Title

GEAR: On optimal decision making with auxiliary data

Permalink

https://escholarship.org/uc/item/7gd7z04s

Journal

Stat, 10(1)

ISSN

2049-1573

Authors

Cai, Hengrui
Song, Rui
Lu, Wenbin

Publication Date

2021-12-01

DOI

10.1002/sta4.399

Peer reviewed

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

GEAR: On Optimal Decision Making with Auxiliary Data

Hengrui Cai* | Rui Song | Wenbin Lu

Department of Statistics, North Carolina State University, Raleigh, USA

Correspondence
*Hengrui Cai, Department of Statistics, North Carolina State University, Raleigh, NC 27695. Email: hcai5@ncsu.edu

Present Address
Department of Statistics, North Carolina State University, Raleigh, NC 27695

Summary

Personalized optimal decision making, finding the optimal decision rule (ODR) based on individual characteristics, has attracted increasing attention recently in many fields, such as education, economics, and medicine. Current ODR methods usually require the primary outcome of interest in samples for assessing treatment effects, namely the experimental sample. However, in many studies, treatments may have a long-term effect, and as such the primary outcome of interest cannot be observed in the experimental sample due to the limited duration of experiments, which makes the estimation of ODR impossible. This paper is inspired to address this challenge by making use of an auxiliary sample to facilitate the estimation of ODR in the experimental sample. We propose an auGmented inverse propensity weighted Experimental and Auxiliary sample-based decision Rule (GEAR) by maximizing the augmented inverse propensity weighted value estimator over a class of decision rules using the experimental sample, with the primary outcome being imputed based on the auxiliary sample. The asymptotic properties of the proposed GEAR estimators and their associated value estimators are established. Simulation studies are conducted to demonstrate its empirical validity with a real AIDS application.

KEYWORDS:
Augmented inverse propensity weighted estimation; Auxiliary data; Individualized treatment rule; Optimal treatment decision making

1 | INTRODUCTION

Personalized optimal decision making, finding the optimal decision rule (ODR) based on individual characteristics to maximize the mean outcome of interest, has attracted increasing attention recently in many fields. Examples include offering customized incentives to increase sales and level of engagement in the area of economics (Turvey 2017), developing an individualized treatment rule for patients to optimize expected clinical outcomes of interest in precision medicine (Chakraborty & Moodie 2013), and designing a personalized advertisement recommendation system to raise the click rates in the area of marketing (Cho, Kim, & Kim 2002).

The general setup for finding the ODR contains three components in an experimental sample (from either randomized trials or observational studies): the covariate information (X), the treatment information (A), and the outcome of interest (Y). However, current ODR methods cannot be applied to cases where treatments have a long-term effect and the primary outcome of interest cannot be observed in the experimental sample. Take the AIDS Clinical Trials Group Protocol 175 (ACTG 175) data (Hammer et al. 1996) as an example. The experiment randomly assigned HIV-infected patients to competitive antiretroviral regimens, and recorded their CD4 count (cells/mm3) and CD8 count over time. A higher CD4 count usually indicates a stronger immune system. However, due to the limitation of the follow-up, the clinical meaningful long-term outcome of interest for the AIDS recovery may be missing for a proportion of patients. Similar problems are also considered in the evaluation of education programs, such as the Student/Teacher Achievement Ratio (STAR) project (Chetty et al. 2011; Word et al. 1990) that studied long-term impacts of early childhood

[0] **Abbreviations:** ACTG 175, AIDS Clinical Trials Group Protocol 175; AIPW, augmented inverse propensity weighted; CI, confidence interval; GEAR, auGmented inverse propensity weighted Experimental and Auxiliary sample-based decision Rule; I.I.D., independent and identically distributed; ODR, optimal decision rule

education on the future income. Due to the heterogeneity in individual characteristics, one cannot find a unified best treatment for all subjects. However, the effects of treatment on the long-term outcome of interest can not be evaluated using the experimental data solely. Hence, deriving an ODR to maximize the expected long-term outcome based on baseline covariates obtained at an early stage is challenging.

This paper is inspired to address the challenge of developing ODR when the long-term outcome cannot be observed in the experimental sample. Although the long-term outcome may not be observed in the experimental sample, we could instead obtain some intermediate outcomes (also known as surrogacies or proximal outcomes, M) that are highly related to the long-term outcome after the treatment was given. For instance, the CD4 and CD8 counts recorded after a treatment is assigned, have a strong correlation with the healthy of the immune system, and thus can be viewed as intermediate outcomes. A natural question is whether an ODR to maximize the expected long-term outcome can be estimated based on the experimental sample (that consists of $\{X, A, M\}$) only. The answer is generally no mainly for two reasons. First, it is common and usually necessary to have multiple intermediate outcomes to characterize the effects of treatment on the long-term outcome. However, when there are multiple intermediate outcomes, it is hard to determine which intermediate outcome or what combination of intermediate outcomes will lead to the best ODR for the long-term outcome. Second, to derive the ODR that maximizes the expected long-term outcome of interest based on the experimental sample, we need to know the relationship between the long-term outcome, intermediate outcomes and baseline covariates, which is generally not practical.

In this work, we propose using an auxiliary data source, namely the auxiliary sample, to recover the missing long-term outcome of interest in the experimental sample, based on the rich information of baseline covariates and intermediate outcomes. Auxiliary data, such as electronic medical records or administrative records, are now widely accessible. These data usually contain rich information for covariates, intermediate outcomes, and the long-term outcome of interest. However, since they are generally not collected for studying treatment effects, treatment information may not be available in auxiliary data. In particular, in this work, we consider the situation that an auxiliary data consisting of $\{X, M, Y\}$ is available, where Y is the long-term outcome of interest. Note it is also impossible to derive ODR based on such auxiliary sample due to missing treatments.

1.1 | Related Works

There is a huge literature on learning the ODR, including Q-learning (Qian & Murphy 2011; Watkins & Dayan 1992; Y. Zhao, Kosorok, & Zeng 2009), A-learning (Murphy 2003; Robins, Hernan, & Brumback 2000; Shi, Fan, Song, & Lu 2018a), value search methods (Nie, Brunskill, & Wager 2020; Wang, Zhou, Song, & Sherwood 2018; B. Zhang, Tsiatis, Laber, & Davidian 2012 2013), outcome weighted learning (Y. Zhao, Zeng, Rush, & Kosorok 2012; Y.-Q. Zhao, Zeng, Laber, & Kosorok 2015; Zhou, Mayer-Hamblett, Khan, & Kosorok 2017), targeted minimum loss-based estimator (van der Laan & Luedtke 2015), and decision list-based methods (Y. Zhang, Laber, Davidian, & Tsiatis 2018; Y. Zhang, Laber, Tsiatis, & Davidian 2015). While none of these methods could derive ODR from the experimental sample with unobserved long-term outcome of interest.

Our considered estimation of the ODR naturally falls in the framework of semi-supervised learning. A large number of semi-supervised learning methods have been proposed for the regression or classification problems (Chakrabortty, Cai, et al. 2018; Chapelle, Scholkopf, & Zien 2009; Chen, Hong, Tarozzi, et al. 2008; Zhu 2005). Recently, Athey, Chetty, Imbens, and Kang (2019) studied the estimation of the average treatment effect under the framework of combining the experimental data with the auxiliary data. They proposed to use the surrogate index and clarified the comparability and surrogacy assumptions, which allowed them to impute the missing outcomes in the experimental data based on the regression model learned from the auxiliary data using baseline covariates and intermediate outcomes. However, as far as we know, no work has been done for estimating the ODR in such a semi-supervised setting.

