $\Delta(\mathbf{X})$.

The next step is to construct residuals from the linear ITRs and the main effects, $r_i = Y_i - \hat{m}(\mathbf{X}_i) - A_i \hat{\boldsymbol{\beta}}_1^T \mathbf{X}_i/2$, for $i = 1, \dots n$. Thus, if the additive approximation (2.4) is well posited, the residual r_i will contain only the error term and any nonlinear components of $\Delta(\mathbf{X})$. Using these residuals as a working response, for each covariate X_j ($j = 1, \dots p$) in $\mathbf{X}$, we fit an inverse propensity weighted penalized smoothing spline model (Wahba (2006)) to estimate the additive nonlinear contribution of each covariate j to $\Delta(\mathbf{X})$. We then construct a design matrix of the fitted values from each of these covariate-specific models to be used as additional features in a final estimate of $\Delta(\mathbf{X})$. Constructing the nonlinear features one variable at a time ensures computational efficiency, while still maintaining an additive model (i.e., no between-variable interactions). In contrast to Tay and Tibshirani (2020), which used a fixed value for degrees of freedom, we use penalized splines with a data-adaptive value for

degrees of freedom (Krivobokova, Crainiceanu and Kauermann (2008)). While using fixed degrees of freedom, as in Tay and Tibshirani (2020), allows one to straightforwardly bound the overall degrees of freedom used, we have found that nonlinear terms for different co-variates can be either under- or oversmoothed if there is a wide degree of variation in the variability of the different additive terms. We have found this problem to be well mitigated by using penalized splines.

Let $\mathbf{G}_i$ denote the vector of nonlinear effects (i.e., the fitted values from the covariate-specific penalized splines) for the ith observation, where the jth entry of $\mathbf{G}_i$ is given by $g_j(X_{ij})$, where $g_j(\cdot)$ is the spline function fit by regressing r_i on the jth covariate X_{ij}. We define the combined coefficients for the linear and nonlinear components as $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_{\text{lin}}^T, \boldsymbol{\beta}_{\text{non}}^T)$ and their estimates as $\widehat{\boldsymbol{\beta}}^T = (\widehat{\boldsymbol{\beta}}_{\text{lin}}^T, \widehat{\boldsymbol{\beta}}_{\text{non}}^T)$. Finally, in order to select the final model, we perform a penalized regression of $Y_i - \hat{m}(\mathbf{X}_i)$ on $\mathbf{X}_i$ and $\mathbf{G}_i$,

$$
(2.7) \quad \begin{aligned}
\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \hat{\pi}(\mathbf{X}_i, A_i)^{-1} &\left\{ Y_i - \hat{m}(\mathbf{X}_i) - \frac{A_i}{2}(\boldsymbol{\beta}_{\text{lin}}^T \mathbf{X}_i + \boldsymbol{\beta}_{\text{non}}^T \mathbf{G}_i) \right\}^2 \\
&+ \lambda_2 \sum_{j=1}^{p} (|\beta_{\text{lin},j}| + \gamma_j |\beta_{\text{non},j}|),
\end{aligned}
$$

where γ_j is a data-driven penalty factor for the jth nonlinear term. In particular, we set $\gamma_j = \min(\sqrt{p}, 1 + s_j^{-1})$ with s_j defined as the standard deviation of the jth nonlinear term across all observations. Such a penalty factor allows us to penalize nonlinear terms based on their complexity, that is, the penalty factor for the jth nonlinear term is smaller for covariates whose estimated nonlinear terms have large variation and the penalty is larger for terms with small variation. Since γ_j is lower bounded by 1, it guarantees the penalization for nonlinear terms is heavier than the corresponding linear term, which agrees with our proposal to have a parsimonious rule. Furthermore, γ_j is upper bounded by $\sqrt{p}$ to prevent the overall penalty from being dominated by extreme penalty factors. The selection of the penalization term λ_2 will be discussed in more detail in Section 2.4; however, it requires great care as to not over-fit the data, since the criterion (2.7) contains several terms ($\hat{m}$ and $\mathbf{G}$) estimated from the data. Given a well-chosen value λ_2, the final estimate for $\Delta(\mathbf{X})$ is expressed as

$$
(2.8) \quad \hat{f}_{\lambda_2}(\mathbf{X}) = \widehat{\boldsymbol{\beta}}_{\text{lin}} \mathbf{X} + \widehat{\boldsymbol{\beta}}_{\text{non}} \mathbf{G},
$$

where inclusion of λ_2 in the notation indicates the dependence of the estimated function on the final tuning parameter. The final treatment rule is then $\hat{d}(\mathbf{X}) = \operatorname{sign}\{\hat{f}_{\lambda_2}(\mathbf{X})\}$. We note that both (2.6) and (2.7) can be optimized with existing software for the LASSO, such as glmnet, and thus each step is computationally simple and expedient.

The proposed procedure has multiple aspects for encouraging parsimonious ITRs and enhancing their interpretability. Following the reluctant additive principle, our method promotes preference for linear terms in two ways: 1.) nonlinear terms are fit using the residuals after removing the linear terms; 2.) nonlinear terms are penalized based on their complexity; therefore, they must contribute substantially to the ITR to survive the final model selection process in Step 3. Furthermore, traditional GAMs allow multiple nonlinear features representing the nonlinear effect of variable j. In contrast, our construction in Step 2 combines them as one feature. This ensures that after variable selection, one variable can contribute at most two terms to the final ITRs: one for the linear effect and the other for the nonlinear effect. This preference for linear terms combined with the simple structure, as shown in (2.8), allows for complexity, while still maintaining an easy-to-understand interpretation, which is a key advantage of our method compared to existing nonlinear ITR learning methods. Now that we have shown the main idea of our approach, we next provide implementation details that are critical to making RAITR work effectively, that is, estimation of $m(\mathbf{X})$ and $\pi(\mathbf{X}, A)$ and the tuning parameter (λ_2) selection.

2.3. *Nuisance parameter estimation with crossfitting.* Since we are focused on identifying individualized treatment rules, the parameters in the main effects $m(\mathbf{X})$ and propensity score $\pi(\mathbf{X}, A)$ models may be considered nuisance parameters. The work by Pan and Zhao (2021) shows that using flexible models for estimating $m(\mathbf{X})$ and $\pi(\mathbf{X}, A)$ is still useful in the sense that well-estimated main effects and propensity scores can lead to improved robustness and efficiency of the ITR estimator. However, in practice, we have seen that using a model that is too flexible for the nuisance parameter can introduce bias. In order to use a flexible model while simultaneously preventing overfitting, we use crossfitting (Chernozhukov et al. (2018)) with a gradient boosting model (Hastie, Tibshirani and Friedman (2009)) to estimate the main effects model. Using complex models for the nuisance parameters does not conflict with our goal of creating interpretable ITRs, since this does not affect the interpretability of the final treatment rules. Once the main effects and propensity score estimates are specified, we implement the augmentation procedure with our approach, as described in Section 2.2. The estimation of nuisance parameters is performed as follows:

(1) Take a K-fold random partition $(I_k)_{k=1}^K$ of observation indices $[n] = 1, ..., n$ such that the size of each fold I_k is $n_K = n/K$. Also, for each $k \in [K] = 1, ..., K$, define the complement set $I_k^c := \{1, ..., n\} \setminus I_k$.