1.2 | Contributions

Our work contributes to the following folds. First, to the best of our knowledge, this is the first work on estimating the heterogeneous treatment effect and developing the optimal decision making for the long-term outcome that cannot be observed in an experiment, by leveraging the idea from semi-supervised learning and extending the framework of Athey et al. (2019). Methodologically, we propose an auGmented inverse propensity weighted Experimental and Auxiliary sample-based decision Rule, named GEAR. This rule maximizes the augmented inverse propensity weighted (AIPW) estimator of the value function over a class of interested decision rules using the experimental sample, with the primary outcome being imputed based on the auxiliary sample. Theoretically, we show that the AIPW estimator under the proposed GEAR is consistent and derive its corresponding asymptotic distribution under certain conditions. A confidence interval (CI) for the estimated value is provided.

The rest of this paper is organized as follows. We introduce the statistical framework for estimating the optimal treatment decision rule using the experimental sample and the auxiliary sample, and associated assumptions in Section 2. In Section 3, we propose our GEAR method and establish consistency and asymptotic distributions of the estimated value functions under the proposed GEAR. Extensive simulations and sensitivity studies are conducted to demonstrate the empirical validity of the proposed method in Section 4, followed by an application to ACTG 175 data in Section 5. We conclude our paper with a discussion in Section 6. The technical proofs are given in the supplementary article.

2 | STATISTICAL FRAMEWORK

2.1 | Experimental Sample and Auxiliary Sample

Suppose there is an experimental sample of interest E. Let X_E denote r-dimensional individual's baseline covariates with the support $\mathbb{X}_E \in \mathbb{R}^r$, and $A_E \in \{0, 1\}$ denote the treatment an individual receives. The long-term outcome of interest Y_E with support $\mathbb{Y}_E \in \mathbb{R}$ cannot be observed, instead we only obtain the s-dimensional intermediate outcomes M_E with support $\mathbb{M}_E \in \mathbb{R}^s$ after a treatment A_E is assigned. Denote N_E as the sample size for the experimental sample, which consists of $\{E_i = (X_{E,i}, A_{E,i}, M_{E,i}), i = 1, \dots, N_E\}$ independent and identically distributed (I.I.D.) across i.

To recover the missing long-term outcome of interest in the experimental sample, we include an auxiliary sample, U, which contains the individual's baseline covariates X_U, intermediate outcomes M_U, and the observed long-term outcome of interest Y_U, with support $\mathbb{X}_U, \mathbb{M}_U, \mathbb{Y}_U$ respectively. However, treatment information is not available in the auxiliary sample. Let N_U denote the sample size for the I.I.D. auxiliary sample that includes $\{U_i = (X_{U,i}, M_{U,i}, Y_{U,i}), i = 1, \dots, N_U\}$.

We use $R = \{E, U\}$ to indicate the missingness and identification of each sample, where $R = E$ implies the experimental sample with missing long-term primary outcome and $R = U$ means the auxiliary sample with missing treatment information. Thus, these two samples can also be rewritten as one joint sample $\{(X_i, R_i, A_i\mathbb{I}_{R_i=E}, M_i, Y_i\mathbb{I}_{R_i=U}), i = 1, \dots, N_E + N_U\}$, where $\mathbb{I}(\cdot)$ is an indicator function.

2.2 | Assumptions

In this subsection, we make five key assumptions in order to introduce the ODR. For the experimental sample, define the potential outcomes $Y_E^*(0)$ and $Y_E^*(1)$ as the long-term outcome that would be observed after an individual receiving treatment 0 or 1, respectively. Let the propensity score as the conditional probability of receiving treatment 1 in the experimental sample, i.e. $\pi(x) = \Pr_E(A_{E,i} = 1 | X_{E,i} = x)$. As standard in causal inference by Rubin (1978), we assume:

(A1). Stable Unit Treatment Value Assumption (SUTVA): $Y_E = A_E Y_E^*(1) + (1 - A_E) Y_E^*(0)$.
(A2). No Unmeasured Confounders Assumption: $\{Y_E^*(0), Y_E^*(1)\} \perp\!\!\!\perp A_E \mid X_E$.
(A3). $0 < \pi(x) < 1$ for all $x \in \mathbb{X}_E$.

To impute the missing long-term outcome in the experimental sample with the assistance of the auxiliary sample, we introduce the following two assumptions, the comparability assumption and the surrogacy assumption.

First, the comparability assumption states that the population distribution of the long-term outcome of interest Y is independent of whether belonging to the experimental sample or the auxiliary sample, given the information of population baseline covariates X and population intermediate outcomes M as follows.

(A4). Comparability Assumption: $Y \perp\!\!\!\perp R \mid X, M$.

Here, (A4) is also known as 'conditional independence assumption' made in Chen et al. (2008), and has an equivalent expression as $Y_E \mid \{M_E, X_E\} \sim Y_U \mid \{M_U, X_U\}$ proposed in Athey et al. (2019). When (A4) holds, we have a direct conclusion of the equality of the conditional mean outcome given baseline covariates and intermediate outcomes in each sample, stated in the following corollary.

Corollary 2.1. (Equal Conditional Mean) Under (A4),

$$\mathbb{E}[Y_E | M_E = m, X_E = x] = \mathbb{E}[Y_U | M_U = m, X_U = x]. \tag{1}$$

Remark 1. It is shown in Section 3 that (A4) can be relaxed to Equation (1) for deriving the proposed method.

We further define the missing at random (MAR) assumption in the joint sample as: $\{Y, A\} \perp\!\!\!\perp R \mid X, M$; and give the following corollary to show the relationship between (A4) and the MAR assumption.

Corollary 2.2. (MAR Assumption)

$$\{Y, A\} \perp\!\!\!\perp R \mid X, M \longrightarrow Y \perp\!\!\!\perp R \mid X, M.$$

Remark 2. Corollary 2.2 is a direct result of joint independence implying marginal independence. Though (A4) is untestable due to the missing long-term outcome in the experimental sample, one can believe (A4) holds if there exists strong evidence about the reasonability of the MAR assumption in the joint sample.

Second, the surrogacy assumption states that the long-term outcome of interest in the experimental sample is independent of the treatment conditional on a set of baseline covariates and intermediate outcomes as below.

(A5). Surrogacy Assumption: $Y_E \perp\!\!\!\perp A_E \mid X_E, M_E$.

Remark 3. The above assumption is first proposed in Athey et al. (2019). The validation of the surrogacy assumption relies on the 'richness' of intermediate outcomes that are highly related to the long-term outcome of interest. Similarly, it is infeasible to check the surrogacy assumption due to the missing long-term outcome in the experimental sample.

We illustrate the statistical framework of the joint sample under above assumptions by a direct acyclic graph in Figure 1. Graphically, A and Y have no common parents except for X, encoding (A2); R and Y have two common parents, X and M, encoding (A4); when fixing X and M, A and Y are independent, encoding (A5).

2.3 | Value Function and Optimal Decision Rule

A decision rule is a deterministic function $\mathsf{d}(\cdot)$ that maps $\mathbb{X}_\mathsf{E}$ to $\{0,1\}$. Define the potential outcome of interest under $\mathsf{d}(\cdot)$ as $\mathsf{Y}^*_\mathsf{E}(\mathsf{d}) = \mathsf{Y}^*_\mathsf{E}(0)\{1 - \mathsf{d}(\mathsf{X}_\mathsf{E})\} + \mathsf{Y}^*_\mathsf{E}(1)\mathsf{d}(\mathsf{X}_\mathsf{E})$, which would be observed if a randomly chosen individual from the experimental sample had received a treatment according to $\mathsf{d}(\cdot)$, where we suppress the dependence of $\mathsf{Y}^*_\mathsf{E}(\mathsf{d})$ on X_E. We then define the value function under $\mathsf{d}(\cdot)$ as the expectation of the potential outcome of interest over the experimental sample as

$$\mathsf{V}(\mathsf{d}) = \mathbb{E}\{\mathsf{Y}^*_\mathsf{E}(\mathsf{d})\} = \mathbb{E}[\mathsf{Y}^*_\mathsf{E}(0)\{1 - \mathsf{d}(\mathsf{X}_\mathsf{E})\} + \mathsf{Y}^*_\mathsf{E}(1)\mathsf{d}(\mathsf{X}_\mathsf{E})].$$

As a result, we have the optimal treatment decision rule (ODR) of interest defined to maximize the value function over the experimental sample among a class of decision rules of interest as $\mathsf{d}^{\mathsf{opt}}(\cdot) = \arg\min_{\mathsf{d}(\cdot)} \mathsf{V}(\mathsf{d})$. Suppose the decision rule $\mathsf{d}(\cdot)$ relies on a model parameter β, denoted as $\mathsf{d}(\cdot) \equiv \mathsf{d}(\cdot;\beta)$. We use a shorthand to write $\mathsf{V}(\mathsf{d})$ as $\mathsf{V}(\beta)$, and define $\beta_0 = \arg\min_\beta \mathsf{V}(\beta)$. Thus, the value function under the true ODR $\mathsf{d}(\cdot;\beta_0)$ is defined as $\mathsf{V}(\beta_0)$.

3 | PROPOSED METHOD

In this section, we detail the proposed method by constructing the AIPW value estimator for the long-term outcome based on two samples. Implementation details are provided to find the ODR. The consistency and asymptotical distribution of the value estimator under our proposed GEAR are presented, followed by its confidence interval. We also provide the inverse propensity-score weighted value estimator and its related theories in Appendix A. All the proofs are provided in the supplementary article.

3.1 | AIPW Estimator for Long-Term Outcome

To overcome the difficulty of estimating the value function due to the missing long-term outcome of interest in the experimental sample, one intuitive way is to impute the missing outcome Y_E with its conditional mean outcome given baseline covariates and intermediate outcomes (total common information available in both samples).

Denote $\mu_\mathsf{E}(\mathsf{m},\mathsf{x}) \equiv \mathbb{E}[\mathsf{Y}_\mathsf{E}|\mathsf{M}_\mathsf{E} = \mathsf{m}, \mathsf{X}_\mathsf{E} = \mathsf{x}]$, and $\mu_\mathsf{U}(\mathsf{m},\mathsf{x}) \equiv \mathbb{E}[\mathsf{Y}_\mathsf{U}|\mathsf{M}_\mathsf{U} = \mathsf{m}, \mathsf{X}_\mathsf{U} = \mathsf{x}]$. Under Corollary 2.1, we have $\mu_\mathsf{E}(\mathsf{m},\mathsf{x}) = \mu_\mathsf{U}(\mathsf{m},\mathsf{x})$. Here, $\mu_\mathsf{E}(\mathsf{m},\mathsf{x})$ is inestimable because of the missing long-term outcome. We instead use $\mu_\mathsf{U}(\mathsf{M}_\mathsf{E},\mathsf{X}_\mathsf{E})$ to impute the missing Y_E and give the following lemma as a middle step to construct the AIPW value estimator for the long-term outcome.

Lemma 1. Under (A1)-(A5), given $\mathsf{d}(\cdot;\beta)$, we have

$$\mathsf{V}(\beta) = \mathbb{E}\left[\frac{\mathbb{I}\{\mathsf{A}_\mathsf{E} = \mathsf{d}(\mathsf{X}_\mathsf{E};\beta)\}\mu_\mathsf{U}(\mathsf{M}_\mathsf{E},\mathsf{X}_\mathsf{E})}{\mathsf{A}_\mathsf{E}\pi(\mathsf{X}_\mathsf{E}) + (1-\mathsf{A}_\mathsf{E})\{1-\pi(\mathsf{X}_\mathsf{E})\}}\right].$$

Next, we propose the AIPW estimator of the value function for the long-term outcome in the experimental sample. To address the difficulty of forming the augmented term when the long-term outcome of interest cannot be observed, we show that augmenting on the missing long-term outcome is equivalent to augmenting on the imputed conditional mean outcome of interest $\mu_\mathsf{U}(\mathsf{M}_\mathsf{E},\mathsf{X}_\mathsf{E})$, by the following lemma.

Lemma 2. Under (A1)-(A5), given $\mathsf{d}(\cdot;\beta)$, we have

$$\mathbb{E}_{\mathsf{Y}_\mathsf{E}|\mathsf{X}_\mathsf{E}}\{\mathsf{Y}_\mathsf{E}|\mathsf{A}_\mathsf{E} = \mathsf{d}(\mathsf{X}_\mathsf{E};\beta),\mathsf{X}_\mathsf{E}\} = \mathbb{E}_{\mathsf{M}_\mathsf{E}|\mathsf{X}_\mathsf{E}}\{\mu_\mathsf{U}(\mathsf{M}_\mathsf{E},\mathsf{X}_\mathsf{E})|\mathsf{A}_\mathsf{E} = \mathsf{d}(\mathsf{X}_\mathsf{E};\beta),\mathsf{X}_\mathsf{E}\},$$

where $\mathsf{E}_{\mathsf{A}|\mathsf{B}}$ means taking expectation with respect to the conditional distribution of A given B.

According to Lemma 1 and Lemma 2, given a decision rule $\mathsf{d}(\cdot;\beta)$, the value function $\mathsf{V}(\beta)$ can be consistently estimated through

$$\mathsf{V}^\star_{\mathsf{n},\mathsf{AIP}}(\beta) = \frac{1}{\mathsf{N}_\mathsf{E}}\sum_{\mathsf{i}=1}^{\mathsf{N}_\mathsf{E}}\left[\nu_\mathsf{i} + \frac{\mathbb{I}\{\mathsf{A}_{\mathsf{E},\mathsf{i}} = \mathsf{d}(\mathsf{X}_{\mathsf{E},\mathsf{i}};\beta)\}\{\mu_\mathsf{U}(\mathsf{M}_{\mathsf{E},\mathsf{i}},\mathsf{X}_{\mathsf{E},\mathsf{i}}) - \nu_\mathsf{i}\}}{\mathsf{A}_{\mathsf{E},\mathsf{i}}\pi(\mathsf{X}_{\mathsf{E},\mathsf{i}}) + (1-\mathsf{A}_{\mathsf{E},\mathsf{i}})\{1-\pi(\mathsf{X}_{\mathsf{E},\mathsf{i}})\}}\right],$$

where $\nu_i \equiv \mathbb{E}\{\mu_U(M_{E,i}, X_{E,i})|A_{E,i} = d(X_{E,i};\beta), X_{E,i}\}$ presents the augmented term. Here, the propensity score π can be estimated in the experimental sample, denoted as $\widehat{\pi}$, and the conditional mean μ_U can be estimated in the auxiliary sample, denoted as $\widehat{\mu}_U$. Then, by replacing the implicit functions in $V^{\star}_{n,AIP}(\beta)$, it is straightforward to give the AIPW estimator of the value function $V(\beta)$ as

$$\widehat{V}_{AIP}(\beta) = \frac{1}{N_E}\sum_{i=1}^{N_E}\left[\widehat{\nu}_i + \frac{\mathbb{I}\{A_{E,i} = d(X_{E,i};\beta)\}\{\widehat{\mu}_U(M_{E,i}, X_{E,i}) - \widehat{\nu}_i\}}{A_{E,i}\widehat{\pi}(X_{E,i}) + (1 - A_{E,i})\{1 - \widehat{\pi}(X_{E,i})\}}\right],$$

where $\widehat{\nu}_i \equiv \widehat{\mathbb{E}}\{\widehat{\mu}_U(M_{E,i}, X_{E,i})|A_{E,i} = d(X_{E,i};\beta), X_{E,i}\}$ is the estimator for ν_i. We define $\widehat{\beta}^G = \arg\max_\beta \widehat{V}_{AIP}(\beta)$, and then propose the GEAR as $d(X;\widehat{\beta}^G)$ with the corresponding estimated value function as $\widehat{V}_{AIP}(\widehat{\beta}^G)$.