(2) For each $k \in K$, using only data in using I_k^c, we construct the propensity score model and the main effect models. We denote these models as $\hat{\pi}^{I_k^c}(\mathbf{X}, A)$ and $\hat{m}^{I_k^c}(\mathbf{X})$, respectively.

Since the construction of the propensity score model and main effect model uses the out-of-fold samples, crossfitting prevents the model overfitting for these nuisance parameters. Though crossfitting to avoid overfitting of nuisance models is important, it is worth noting that the objective for the two nuisance models differs: optimizing fit (minimizing prediction error) in the main effect model can reduce variance and improve ITRs, whereas optimizing fit in the propensity score may lead to instability: the goal of this model is to achieve covariate balance between treatment groups, and ideally, minimizing covariate imbalance should be the objective of this model fit.

Once we estimate the nuisance parameters, to implement RAITR, we simply replace $\hat{\pi}(\mathbf{X}, A)$ and $\hat{m}(\mathbf{X})$ in Algorithm 1 with $\hat{\pi}^{I_k^c}(\mathbf{X}, A)$ and $\hat{m}^{I_k^c}(\mathbf{X})$, respectively. Adapting Algorithm 1 to use with estimators arising from out-of-fold data requires care; therefore, we explicitly show an algorithm (Algorithm 2) in the Supplementary Material (Maronge, Huling and Chen (2023)).

Now that we have described fitting the RAITR model, including the nuisance parameters, the final requirement is selecting the tuning parameter, λ_2, for the penalized regression of the final model. Next, we specify the criteria for selecting this parameter.

2.4. *Model selection criterion.* A natural choice for selecting λ_2 (and the corresponding ITR) is to perform K-fold cross-validation of the LASSO model in (2.7). However, an extra stage of cross-validation would increase required computation. Moreover, as we show in simulations and data example, cross-validation tends to overselect nonlinear effects. Lastly, conducting cross-validation for the final model can attenuate the benefit of crossfitting; that is, the estimation error of nuisance parameters and parameters of interest in the ITR would no longer be orthogonal, unless complex nested cross-validation strategy is used. Hence, we consider an alternative approach for selecting λ_2 by the concordance information criterion (CIC; Shi, Song and Lu (2021)). From our experimentation CIC performs favorably to cross-validation to control bias in the final ITRs of RAITR. As an added feature, the CIC does not require splitting the data into folds, thereby reducing computational load. The CIC is denoted

$$\mathrm{CIC}(\hat{f}_\lambda) = nC(\hat{f}_\lambda) - \kappa_n DF(\hat{f}_\lambda)),$$

where $C(\hat{f}_\lambda)$ is the concordance function for the ITR $(\text{sign}(\hat{f}_\lambda))$ governed by a tuning parameter λ that drives the complexity of the ITR, κ_n is a constant depending on n and/or p, and $DF(\hat{f}_\lambda)$ is the degrees of freedom of the corresponding ITR. In our setup we use $\lambda = \lambda_2$. Choices for the value of κ_n and $DF(\hat{f}_\lambda)$ may affect the performance of CIC, and optimal choices are still open research problems; however, we aim to provide recommendations based on numerical results. The CIC is composed of two terms: the first term is the concordance function, and the second is a penalty for ITR complexity, similar to BIC. The concordance function is defined as

$$C(\hat{f}_\lambda) = E\big[\{Y_i(1) - Y_i(-1)\} - \{Y_j(1) - Y_j(-1)\}\big] \times \mathbb{1}\{\hat{f}_\lambda(\mathbf{X}_i) > \hat{f}_\lambda(\mathbf{X}_j)\}.$$

Let $A = -1$ denote the reference treatment (either control, placebo, or active comparator treatment) and $A = 1$ denote the treatment of interest, the concordance is a measure of the degree to which patients who are more likely to be assigned to the treatment of interest also have larger individual treatment effects (relative to the reference treatment $A = -1$). The empirical estimator of the concordance is given by

$$
\begin{aligned}
(2.9) \quad \hat{C}(\hat{f}_\lambda) = \frac{1}{n(n-1)} \sum_{i \neq j} \Bigg[& \frac{\{Y_i - E(Y_i|\mathbf{X}_i = \mathbf{x}_i, A_i = -1)\}\{0.5A_i + 0.5 - \pi(\mathbf{x}_i, A_i = 1)\}}{\pi(\mathbf{x}_i, A_i = a_i)} \\
& - \frac{\{Y_j - E(Y_j|\mathbf{X}_i = \mathbf{x}_i, A_i = -1)\}\{0.5A_j + 0.5 - \pi(\mathbf{x}_j, A_j = 1)\}}{\pi(\mathbf{x}_j, A_j = a_j)} \Bigg] \\
& \times \mathbb{1}\{\hat{f}_\lambda(\mathbf{x}_i) > \hat{f}_\lambda(\mathbf{x}_j)\}.
\end{aligned}
$$

We also consider the doubly-robust estimator,

$$
\begin{aligned}
(2.10) \quad & \hat{C}_{\mathrm{DR}}(\hat{f}_\lambda) \\
& = \frac{1}{n(n-1)} \sum_{i \neq j} \\
& \times \Bigg[\frac{\{Y_i - E(Y_i|\mathbf{X}_i = \mathbf{x}_i, A_i = -1)\}\{0.5A_i + 0.5 - \pi(\mathbf{x}_i, A_i = 1)\}(0.5A_j + 0.5)}{\pi(\mathbf{x}_i, A_i = a_i)\pi(\mathbf{x}_j, A_j = 1)} \\
& \quad - \frac{\{Y_j - E(Y_j|\mathbf{X}_j = \mathbf{x}_j, A_j = -1)\}\{0.5A_j + 0.5 - \pi(\mathbf{x}_j, A_j = 1)\}(0.5A_i + 0.5)}{\pi(\mathbf{x}_j, A_j = a_j)\pi(\mathbf{x}_i, A_i = 1)} \Bigg] \\
& \times \mathbb{1}\{\hat{f}_\lambda(\mathbf{x}_i) > \hat{f}_\lambda(\mathbf{x}_j)\}.
\end{aligned}
$$

$\hat{C}_{\mathrm{DR}}(\hat{f}_\lambda)$ is doubly robust in that it is a consistent estimator of $C(\hat{f}_\lambda)$ when either the models for $\pi(\mathbf{x}, A)$ or $E(Y_i|\mathbf{X} = \mathbf{x}, A_i = -1)$ is correctly specified. Although Fan et al. (2017) and Shi, Song and Lu (2021) studied the theoretical properties of these estimators using κ_n of a certain order, we consider the numerical performance of different choices of κ_n. Particularly, we show two possible choices $\kappa_n = \log n$ or $\kappa_n = 2$, which were advocated by Shi, Song and Lu (2021) to mimic AIC and BIC for likelihood-based model selection. In the context of RAITR, we maximize the CIC to select the tuning parameter in our final model. Note that we can plug in the crossfitted version of the main effect model $(m^{I_k^c})$ and propensity score model $(\pi^{I_k^c})$ into ((2.9) or (2.10)) such that the resulting criterion may benefit from the nuisance parameter estimation described in Section 2.3. Obtaining a precise estimate of $DF(f_\lambda)$ is challenging in our setting and is an area of open work. We use a simple approximation that we have found is effective in practice. The choice of $DF(f_\lambda)$ is $\sum_{j=1}^p \mathbb{1}\{\hat{\beta}_{\mathrm{lin},j} \neq 0\} + \sum_{j=1}^p \mathbb{1}\{\hat{\beta}_{\mathrm{non},j} \neq 0\}\widehat{\mathrm{DF}}(g_j)$, where $\widehat{\mathrm{DF}}(g_j)$ is the estimated degrees of freedom for the estimated smoothing spline for the jth variable.

In our proposal we use $\lambda = \lambda_2$ and do not select λ_1 with the CIC metric, as λ_1 is set in the first stage. We do not choose λ_1 with CIC because we do not use the linear terms by themselves as our final treatment rule. Instead, our choice of λ_1 emphasizes accuracy in predicting the linear components of the interactions, whereas our choice of λ_2 emphasizes performance of the estimated ITRs resulting from our entire proposed estimation procedure.

3. Simulation. In this section we study the performance of our method and existing methods in a number of settings. In each setting we simulate n i.i.d samples $(Y_i, \mathbf{X}_i, A_i)$. Each X_j ($j = 1 \ldots p$) is distributed as $X_j \sim \text{Unif}(-2, 2)$, where "Unif" represents the continuous uniform distribution. The treatment A follows a binary distribution (-1 or 1), where $P(A = 1|\mathbf{X}) = \exp(\mathbf{X}^T \beta)/\{1 + \exp(\mathbf{X}^T \beta)\}$, with $\beta = (1, -1, 0.5, 0, 0, \ldots, 0)^T$. The response is normally distributed as $N(m(\mathbf{X}) + (A/2) \times \Delta(\mathbf{X}), 4)$. For these simulations $m(\mathbf{X})$ is given by the nonlinear function $m(\mathbf{X}) = -c \sum_{j=1}^{5}\{X_j + (2/3)(2X_j^2 - 1)\}$. We adjust the value of c to adjust the relative effect size of $\Delta(\mathbf{X})$ to the remaining components in the response function. For each data-generating model we investigate two values of c: the smaller value indicates a "large" effect size, and the larger value indicates a "small" effect size. We consider multiple models for $\Delta(\mathbf{X})$ with various degrees of complexity. The goal is to show that, in cases where $\Delta(\mathbf{X})$ is relatively simple (e.g., linear), our approach does not sacrifice simplicity; however, in complicated cases (e.g., highly nonlinear, polynomial, or tree-based models), RAITR is able to capture these complexities. The particular choices of $\Delta(\mathbf{X})$ are given as below:

1. Linear: $\Delta(\mathbf{X}) = \mathbf{X}^T \beta$, $\beta = (1, -1, 0.5, 1, -1, -0.5, 0, 0, \ldots, 0)^T$, $c = 0.1$ or 3.
2. Highly Nonlinear: $\Delta(\mathbf{X}) = X_1^3 + |X_3| \exp\{X_5\} + 5\sin(2\pi X_7) + 5\cos(2\pi X_6) - (X_4 + X_8)^2 + 3|X_2 + X_5|$, $c = 1$ or 8.
3. Tree: $\Delta(\mathbf{X}) = 3\{\mathbb{1}(X_1 + 0.5 > 0) \times \text{sign}(X_2 - 0.5)\} + 2.5\{\mathbb{1}(X_1 + 0.5 < 0) \times \text{sign}(X_4 - 0.5)\} + 0.5$, $c = 1$ or 5.
4. Polynomial: $\Delta(\mathbf{X}) = 0.2 + X_1^2 + X_2^2 - X_3^2 - X_4^2$, $c = 0.1$ or 2.

The optimal ITRs are additive in $\mathbf{X}$ in the linear and polynomial settings but are not additive in the highly nonlinear and tree settings.

We train the ITRs with various sample sizes ($n = 250, 500, 1000$) and covariate dimension ($p = 100, 500, 1000$) combinations, then evaluate the ITR on a testing set of size 10,000. We report two commonly used metrics for ITR performance evaluation: the value function of the estimated rule and the agreement between the estimated rule and the optimal rule. In particular, for a given rule $\hat{d}(\mathbf{X})$, we can directly estimate $E[Y(\hat{d}(\mathbf{X}))]$ and $E[\hat{d}(\mathbf{X}) = d^*(\mathbf{X})]$ by evaluating their empirical versions on the testing data.

We estimate ITRs using two versions of a strictly linear rule based on the weighted learning (Chen et al. (2017)) with the least square loss (i.e., using only the first three steps of the algorithm in Supplementary Material Section A to estimate ITRs), generalized random forests (GRF; Athey, Tibshirani and Wager (2019), Wager and Athey (2018)), sparse additive model for treatment effect-modifier selection (samTEMsel; Park et al. (2022)), and different variations of our method, RAITR. We did not include support vector machine-based nonlinear ITR learning methods, such as Zhao et al. (2012), Zhou et al. (2017), in the comparison, as they are not suited for the high dimensional covariate settings considered in this paper. Additional information about the simulation scenarios, including computational load, covariate balancing, and a formal definition of "small" and "large" effect sizes can be found in Supplementary Material Sections B.1, B.2, and B.3, respectively.