3.2 | Implementation Details

3.2.1 | Class of Decision Rules

The GEAR can be searched within a pre-specified class of decision rules. Popular classes include generalized linear rules, fixed depth decision trees, threshold rules, and so on (Athey & Wager 2017; Rai 2018; B. Zhang et al. 2012). In this paper, we focus on the class of generalized linear rules. Specifically, suppose the decision rule takes a form as $d(X_E;\beta) \equiv \mathbb{I}\{g(X_E)^\top\beta > 0\}$, where $g(\cdot)$ is an unknown function. We use $\phi_X(\cdot)$ to denote a set of basis functions of $\mathbb{X}_E$ with length v, which are "rich" enough to approximate the underlying function $g(\cdot)$. Thus, the GEAR is found within a class of $\mathbb{I}\{\phi_X(X_E)^\top\beta > 0\}$. For notational simplicity, we include 1 in $\phi_X(\cdot)$ so that $\beta \in \mathbb{R}^{v+1}$. With subject to $||\beta||_2 = 1$ for identifiability purpose, the maximizer for $\widehat{V}_{AIP}(\beta)$ can be solved using any global optimization algorithm. In our implementation, we apply the heuristic algorithm to search for the GEAR. The architecture of the proposed GEAR is illustrated in Figure 2.

3.2.2 | Estimation Models

The conditional mean of the long-term outcome $\mu_U(m,x)$ can be estimated through any parametric or nonparametric model. In practice, we assume $\mu_U(m,x)$ can be determined by a flexible basis function of baseline covariates and intermediate outcomes, to fully capture the underlying true model. Similarly, one can use a flexible basis function of baseline covariates and the treatment to model the augmented term as well as the propensity score function. Note that any machine learning tools such as Random Forest or Deep Learning can be applied to model terms in the proposed AIPW estimator. Our theoretical results still hold under these nonparametric models as long as the regressors have desired convergence rates (see results established in Farrell, Liang, and Misra (2018); Wager and Athey (2018)).

3.2.3 | Estimation of the Augmented Term

To estimate the augmented term ν_i, we need three steps as follows.

Step 1. First, we model $\mu_U(m,x)$ through the auxiliary sample $\{X_U, M_U, Y_U\}$ as $\widehat{\mu}_U(m,x)$.

Step 2. Second, we plug $\{M_E, X_E\}$ of the experimental sample into $\widehat{\mu}_U(m,x)$ and get $\widehat{\mu}_U(M_E, X_E)$ as the conditional mean outcome of interest to impute the missing Y_E.

Step 3. At last, we fit $\widehat{\mu}_U(M_E, X_E)$ on $\{A_E, X_E\}$ in the experimental sample, and get $\widehat{\nu}_i$.

3.3 | Theoretical Properties

We next show the consistency and asymptotic normality of our proposed AIPW estimator. Its asymptotic variance can be decomposed into two parts, corresponding to the estimation variances from two independent samples. As mentioned in Section 3.2, our AIPW estimator can handle various machine learning or parametric estimators as long as regressors have desired convergence rates. To derive an explicit variance form, we next focus on parametric models.

We posit parametric models for $\pi(x) \equiv \pi(x;\gamma)$ and $\mu_U(m,x) \equiv \mu_U(m,x;\lambda)$ with true model parameters γ and λ. Let $\phi_X(X)$ and $\phi_M(M)$ to represent appropriate basis functions for X and M, respectively. Without loss of generality, we posit basis model for the augmented term such that $\mathbb{E}\{\mu_U(m,x;\lambda)|A = 0, X = x\} \equiv \phi_X(x)^\top\theta_0$, and $\mathbb{E}\{\mu_U(m,x;\lambda)|A = 1, X = x\} \equiv \phi_X(x)^\top\theta_1$ with true model parameters θ_0 and θ_1.

3.4 | Technical Conditions

The following conditions are needed to derive our theoretical results:
(A6). Suppose the density of covariates $f_X(x)$ is bounded away from 0 and ∞ and is twice continuously differentiable with bounded derivatives.

(A7). Both $\pi(x;\gamma)$ and $\mu_U(m,x;\lambda)$ are smooth bounded functions, with their first derivatives exist and bounded.
(A8). Model for $\mu_U(m,x;\lambda)$ is correctly specified.
(A9). Denote $t = \sqrt{N_E/N_U}$ and assume $0 < t < +\infty$.
(A10). The true value function $V(\beta)$ is twice continuously differentiable at a neighborhood of β_0.
(A11). Either the model of the propensity score or the model of the augmented term is correctly specified.

Here, (A6) and (A10) are commonly imposed to establish the inference for value search methods (Wang et al. 2018; B. Zhang et al. 2012). (A7) is assumed for desired convergence rates of $\widehat{\pi}$ and $\widehat{\mu}_U$. To apply machine learning tools, similar assumption is required (see more details in Farrell et al. (2018); Wager and Athey (2018)). From (A8), we can replace the missing long-term outcome with its imputation, and thus the consistency holds. Evaluations are provided in Section 4.2 to examine the proposed method when (A8) is violated. (A9) states that the sizes of two samples are comparable, which prevents the asymptotic variance from blowing up when combining two samples in semi-supervised learning (Chakrabortty et al. 2018; Chen et al. 2008). (A11) is included to establish the doubly robustness of the value estimator, which is commonly used in the literature of doubly robust estimator (Dudík, Langford, & Li 2011; B. Zhang et al. 2012 2013).

3.5 | Theoretical Results

The following theorem gives the consistency of our AIPW estimator of the value function to the true value function.

Theorem 3.1. (Consistency) Under (A1)-(A9) and (A11),

$$\widehat{V}_{AIP}(\beta) = V(\beta) + o_p(1), \quad \forall \beta.$$

Remark 4. When the model for $\mu_U(m,x)$ is correctly specified, our AIPW estimator is doubly robust given either the model of the propensity score or the model of the augmented term is correct. To prove the theorem, we establish the theoretical results with their proofs for the inverse propensity-score weighted estimator as a middle step. See more details in Appendix A.

To establish the asymptotic normality of $\widehat{V}_{AIP}(\widehat{\beta}^G)$, we first show the estimator $\widehat{\beta}^G$ has a cubic rate towards the true β_0.

Lemma 3. Under (A1)-(A11), we have

$$N_E^{1/3}||\widehat{\beta}^G - \beta_0||_2 = O_p(1), \tag{2}$$

where $||\cdot||_2$ is the L_2 norm, and $O_p(1)$ means the random variable is stochastically bounded.

Based on Lemma 3, we next give the asymptotic normality of $\sqrt{N_E}\{\widehat{V}_{AIP}(\widehat{\beta}^G) - V(\beta_0)\}$ in the following theorem.

Theorem 3.2. (Asymptotic Distribution) Under (A1)-(A11),

$$\sqrt{N_E}\{\widehat{V}_{AIP}(\widehat{\beta}^G) - V(\beta_0)\} \xrightarrow{\mathcal{D}} N(0, \sigma^2_{AIP}), \tag{3}$$

where $\sigma^2_{AIP} = t\sigma^2_U + \sigma^2_E$, $\sigma^2_U = \mathbb{E}[\{\xi_i^{(U)}\}^2]$, and $\sigma^2_E = \mathbb{E}[\{\xi_i^{(E)}\}^2]$. Here, $\xi_i^{(E)}$ and $\xi_i^{(U)}$ are the I.I.D. terms in the experimental sample and auxiliary sample, respectively.

Remark 5. From Theorem 3.2, the asymptotic variance of the AIPW estimator has an additive form that consists of the estimation error from each sample. Proportion of these two estimation variances is controlled by the sample ratio. In reality, N_U is usually larger than N_E. When $N_U/N_E \to \infty$, we have $t \to 0$, and thus the estimation error from auxiliary sample can be ignored. Our result under this special case is supported by Chakrabortty et al. (2018) where they considered $N_U/N_E \to \infty$ for a regression problem.

3.6 | Variance Estimation and Confidence Interval

Next, we give explicit form of $\xi_i^{(E)}$ and $\xi_i^{(U)}$ from the proof of Theorem 3.2 to estimate σ_{AIP}. Denote $\dot{\pi}(x;\gamma) \equiv \partial\pi(x;\gamma)/\partial\gamma$ and $\dot{\mu}_U(m,x;\lambda) \equiv \partial\mu_U(m,x;\lambda)/\partial\lambda$. Let

$$\begin{aligned} H_1 &\equiv \lim_{N_E \to +\infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \phi_X(X_{E,i}) \dot{\pi}(X_{E,i};\gamma)^\top, H_2 \equiv \lim_{N_U \to +\infty} \frac{1}{N_U} \sum_{i=1}^{N_U} \begin{bmatrix} \phi_X(X_{U,i}) \\ \phi_M(M_{U,i}) \end{bmatrix} \dot{\mu}_U(M_{U,i}, X_{U,i};\lambda)^\top, \\ H_3 &\equiv \lim_{N_E \to +\infty} \frac{1}{N_E} \sum_{i=1}^{N_E} (1 - A_{E,i}) \phi_X(X_{E,i}) \phi_X(X_{E,i})^\top, H_4 \equiv \lim_{N_E \to +\infty} \frac{1}{N_E} \sum_{i=1}^{N_E} A_{E,i} \phi_X(X_{E,i}) \phi_X(X_{E,i})^\top, \end{aligned}$$

$$
\begin{aligned}
G_1 &\equiv \lim_{N_E\to+\infty}\frac{1}{N_E}\sum_{i=1}^{N_E}\frac{r_i(1-2A_{E,i})\dot{\pi}(X_{E,i};\gamma)\mu_U(M_{E,i},X_{E,i};\lambda)}{s_i^2}, G_2 \equiv \lim_{N_E\to+\infty}\frac{1}{N_E}\sum_{i=1}^{N_E}\frac{r_i}{s_i}\dot{\mu}_U(M_{E,i},X_{E,i};\lambda),\\
G_3 &\equiv \lim_{N_E\to+\infty}\frac{1}{N_E}\sum_{i=1}^{N_E}-\frac{r_iq_i(1-2A_{E,i})\dot{\pi}(X_{E,i};\gamma)}{s_i^2}, G_4 \equiv \lim_{N_E\to+\infty}\frac{1}{N_E}\sum_{i=1}^{N_E}\left[1-\frac{r_i}{s_i}\right]\phi_X(X_{E,i})\{1-d(X_{E,i};\beta_0)\},\\
G_5 &\equiv \lim_{N_E\to+\infty}\frac{1}{N_E}\sum_{i=1}^{N_E}\left[1-\frac{r_i}{s_i}\right]\phi_X(X_{E,i})d(X_{E,i};\beta_0),
\end{aligned}
$$

where $r_i \equiv \mathbb{I}\{A_{E,i}=d(X_{E,i};\beta_0)\}, s_i \equiv A_{E,i}\pi(X_{E,i};\gamma)+(1-A_{E,i})\{1-\pi(X_{E,i};\gamma)\}$, and $q_i \equiv \phi_X(X_{E,i})^\top\theta_0+\phi_X(X_{E,i})^\top(\theta_1-\theta_0)d(X_{E,i};\beta_0)$.

Then, the I.I.D. term in the experimental sample is

$$
\begin{aligned}
\xi_i^{(E)} \equiv &\frac{r_i\{\mu_U(M_{E,i},X_{E,i};\lambda)-\nu_i^*\}}{s_i}+\nu_i^*-V(\beta_0)+(G_1^\top+G_3^\top)H_1^{-1}\phi_X(X_{E,i})\{A_{E,i}-\pi(X_{E,i};\gamma)\}\\
&+G_5^\top H_4^{-1}\phi_X(X_{E,i})A_{E,i}\{\mu_U(M_{E,i},X_{E,i};\lambda)-\phi_X(X_{E,i})^\top\theta_1\}+G_4^\top H_3^{-1}\phi_X(X_{E,i})(1-A_{E,i})\{\mu_U(M_{E,i},X_{E,i};\lambda)-\phi_X(X_{E,i})^\top\theta_0\},
\end{aligned}
$$

for $\nu_i^* \equiv E\{\mu_U(M_{E,i},X_{E,i};\lambda)|A_{E,i}=d(X_{E,i};\beta_0),X_{E,i}\}$. And the I.I.D. term in the auxiliary sample corresponds to

$$
\xi_i^{(U)} \equiv G_2^\top H_2^{-1}\begin{bmatrix}\phi_X(X_{U,i})\\ \phi_M(M_{U,i})\end{bmatrix}\{Y_{U,i}-\mu_U(M_{U,i},X_{U,i};\lambda)\}.
$$

By plugging the estimations into the pre-specified models, we could obtain the estimated $\widehat{\xi}_i^{(E)}$ and $\widehat{\xi}_i^{(U)}$. Then the variance σ_E^2 and σ_U^2 can be consistently estimated by $\widehat{\sigma}_E^2 = N_E^{-1}\sum_{i=1}^{N_E}\{\widehat{\xi}_i^{(E)}\}^2$ and $\widehat{\sigma}_U^2 = N_U^{-1}\sum_{i=1}^{N_U}\{\widehat{\xi}_i^{(U)}\}^2$, respectively. Thus, we can estimate σ_{AIP} through

$$
\widehat{\sigma_{AIP}} \equiv \sqrt{t\widehat{\sigma}_U^2+\widehat{\sigma}_E^2}, \tag{4}
$$

based on Theorem 3.2. Therefore, a two-sided $1-\alpha$ confidence interval (CI) for $V(\beta_0)$ under the GEAR is

$$
\left[\widehat{V}_{AIP}(\widehat{\beta}^G)-\frac{z_{\alpha/2}\widehat{\sigma_{AIP}}}{\sqrt{N_E}},\ \widehat{V}_{AIP}(\widehat{\beta}^G)+\frac{z_{\alpha/2}\widehat{\sigma_{AIP}}}{\sqrt{N_E}}\right], \tag{5}
$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2-$th quantile of a standard normal distribution.

4 | SIMULATION STUDIES

In this section, we evaluate the proposed method when the model of the conditional mean of the long-term outcome is correctly specified and misspecified in Section 4.1 and Section 4.2, respectively. Additional sensitivity studies of the assumption violation are provided in Section 4.3.