In Figures 1 to 4, we show results across 100 replicates in terms of the agreement. In Supplementary Material Section B.3, we also display results in terms of the value function. For linear ITRs and RAITR, we either select tuning parameters using cross-validation (CV)
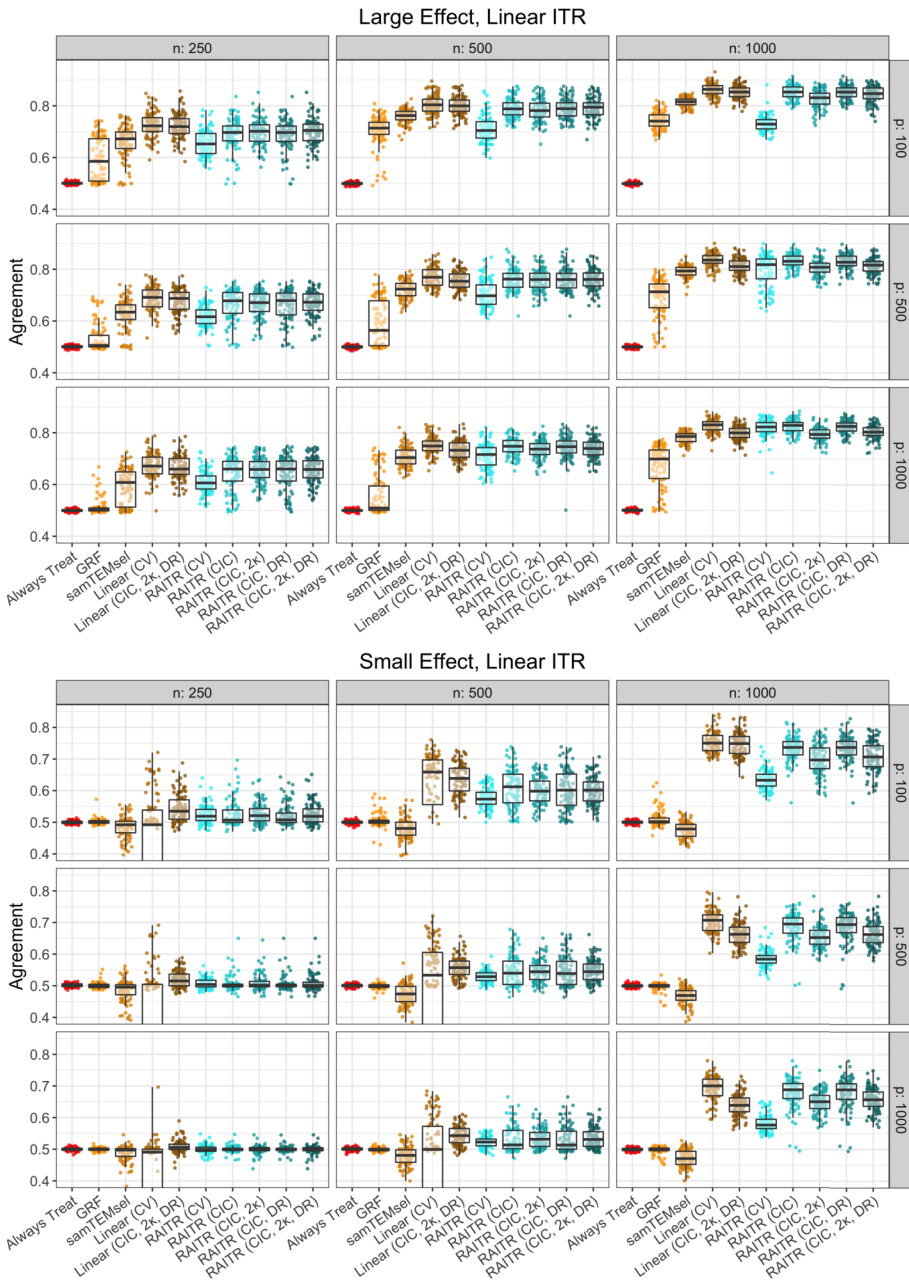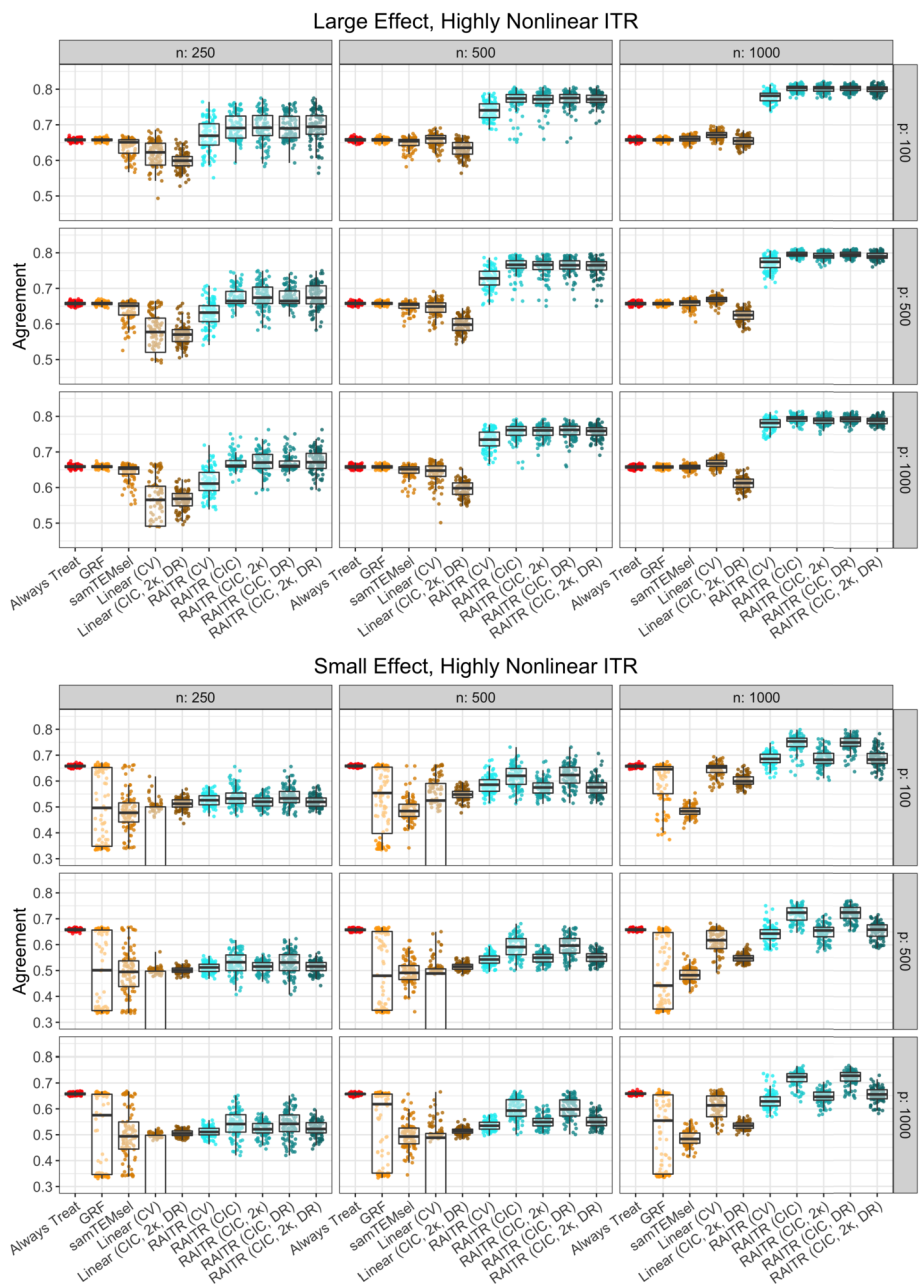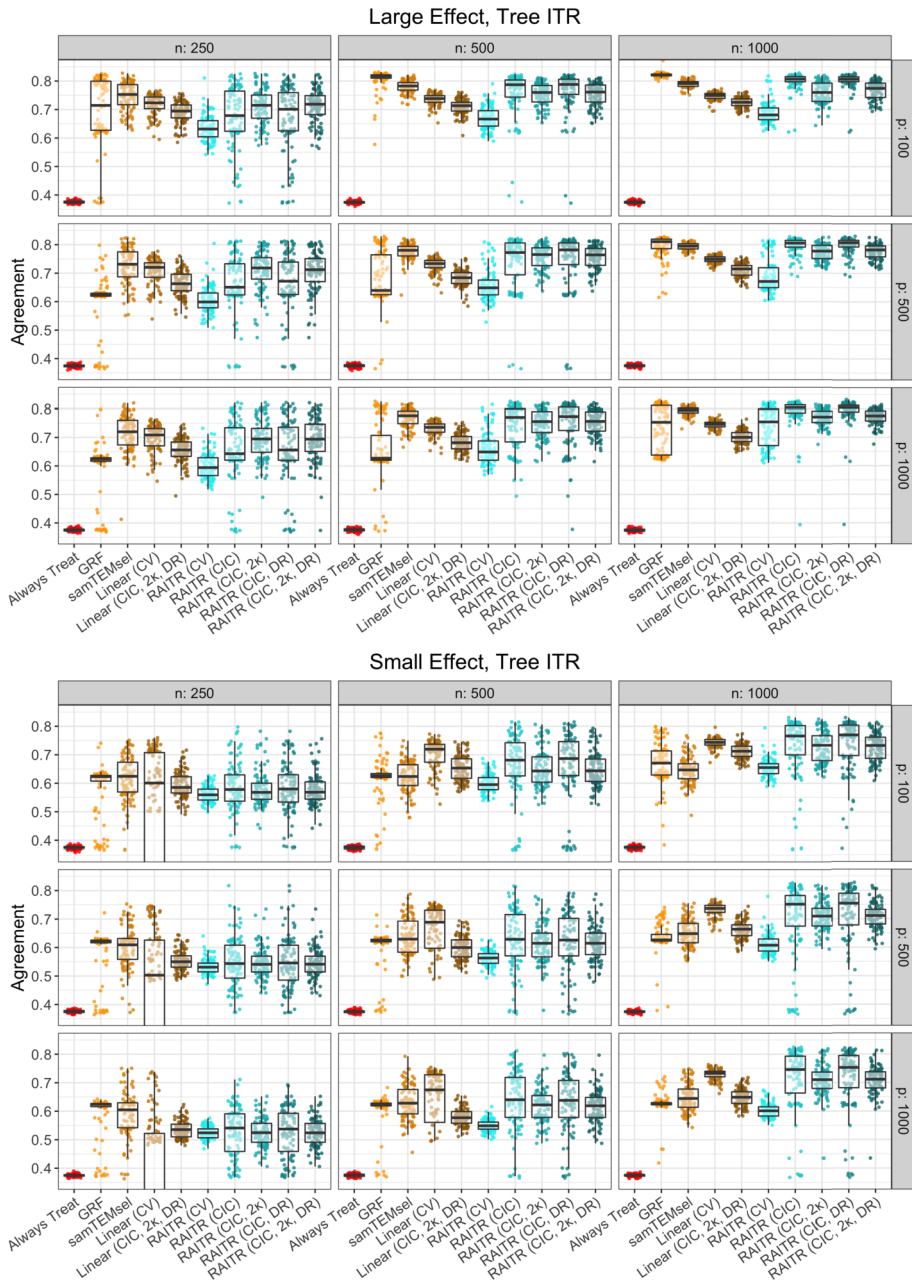
FIG. 1. *Top*: *Results for agreement for large effect size, linear data-generating model for* $n = 250, 500, 1000$ *and* $p = 100, 500, 1000$ *for RAITR, linear ITRs, GRF, and samTEMsel. Bottom*: *As for top but small effect size. The vertical axis limits of the bottom plot have been modified for clarity.*

or the concordance information criterion (CIC). When CIC is used, we either set $\kappa_n = \log n$ or $\kappa_n = 2$ (2κ), and either use the standard estimate for concordance or the doubly-robust (DR) version. For other methods the tuning parameters are selected by CV.

Regarding the linear setting, unsurprisingly, the linear ITR analyses perform best, but RAITR, particularly with tuning parameters selected using CIC, performs comparably to the linear rules. Using RAITR with final models selected by CV generally performs poorly because the model over-selects nonlinear terms. The samTEMsel method generally performs relatively poorly, compared to linear ITRs, suggesting a tendency for it to select overly com-

FIG. 2. *Top*: *Results for agreement for large effect size, highly nonlinear data-generating model for* $n = 250, 500, 1000$ *and* $p = 100, 500, 1000$ *for RAITR, linear ITRs, GRF, and samTEMsel. Bottom*: *As for top but small effect size. The vertical axis limits of the bottom plot have been modified for clarity.*

plex ITRs when the underlying data structure is simple. Similarly, GRF struggles, due to the linear ITR structure, and is particularly poor in high-dimensional settings, as it does not have an explicit variable selection component.

In the highly nonlinear setting, RAITR with CIC consistently outperforms all other methods, particularly with a large effect size. Due to the highly complex and nonadditive nature of the ITR in this setting, there appears to be an upper bound on how well RAITR can perform. Despite this, in the "large" effect setting, RAITR is desirable compared to the other nonlinear
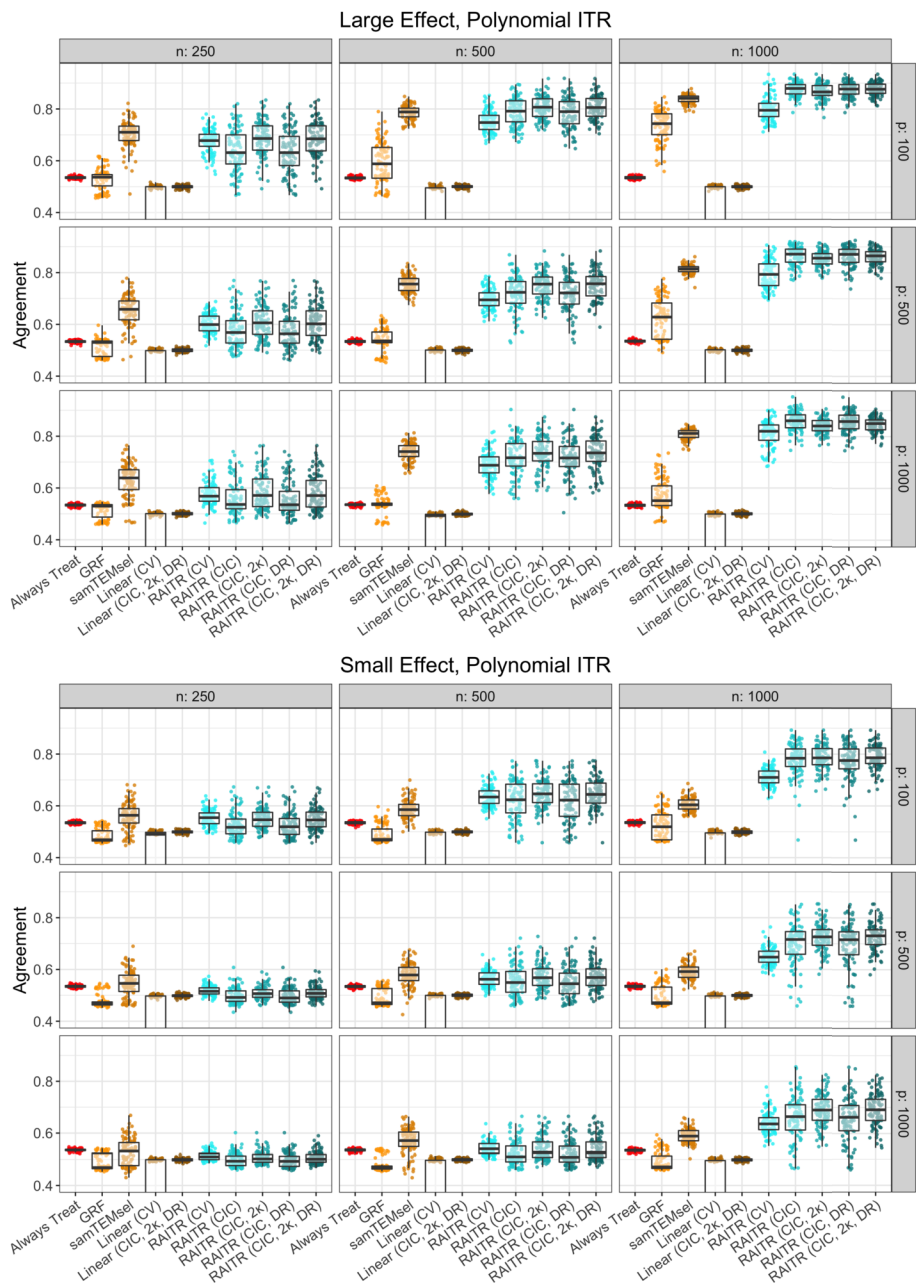
FIG. 3. *Top*: *Results for agreement for large effect size, tree data-generating model for n* = 250, 500, 1000 *and* p = 100, 500, 1000 *for RAITR, linear ITRs, GRF, and samTEMsel. Bottom*: *As for top but small effect size.*

ITRs. In the "small effect" setting, no method except RAITR (with a large enough sample size) does better than the naive "always treat" rule.