4.1 | Evaluation under Correctly Specified Model

Simulated data, including baseline covariates $X=[X^{(1)},X^{(2)},\cdots,X^{(r)}]^\top$, the treatment A, intermediate outcomes $M=[M^{(1)},M^{(2)},\cdots,M^{(s)}]^\top$, and the long-term outcome Y, are generated from the following model:

$$
\begin{aligned}
X^{(1)},X^{(2)},\cdots,X^{(r)} &\overset{iid}{\sim} \text{Uniform}[-1,1], \quad A \overset{iid}{\sim} \text{Bernoulli}(0.5),\\
M = H^M(X)+AC^M(X)+\epsilon^M&, \ \ Y = H^Y(X)+C^Y(X,M)+\epsilon^Y,
\end{aligned}
$$

where ϵ^M and ϵ^Y are random errors following $N(0,0.5)$. Here, A in the auxiliary sample is used only for generating intermediate outcomes such that the comparability assumption is satisfied. Note that Y is generated for the auxiliary sample only. Given X and M, we can see Y is independent of A, which indicates the surrogacy assumption.

Set $r=4$ and $s=2$. We consider following two scenarios with different $H^M(\cdot)$, $C^M(\cdot)$, $H^Y(\cdot)$, and $C^Y(\cdot)$.

$$
\textbf{S1}:\begin{cases}H^M(X)=\begin{bmatrix}X^{(3)}\\X^{(1)}\end{bmatrix}, C^M(X)=\begin{bmatrix}4\{X^{(1)}-X^{(2)}\}\\4\{X^{(4)}-X^{(3)}\}\end{bmatrix},\\ H^Y(X)=-1+X^{(2)}+X^{(4)}, C^Y(X,M)=M^{(1)}+M^{(2)}.\end{cases}
\quad
\textbf{S2}:\begin{cases}H^M(X)=\begin{bmatrix}\{X^{(1)}\}^2X^{(3)}+\sin\{X^{(4)}\}\\ \{X^{(1)}\}^3-\{X^{(2)}-X^{(4)}\}^2\end{bmatrix},\\ C^M(X)=\begin{bmatrix}4\{X^{(1)}-X^{(2)}\}\\4\{X^{(4)}-X^{(3)}\}\end{bmatrix},\\ H^Y(X)=-1+X^{(2)}+X^{(4)}, C^Y(X,M)=M^{(1)}+M^{(2)}.\end{cases}
$$

Under Scenario 1 and 2, we have the parameter of the true ODR as $\beta_0 = [0,0.5,-0.5,-0.5,0.5]^\top$ with subject to $||\beta_0||_2 = 1$, which can be easily solved based on the function $C^M(\cdot)$ that describes the treatment-covariates interaction. The true value $V(\beta_0)$ can be calculated by Monte Carlo approximations, as listed in Table 1. We consider $N_U = 400$ for the auxiliary sample and allow N_E chosen from the set $\{200,400,800\}$ in the experimental sample.

To apply the GEAR, we model the conditional mean outcome $\mu_U(m, x)$ and the augmented term v_i in the auxiliary data via a linear regression. Here, the model of $\mu_U(m, x)$ is correctly specified by noting that Y is linear in $\{X, M\}$ under Scenario 1 and 2. The GEAR is searched within a class of $d(X_E; \beta) = \mathbb{I}(X_E^\top \beta > 0)$ subjecting to $||\beta||_2 = 1$, through Genetic Algorithm provided in R package `rgenound`, where we set 'optim.method' = 'Nelder-Mead', 'pop.size' = 3000, 'domain'=[-10,10], and 'starting.values' as a zero vector. Results are summarized in Table 1, including the estimated value under the estimated rule $\widehat{V}_{AIP}(\widehat{\beta}^G)$ and its standard error $SE\{\widehat{V}_{AIP}\}$, the estimated standard deviation $\mathbb{E}\{\widehat{\sigma}_{AIP}\}$ by Equation (4), the value under the estimated rule $V(\widehat{\beta}^G)$ by plugging the GEAR into the true model, the empirical coverage probabilities (CP) for 95% CI constructed by Equation (5), the rate of the correct decision (RCD) made by the GEAR, and the L_2 loss of $\widehat{\beta}^G$ ($||\widehat{\beta}^G - \beta_0||_2$), aggregated over 500 simulations.

From Table 1, it is clear that both the estimated GEAR and its estimated value approach to the true as the sample size N_E increases in all scenarios. Specifically, our proposed GEAR method achieves $V(\widehat{\beta}^G) = 0.86$ in Scenario 1 ($V(\beta_0) = 0.87$) and $V(\widehat{\beta}^G) = 0.19$ in Scenario 2 ($V(\beta_0) = 0.20$) when $N_E = 800$. Notice that the ℓ_2 loss of $\widehat{\beta}^G$ decays at a rate that is approximately proportional to $N_E^{-1/3}$, which verifies our theoretical findings in Lemma 3. Moreover, the average rate of the correct decision made by the GEAR increases with N_E increasing. In addition, there are two findings that help to verify Theorem 3.2. First, the estimated standard deviation of value function is close to the standard error of the estimated value function, and gets smaller as the sample size N_E increases. Second, the empirical coverage probabilities of the proposed 95% CI approach to the nominal level in all settings. Note that there is no strictly increasing trend of the empirical coverage probabilities due to the fixed sample size $N_U = 400$.

4.2 | Evaluation under Model Misspecification

We consider more general settings to examine the proposed method when the model of $\mu_U(m, x)$ is misspecified. The data is generated from the same model in Section 4.1.We fix

$$\left\{ H^M(X) = \begin{bmatrix} X^{(3)} \\ X^{(1)} \end{bmatrix}, C^M(X) = \begin{bmatrix} 4\{X^{(1)} - X^{(2)}\} \\ 4\{X^{(4)} - X^{(3)}\} \end{bmatrix}, \right.$$

and set following three scenarios with different $H^Y(\cdot)$ and $C^Y(\cdot)$.

$$\textbf{S3}: \left\{ \begin{aligned} H^Y(X) &= \{X^{(1)} + X^{(3)}\}\{X^{(1)}\}^2 \\ &\quad + \sin\{X^{(4)}\} - \{X^{(2)} - X^{(4)}\}^2, \\ C^Y(X, M) &= M^{(1)} + M^{(2)}. \end{aligned} \right. \quad \textbf{S4}: \left\{ \begin{aligned} H^Y(X) &= \{X^{(1)}\}^3 + \{X^{(2)}\}^2 + X^{(3)}, \\ C^Y(X, M) &= M^{(1)} + X^{(4)} M^{(2)}. \end{aligned} \right. \quad \textbf{S5}: \left\{ \begin{aligned} H^Y(X) &= X^{(2)} - \{X^{(4)}\}^2, \\ C^Y(X, M) &= 0.25\{M^{(1)} - X^{(3)}\}^2 + M^{(2)}. \end{aligned} \right.$$

Under Scenario 3, we have the true ODR is still linear while the true ODRs for Scenario 4 and 5 are non-linear due to their $C^Y(\cdot)$ involving covariates-surrogacy interaction. Table 2 lists the true value $V(\beta_0)$ for each scenario.