In the tree setting, GRF outperforms all other methods for some scenarios but performs poorly when the dimension is large, again due to its lack of an explicit variable selection component. Using samTEMsel or RAITR with CIC consistently performs well. In the polynomial setting, RAITR and samTEMsel perform much better than all other methods, with RAITR performing the best when the sample size is large enough to support the augmentation procedure.

FIG. 4. *Top*: *Results for agreement for large effect size, polynomial data-generating model for* $n = 250, 500, 1000$ *and* $p = 100, 500, 1000$ *for RAITR, linear ITRs, GRF, and samTEMsel. Bottom*: *As for top but small effect size.*

In general, RAITR with CIC regularly performs at least as well, and many times, better than the other approaches considered. A caveat is that the augmentation procedure tends to require a large sample size to perform well. However, even in the settings with smaller sample size, RAITR performs comparably to other approaches and in the settings with large sample size often performs much better. In many cases, when $\kappa_n = \log n$, RAITR performs better on average than when $\kappa_n = 2$; however, it appears that the 2κ version can better protect against poor performance. Therefore, the choice of κ_n should be made relative to the application of interest. In these simulations, RAITR with doubly-robust CIC performed comparably
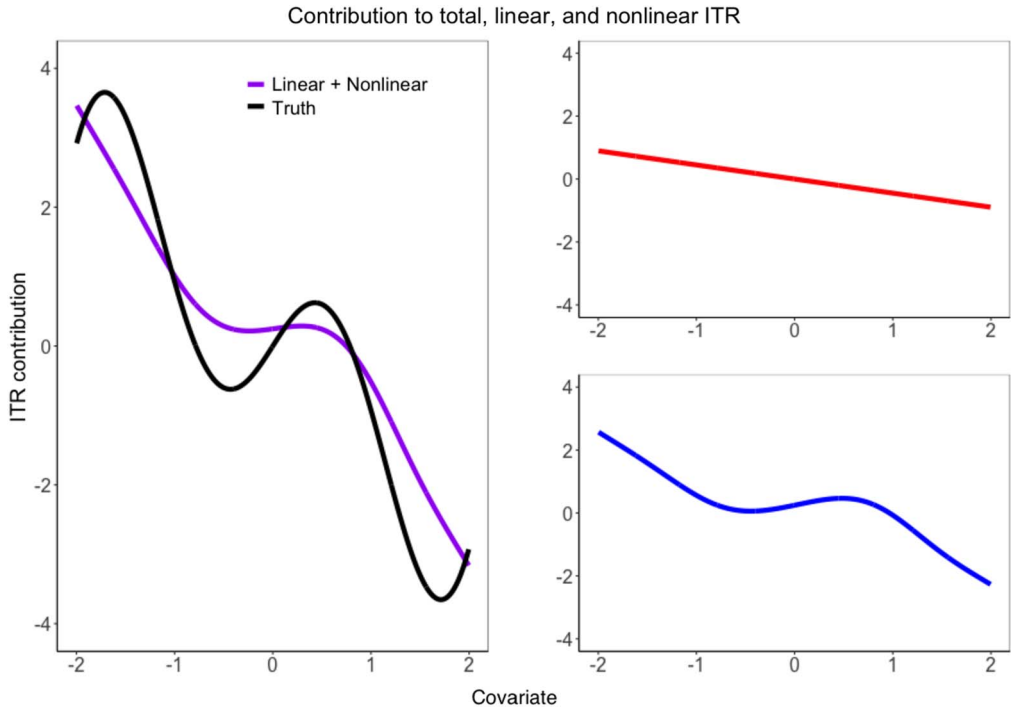
Contribution to total, linear, and nonlinear ITR

FIG. 5. *Example of resulting ITR from RAITR (CIC, 2κ, DR) from one repetition of the highly nonlinear simulation setting (left) as well as contribution from the linear component (top-right) and nonlinear component (bottom-right).*

to the standard version. Additionally, RAITR performs well in the "small effect" setting, particularly when the sample size is large enough to support augmentation. Finally, increasing dimensionality can cause deterioration in performance for other methods, such as GRF, samTEMsel, and linear rules, but RAITR appears to be more robust to this problem. Notably, the results for the linear ITR and highly nonlinear ITR settings jointly emphasize RAITR's ability to balance performance and return either simple ITRs or sufficiently complex ITRs when necessary. Note that Shi, Song and Lu (2021) also proposed the value information criterion (VIC) for ITR model selection as well as CIC. In their study Shi, Song and Lu (2021) observed that CIC performed better than VIC empirically. We also observed a similar phenomenon in the early development of our method. As such, we focus on comparing the results of RAITR by using either CIC or the commonly used method of CV as the model selection criterion.

As a final point of emphasis, a key reason for RAITR to yield a highly interpretable result is that each covariate can only contribute at most two terms to the final ITR. This is demonstrated in Figure 5. For demonstration we generate data using a similar model to the highly nonlinear setting, as described before; except now, we have $\Delta(\mathbf{X}) = (1/2)\cos(2X_1)(2X_1)$, $n = 1000$, and $p = 10$. We show the contribution of X_1 to the resulting ITR from one repetition of RAITR (CIC, 2κ, DR). On the left we plot the covariate vs. $\hat{f}(\mathbf{X}) = \hat{\boldsymbol{\beta}}_{\text{lin}}\mathbf{X} + \hat{\boldsymbol{\beta}}_{\text{non}}\mathbf{G}$ as well as the true ITR. On the top-right and bottom-right, we show only the linear and nonlinear contribution, respectively. The figure demonstrates that RAITR allows us to visualize the contribution of a covariate to an ITR in a simple manner, which results in a higher level of interpretability and can be quite useful in practice.

4. Analysis of GDSC data. The Genomics of Drug Sensitivity in Cancer (GDSC) study measures the response of 1018 human cancer cell lines to 265 anticancer drugs (Yang et al.

(2012)). In particular, it contains detailed genetic information, including gene expression and chromosomal copy number, pharmacological information and treatment responses on human cancer cell lines. As noted by Iorio et al. (2016), since cancer cell lines are replicable, we may test multiple drugs on the same cell line in order to compare drug sensitivity to a response of interest. This makes the GDSC data and similar datasets, such as Cancer Cell Line Encyclopedia (Ghandi et al. (2019)), a particularly interesting resource for benchmarking the performance of different ITRs.

In this paper, we focus on the outcome IC_{50}, which is the drug concentration that reduces viability by 50%. This measurement was collected on each of the 1018 cancer cell lines. A smaller value for IC_{50} means the drug is more effective. Hence, to be consistent with the notation of a large value is desirable, we use $-\log(IC_{50})$ the log transformation is done following the procedure used by Yang et al. (2012) as the outcome. We use the gene expression data (17,419 genes in total) of each cancer line as the predictors, and our aim is to recommend the drug that is more effective between a pair of drugs. Gene expression was chosen because it was the most predictive data type for drug sensitivity in the GSDC study (Iorio et al. (2016)).

The minimum and maximum of the mean $\log(IC_{50})$ of the considered drugs are -5.59 and 8.00, respectively. There are 133 drugs with the mean $\log(IC_{50})$ less than 2. Among these drugs with relatively small $\log(IC_{50})$, there are 235 drug pairs with similar $\log(IC_{50})$ values among cancer cells: that is, the empirical probability of one drug being the optimal treatment is greater than 0.35 and less than 0.65. Lastly, we select the 26 drug pairs out of 235 drug pairs that are most likely to have a nonlinear relationship between $\log(IC_{50})$ and the gene expression level. The complete results for all 26 pairs are given in Supplementary Material Section C. Here we show results from six drug pairs that are representative: (1) SN-38 and Vinblastine, (2) JQ1 and Navitoclax, (3) IGFR_3801 and CD532, (4) AZD2014 and IOX2, (5) Cytarabine and IMD-0354, and (6) Refametinib and AZD6738. These drug combinations were chosen because they gave a large variety in performance for the methods under consideration.

We use the median absolute deviation (MAD) to filter out the gene expressions with low variability (MAD < 1), resulting in 1486 genes for the downstream analyses. The number of cells (sample size) varies for different drug pairs due to different degree of outcomes missing—a full table of number of cells for each of the 26 drug pairs is given in Table C.1 of the Supplementary Material but is approximately 900 cells for each drug pair.

We randomly split the data into training (two-thirds) and testing (one-third), then evaluate trained models on the testing data. We repeat this procedure 100 times. Even though we can observe the response for both treatments, we construct the training data to be more comparable to learning ITRs in a standard clinical setting. In particular, for each cell we randomly select one treatment as the "observed" treatment and the corresponding $-\log(IC_{50})$ as the "observed" outcome and discard the remaining treatment and outcome-related information for model training. On the testing data, we can directly calculate the empirical agreement and value function of the trained model on the testing data, as we did in the simulation study. Typically, we would need to estimate the propensity score and/or outcome regression function for calculating the empirical agreement and the value function, but due to the special structure of the GDSC data, this is not required. Finally, we report the average agreement and value function over the 100 replications. A selection of the results with respect to the agreement is given in Table 1. The complete results with the agreement and the value function for all 26 drug pairs are given in Supplementary Material Section C.

As seen in Table 1, RAITR, particularly RAITR (CIC, 2κ, DR), performs well across all drug pairs, which indicates a high level of flexibility. In two out of the six reported drug pairs RAITR (CIC, 2κ, DR) performs best and in the remaining settings still performs strongly.

TABLE 1

Results from GDSC analysis for a selection of drug pairs. In one row linear method is denoted L. to increase clarity. Results are given in terms of mean agreement across 100 replicates with the associated standard errors in parentheses. The highest performing method(s) for each drug is in bold

	SN-38 and Vinblastine	JQ1 and Navitoclax	IGFR_3801 and CD532	AZD2014 and IOX2	Cytarabine and IMD-0354	Refametinib and AZD6738
Method	Mean agreement (standard deviation), 100 replicates					
Always Treat	0.517 (0.022)	0.593 (0.025)	0.599 (0.025)	0.536 (0.023)	0.538 (0.025)	0.496 (0.024)
GRF	0.527 (0.030)	**0.659** (0.025)	0.654 (0.029)	0.534 (0.040)	0.605 (0.034)	0.681 (0.026)
samTEMsel	0.562 (0.043)	0.647 (0.025)	0.656 (0.024)	0.528 (0.041)	0.576 (0.035)	0.687 (0.027)
Linear (CV)	0.627 (0.048)	0.618 (0.042)	0.653 (0.042)	0.531 (0.041)	0.603 (0.036)	**0.697** (0.026)
L. (CIC, 2κ, DR)	**0.630** (0.034)	0.621 (0.027)	0.661 (0.026)	**0.551** (0.028)	0.609 (0.029)	0.696 (0.025)
RAITR (CV)	0.613 (0.037)	0.630 (0.033)	0.655 (0.035)	0.543 (0.037)	0.599 (0.034)	0.687 (0.029)
RAITR (CIC)	0.600 (0.052)	0.631 (0.042)	0.645 (0.044)	0.538 (0.036)	0.590 (0.033)	0.677 (0.037)
RAITR (CIC, 2κ)	0.627 (0.038)	0.641 (0.032)	**0.667** (0.029)	0.549 (0.031)	0.612 (0.026)	**0.697** (0.026)
RAITR (CIC, DR)	0.599 (0.054)	0.631 (0.039)	0.647 (0.038)	0.539 (0.037)	0.592 (0.032)	0.680 (0.036)
RAITR (CIC, 2κ, DR)	0.628 (0.038)	0.642 (0.031)	**0.667** (0.029)	0.550 (0.030)	**0.613** (0.024)	0.696 (0.026)

Each of the other methods performs well for particular drug pairs but also performs poorly in other pairs. The high degree of flexibility, together with the aforementioned interpretability makes RAITR a desirable method for ITR identification.

5. Discussion. We propose a new method, RAITR, for learning flexible yet interpretable nonlinear ITRs based on a reluctant generalized additive model. RAITR is designed with high dimensionality in mind and efficiently adapts to the smoothness of the underlying ITR. RAITR is parsimonious by supplying a simple modeling framework that only includes complex terms that are essential for the ITR. Even when complex, nonlinear terms are selected; our framework allows for a simple data-visualization tool in order to increase interpretability and collaboration. On the other hand, using the reluctant additive framework creates challenges, which we have addressed by combining crossfitting with CIC, a modern information-based model selection criterion.

As is seen in both the data analysis and simulations, there are times when the linear ITR is enough for an accurate treatment decision. This is because, even if the underlying heterogeneous treatment effect is highly nonlinear, the resulting optimal ITR may be well approximated by a linear ITR. As such, it is imperative to have highly flexible methods which perform well in simpler settings, such as RAITR. Despite the demonstrated advantages of RAITR, one key area of concern with respect to performance is that the augmentation procedure seems to require a relatively large sample size, particularly in high-dimensional settings, to perform well. For this reason and in extremely small sample size settings, we would recommend either using a simpler parametric model for augmentation or omitting the augmentation step (omitting would not impact bias). Furthermore, because our method is designed to "fall back" to a linear ITR when nonlinear terms are not warranted by the data, our approach is almost always competitive with linear ITR approaches. Lastly, regarding when it is appropriate to estimate an ITR, the general principle would also apply: if the sample size is simply too small, it may not feasible to estimate an ITR using any method. If the dimensionality is extremely high, it is advisable to use a variable screening procedure to eliminate variables that do not have marginal interactions with the treatment.

Note that the current implementation of RAITR does not enforce the strong heredity property, which is the notion of including a nonlinear term in the model only when the corresponding linear term has also been included. Even though the chance of violation of strong heredity property with RAITR is low due to how the nonlinear terms are constructed, incorporating strong heredity would be an important future area of research. For example, one approach would be to use the penalty proposed in Bian et al. (2023) to ensure strong heredity. Moreover, even though RAITR is motivated by learning ITRs for observational studies, it is directly applicable to data from controlled experiments and randomized trials. Due to a commonly observed phenomenon of variance reduction when known propensity scores are estimated, the estimated propensity score would still be used in RAITR, since even a well-executed randomized trial could have finite sample covariate imbalance, and estimating average treatment effect has shown that it is beneficial to use the estimated (correctly-specified) propensity score, even if the true treatment assignment mechanism is known (Hirano, Imbens and Ridder (2003)).