We apply the proposed GEAR with the tensor-product B-splines for Scenario 3-5, respectively. Specifically, we first model $\mu_U(m, x)$ with the tensor-product B-splines of $\{X_U, M_U\}$ in the auxiliary sample. The degree and knots for the B-splines are selected based on five-fold cross validation to minimize the least square error of the linear regression. Then, we search the GEAR within the class of $\mathbb{I}\{\phi_X(X_E)^\top \beta > 0\}$, where $\phi_X(\cdot)$ is the polynomial basis with degree=2. Here, the augmented term is fitted by a linear regression of $\widehat{\mu}_U(M_E, X_E)$ on $\{A_E, \phi_X(X_E)\}$. We name the above procedure as 'GEAR-Bspline'. For comparison, we also apply the linear procedure described in Section 4.1 as 'GEAR-linear' without taking any basis. One may note both procedures model $\mu_U(m, x)$ incorrectly. Reported in Table 2 are the empirical results under GEAR-Bspline and GEAR-linear aggregated over 500 simulations.

It can be seen from Table 2 that the GEAR-Bspline procedure performs reasonably better than the linear procedure under non-linear decision rules. Specifically, in Scenario 3 with only the baseline function $H^Y(\cdot)$ non-linear in X, GEAR-linear performs comparable to GEAR-Bspline, as the linear model can well approximate the non-linear baseline function. In Scenario 4 and 5 with more complex non-linear function $C^Y(\cdot)$, GEAR-Bspline outperforms GEAR-linear in terms of smaller bias and higher empirical coverage probabilities of the 95% CI. For example, GEAR-Bspline achieves $V(\widehat{\beta}^G) = 2.43$ in Scenario 4 ($V(\beta_0) = 2.59$) with coverage probability 92.0% and $V(\widehat{\beta}^G) = 2.77$ in Scenario 5 ($V(\beta_0) = 3.03$) with coverage probability 92.4% when $N_E = N_U$, while GEAR-linear can hardly maintain an empirical coverage probability over one third in Scenario 5 due to the severe model misspecification. Note that because of the interaction between X and M in $C^Y(\cdot)$, the model assumption is still mildly violated even applying the GEAR-Bspline method. Thus, the empirical coverage probabilities of the 95% CI decreases as the sample size N_E increases.

4.3 | Sensitivity Studies

In this section, we investigate the finite sample performance of the proposed GEAR when the surrogacy assumption is violated in different extent, i.e. part of the information related to the long-term outcome cannot be collected or captured through intermediate outcomes. We consider the following Scenario 6 with $r = 2$ and $s = 2$.

$$\textbf{S6}: \left\{ H^M(X) = \begin{bmatrix} 0 \\ X^{(1)} \end{bmatrix}, C^M(X) = \begin{bmatrix} -0.5 + 0.4X^{(1)} - 0.6X^{(2)} \\ 0.5 + 0.6X^{(1)} - 0.4X^{(2)} \end{bmatrix}, H^Y(X) = X^{(2)}, C^Y(X, M) = M^{(1)} + M^{(2)}, \right.$$

where the true parameter of the ODR is $\beta_0 = [0, 1/\sqrt{2}, -1/\sqrt{2}]^\top$ with the true value 0.333. We use the following $M_{par}^{(1)}$ as one contaminated intermediate outcome we collected instead of the original $M^{(1)}$ as $M_{par}^{(1)} = M^{(1)} + A(1 - l)\{-0.5 + 0.4X^{(1)}\}$, where the parameter l chosen from $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$ reflects the uncollected information related to the long-term outcome. When $l = 1$, we have the information of intermediate outcomes is fully collected. However, under $l \in \{0, 0.2, 0.4, 0.6, 0.8\}$, the surrogacy assumption cannot hold anymore, since the long-term outcome is still dependent on the treatment given the information of M and X.

Following the same estimation procedure as described in Section 4.1, we summarize the simulation results over 500 replications in Table 3 for $l = \{0, 0.4, 0.8\}$. Figure 3 and Figure 4 show how the bias of $V(\widehat{\beta}^G)$ towards the true value and the average rate of the correct decision made by the GEAR change as the parameter l (that indicates the uncollected information of intermediate outcomes) changes, respectively. Based on the results, our proposed method still has a reasonable performance when the surrogacy assumption is mildly violated. Specifically, the proposed GEAR achieves $V(\widehat{\beta}^G) = 0.314$ in Scenario 6 ($V(\beta_0) = 0.333$) with an empirical coverage probability as 90.8% under $l = 0.8$ and $N_E = 800$. In addition, it is clear that including more intermediate outcomes that are highly correlated to the long-term outcome, could help to explain the treatment effect on the long-term outcome according to Figure 3 and Figure 4. Similarly, to conduct sensitivity analysis when the comparability assumption is violated, we consider the settings in Scenarios 1 and 2 but with $C^Y(X, M) = M^{(1)} + 0.8M^{(2)}$ for the auxiliary sample. The results are summarized in Table 4.

5 | REAL DATA ANALYSIS

In this section, we illustrate our proposed method by application to the AIDS Clinical Trials Group Protocol 175 (ACTG 175) data. There are 1046 HIV-infected subjects enrolled in ACTG 175, who were randomized to two competitive antiretroviral regimens in equal proportions (Hammer et al. 1996): zidovudine (ZDV) + zalcitabine (ddC), and ZDV+didanosine (ddI). Denote 'ZDV+ddC' as treatment 0, versus 'ZDV+ddI' as treatment 1. Here, since the long-term AIDS recovery is not recorded in the dataset, we use the mean CD4 count (cells/mm3) at 96 ± 5 weeks as the long-term outcome of interest (Y). A higher CD4 count usually indicates a stronger immune system. However, about one-third of the patients who received treatment 0 or 1 have a missing long-term outcome, which forms the experimental sample of interest. Due to the limited availability of AIDS electronic medical records data, in this paper we use the rest complete dataset in ACTG 175 as the auxiliary sample by ignoring its treatment information to just demonstrate our method.

To be specific, in the experimental sample ($N_E = 376$), 187 patients were assigned to treatment 0 and 189 patients to treatment 1. The propensity score function is estimated through a logistic regression in the experimental sample. The auxiliary sample consists of $N_U = 670$ subjects with observed long-term outcome. We consider $r = 12$ baseline covariates used in (Tsiatis, Davidian, Zhang, & Lu 2008): 1) four continuous variables: age (years), weight (kg), CD4 count (cells/mm3) at baseline, and CD8 count (cells/mm3) at baseline; 2) eight categorical variables: hemophilia, homosexual activity, history of intravenous drug use, Karnofsky score (scale of 0-100), race (0=white, 1=non-white), gender (0=female), antiretroviral history (0=naive, 1=experienced), and symptomatic status (0=asymptomatic). Intermediate outcomes contain CD4 count at 20 ± 5 weeks and CD8 count at 20 ± 5 weeks. It can be shown in the auxiliary data that intermediate outcomes are highly related to the long-term outcome via a linear regression of Y_U on $\{X_U, M_U\}$. We apply our proposed 'GEAR-linear' and 'GEAR-Bspline' described in Section 4.2 to the ACTG 175 data, respectively. Here, to avoid the curse of high dimensionality, we only take the polynomial basis on the continuous variables with degree as 2. Reported in Table 5 are the estimated mean outcome for each treatment as $\widehat{V}_{AIP}(0)$ and $\widehat{V}_{AIP}(1)$, the estimated value $\widehat{V}_{AIP}(\widehat{\beta}^G)$ with its estimated standard deviation $\widehat{\sigma}_{AIP}$, the 95% CI for the estimated value, and the number of assignments for each treatment.

It is clear that the proposed GEAR estimation procedure with the B-spline performs reasonably better than the linear procedure. Next, we focus on the results obtained from the GEAR-Bspline method in the experimental sample of interest. Our proposed GEAR-Bspline method achieves a value of 344.2 with a smaller standard deviation as 9.9 comparing to GEAR-linear (10.1) in the experimental sample. The GEAR with B-spline assigns 187 patients to 'ZDV+ddI' and 189 patients to 'ZDV+ddC', which is consistent with the competitive nature of these two treatments.