In this work we focused on extending linear ITRs to nonlinear ITRs under the weighted learning framework; however, our method can be readily applicable to other ITR learning frameworks whose current implementations are focused on learning linear ITRs such as A-learning (Chen et al. (2017), Nie and Wager (2021)), dynamic weighted ordinary least squares (Wallace and Moodie (2015)), and more. Moreover, the RAITR approach was designed with continuous outcomes in mind, but the justification of the loss does not require the outcome to be continuous and could thus be applied to count or binary outcomes. However, a similar approach could be developed for other types of responses using the general class of loss functions explored in Chen et al. (2017) and Huling and Yu (2021). Other areas of future research include extending RAITR for multicategory treatments (Qi et al. (2020)) or continuous treatments (Chen, Zeng and Kosorok (2016)) and allowing for the inclusion of pairwise interaction effects via the reluctant interaction modeling framework (Yu, Bien and Tibshirani (2019)).

SUPPLEMENTARY MATERIAL

Supplement to "A reluctant additive model framework for interpretable nonlinear individualized treatment rules" (DOI: 10.1214/23-AOAS1767SUPPA; .pdf). Algorithm 2, additional simulation and real data analysis results, and information on computation times.

Code (DOI: 10.1214/23-AOAS1767SUPPB; .zip). The software for the proposed method in R.

REFERENCES

ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963 https://doi.org/10.1214/18-AOS1709

BIAN, Z., MOODIE, E. E. M., SHORTREED, S. M. and BHATNAGAR, S. (2023). Variable selection in regression-based estimation of dynamic treatment regimes. *Biometrics* **79** 988–999. MR4606331 https://doi.org/10.1111/biom.13608

CHEN, G., ZENG, D. and KOSOROK, M. R. (2016). Personalized dose finding using outcome weighted learning. *J. Amer. Statist. Assoc.* **111** 1509–1521. MR3601705 https://doi.org/10.1080/01621459.2016.1148611

CHEN, S., TIAN, L., CAI, T. and YU, M. (2017). A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics* **73** 1199–1209. MR3744534 https://doi.org/10.1111/biom.12676

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

FAN, C., LU, W., SONG, R. and ZHOU, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1565–1582. MR3731676 https://doi.org/10.1111/rssb.12216

GHANDI, M., HUANG, F. W., JANÉ-VALBUENA, J., KRYUKOV, G. V., LO, C. C., MCDONALD, E. R., BARRETINA, J., GELFAND, E. T., BIELSKI, C. M. et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569** 503–508.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826 https://doi.org/10.1111/1468-0262.00442

HULING, J. D. and YU, M. (2021). Subgroup identification using the personalized package. *J. Stat. Softw.* **98** 1–60.

IORIO, F., KNIJNENBURG, T. A., VIS, D. J., BIGNELL, G. R., MENDEN, M. P., SCHUBERT, M. et al. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell* **166** 740–754.

KRIVOBOKOVA, T., CRAINICEANU, C. M. and KAUERMANN, G. (2008). Fast adaptive penalized splines. *J. Comput. Graph. Statist.* **17** 1–20. MR2424792 https://doi.org/10.1198/106186008X287328

LIANG, M., YE, T. and FU, H. (2018). Estimating individualized optimal combination therapies through outcome weighted deep learning algorithms. *Stat. Med.* **37** 3869–3886. MR3873688 https://doi.org/10.1002/sim.7902

LIN, Y. and ZHANG, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* **34** 2272–2297. MR2291500 https://doi.org/10.1214/009053606000000722

MARONGE, J. M., HULING, J. D. and CHEN, G. (2023). Supplement to "A reluctant additive model framework for interpretable nonlinear individualized treatment rules." https://doi.org/10.1214/23-AOAS1767SUPPA, https://doi.org/10.1214/23-AOAS1767SUPPB

MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Ann. Statist.* **37** 3779–3821. MR2572443 https://doi.org/10.1214/09-AOS692

MI, X., ZOU, F. and ZHU, R. (2019). Bagging and deep learning in optimal individualized treatment rules. *Biometrics* **75** 674–684. MR3999189 https://doi.org/10.1111/biom.12990

MURDOCH, W. J., SINGH, C., KUMBIER, K., ABBASI-ASL, R. and YU, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **116** 22071–22080. MR4030584 https://doi.org/10.1073/pnas.1900654116

NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133 https://doi.org/10.1093/biomet/asaa076

PAN, Y. and ZHAO, Y.-Q. (2021). Improved doubly robust estimation in learning optimal individualized treatment rules. *J. Amer. Statist. Assoc.* **116** 283–294. MR4227694 https://doi.org/10.1080/01621459.2020.1725522

PARK, H., PETKOVA, E., TARPEY, T. and OGDEN, R. T. (2022). A sparse additive model for treatment effect-modifier selection. *Biostatistics* **23** 412–429. MR4424978 https://doi.org/10.1093/biostatistics/kxaa032

PETERSEN, A. and WITTEN, D. (2019). Data-adaptive additive modeling. *Stat. Med.* **38** 583–600. MR3902599 https://doi.org/10.1002/sim.7859

PETERSEN, A., WITTEN, D. and SIMON, N. (2016). Fused lasso additive model. *J. Comput. Graph. Statist.* **25** 1005–1025. MR3572026 https://doi.org/10.1080/10618600.2015.1073155

QI, Z., LIU, D., FU, H. and LIU, Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *J. Amer. Statist. Assoc.* **115** 678–691. MR4107672 https://doi.org/10.1080/01621459.2018.1529597

QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351 https://doi.org/10.1214/10-AOS864

RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 1009–1030. MR2750255 https://doi.org/10.1111/j.1467-9868.2009.00718.x

RUBIN, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100** 322–331. MR2166071 https://doi.org/10.1198/016214504000001880

RUDIN, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1** 206.

SHI, C., SONG, R. and LU, W. (2021). Concordance and value information criteria for optimal treatment decision. *Ann. Statist.* **49** 49–75. MR4206669 https://doi.org/10.1214/19-AOS1908

TAY, J. K. and TIBSHIRANI, R. (2020). Reluctant generalized additive modeling. *Int. Stat. Rev.* **88** S205–S224.

TIAN, L., ALIZADEH, A. A., GENTLES, A. J. and TIBSHIRANI, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *J. Amer. Statist. Assoc.* **109** 1517–1532. MR3293607 https://doi.org/10.1080/01621459.2014.951443

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

VAYENA, E., BLASIMME, A. and COHEN, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Med.* **15** e1002689.

WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. MR3862353 https://doi.org/10.1080/01621459.2017.1319839

WAHBA, G. (2006). *Splines in Nonparametric Regression* **4**. Wiley Online Library, London, UK.

WALLACE, M. P. and MOODIE, E. E. M. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* **71** 636–644. MR3402599 https://doi.org/10.1111/biom.12306

YANG, W., SOARES, J., GRENINGER, P., EDELMAN, E. J., LIGHTFOOT, H., FORBES, S., BINDAL, N., BEARE, D., SMITH, J. A. et al. (2012). Genomics of drug sensitivity in cancer (gdsc): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41** D955–D961.

YU, G., BIEN, J. and TIBSHIRANI, R. (2019). Reluctant interaction modeling. Preprint. Available at: arXiv:1907.08414.

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898 https://doi.org/10.1080/01621459.2012.695674

ZHOU, X., MAYER-HAMBLETT, N., KHAN, U. and KOSOROK, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *J. Amer. Statist. Assoc.* **112** 169–187. MR3646564 https://doi.org/10.1080/01621459.2015.1093947

ZHU, R., ZHAO, Y.-Q., CHEN, G., MA, S. and ZHAO, H. (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics* **73** 391–400. MR3665956 https://doi.org/10.1111/biom.12593