6 | DISCUSSION

In this paper, we proposed a new personalized optimal decision policy when the long-term outcome of interest cannot be observed. Theoretically, we gave the cubic convergence rate of our proposed GEAR, and derived the consistency and asymptotical distributions of the value function under the GEAR. Empirically, we validated our method, and examined the sensitivity of our proposed GEAR when the model is misspecified or when assumptions are violated. There are several other possible extensions we may consider in future work. First, we only consider two treatment options in this paper, while in applications it is common to have more than two options for decision making. Thus, a more general method with multiple treatments or even continuous decision marking is desirable. Second, we can extend our work to dynamic decision making, where the ultimate outcome of interest cannot be observed in the experimental sample but can be found in some auxiliary dataset.

DATA AVAILABILITY STATEMENT

Simulated data and the associated R codes for implementing the proposed GEAR method are provided in the Supporting Information.

References

Athey, S., Chetty, R., Imbens, G. W., & Kang, H. (2019). *The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely* (Tech. Rep.). National Bureau of Economic Research.

Athey, S., & Wager, S. (2017). Efficient policy learning. *arXiv preprint arXiv:1702.02896*.

Chakrabortty, A., Cai, T., et al. (2018). Efficient and adaptive linear regression in semi-supervised settings. *The Annals of Statistics*, *46*(4), 1541–1572.

Chakraborty, B., & Moodie, E. (2013). *Statistical methods for dynamic treatment regimes*. Springer.

Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, *20*(3), 542–542.

Chen, X., Hong, H., Tarozzi, A., et al. (2008). Semiparametric efficiency in gmm models with auxiliary data. *The Annals of Statistics*, *36*(2), 808–843.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly journal of economics*, *126*(4), 1593–1660.

Cho, Y. H., Kim, J. K., & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications*, *23*(3), 329–342.

Davidian, M., Tsiatis, A. A., & Leon, S. (2005). Semiparametric estimation of treatment effect in a pretest–posttest study with missing data. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *20*(3), 261.

Dudík, M., Langford, J., & Li, L. (2011). Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601*.

Farrell, M. H., Liang, T., & Misra, S. (2018). Deep neural networks for estimation and inference. *arXiv preprint arXiv:1809.09953*.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., … others (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, *335*(15), 1081–1090.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 331–355.

Nie, X., Brunskill, E., & Wager, S. (2020). Learning when-to-treat policies. *Journal of the American Statistical Association*, 1–18.

Qian, M., & Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of statistics*, *39*(2), 1180.

Rai, Y. (2018). Statistical inference for treatment assignment policies. *Unpublished Manuscript*.

Robins, J., Hernan, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiol.*, *11*, 550–560.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, *6*, 34–58.

Shi, C., Fan, A., Song, R., & Lu, W. (2018a). High-dimensional a-learning for optimal dynamic treatment regimes. *Annals of statistics*, *46*(3), 925–957.

Tsiatis, A. A., Davidian, M., Zhang, M., & Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in medicine*, *27*(23), 4658–4677.

Turvey, R. (2017). *Optimal pricing and investment in electricity supply: An esay in applied welfare economics*. Routledge.

van der Laan, M. J., & Luedtke, A. R. (2015). Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of causal inference*, *3*(1), 61–95.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Wang, L., Zhou, Y., Song, R., & Sherwood, B. (2018). Quantile-optimal treatment regimes. *Journal of the American Statistical Association*, *113*(523), 1243–1254.

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, *8*(3-4), 279–292.

Word, E. R., et al. (1990). The state of tennessee's student/teacher achievement ratio (star) project: Technical report (1985-1990).

Zhang, B., Tsiatis, A. A., Laber, E. B., & Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, *68*, 1010–1018.

Zhang, B., Tsiatis, A. A., Laber, E. B., & Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, *100*, 681–694.

Zhang, Y., Laber, E. B., Davidian, M., & Tsiatis, A. A. (2018). Estimation of optimal treatment regimes using lists. *J. Amer. Statist. Assoc.*, *113*(524), 1541–1549. doi: 10.1080/01621459.2017.1345743

Zhang, Y., Laber, E. B., Tsiatis, A., & Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, *71*(4), 895–904. doi: 10.1111/biom.12354

Zhao, Y., Kosorok, M. R., & Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in medicine*, *28*(26), 3294–3315.

Zhao, Y., Zeng, D., Rush, A. J., & Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.*, *107*(499), 1106–1118. doi: 10.1080/01621459.2012.695674

Zhao, Y.-Q., Zeng, D., Laber, E. B., & Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.*, *110*(510), 583–598. doi: 10.1080/01621459.2014.937488

Zhou, X., Mayer-Hamblett, N., Khan, U., & Kosorok, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, *112*(517), 169–187.

Zhu, X. J. (2005). *Semi-supervised learning literature survey* (Tech. Rep.). University of Wisconsin-Madison Department of Computer Sciences.

SUPPORTING INFORMATION

The following supporting information is available as part of the online article:

S1. Technical Proofs. The technical proofs for all the established theoretical results.

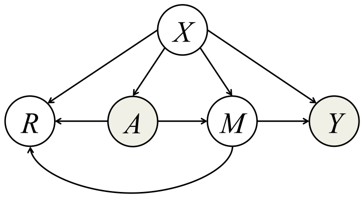

FIGURE 1 A direct acyclic graph illustrating assumptions (A2), (A4), and (A5) in the joint sample. White nodes represent observed variables, and grey nodes are variables with missing values.

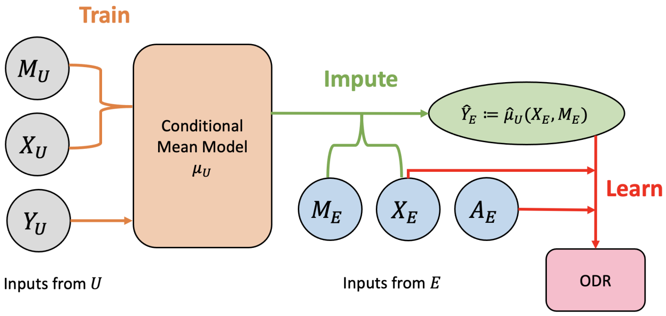

FIGURE 2 The architecture of the GEAR.

TABLE 1 Empirical results under the GEAR for Scenario 1 and 2.

	Scenario 1			**Scenario 2**		
$\mathsf{N}_\mathsf{E} =$	200	400	800	200	400	800
$\mathsf{V}(\beta_0)$		0.87			0.20	
$\widehat{\mathsf{V}}_{\mathsf{AIP}}(\widehat{\beta}^{\mathsf{G}})$	0.89	0.89	0.88	0.24	0.24	0.22
$\mathsf{SE}\{\widehat{\mathsf{V}}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathbb{E}\{\widehat{\sigma}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathsf{V}(\widehat{\beta}^{\mathsf{G}})$	0.85	0.86	0.86	0.18	0.18	0.19
CP (%)	94.6	94.8	94.8	95.0	94.4	94.8
RCD (%)	95.9	96.6	97.3	95.0	95.8	96.7
$\|\|\widehat{\beta}^{\mathsf{G}} - \beta_0\|\|_2$	0.12	0.09	0.07	0.14	0.11	0.09

TABLE 2 Empirical results under the GEAR for Scenario 3-5.

	GEAR-Linear			**GEAR-Bspline**		
$\mathsf{N}_\mathsf{E} =$	200	400	800	200	400	800
S3	$\mathsf{V}(\beta_0)$	=	1.20			
$\widehat{\mathsf{V}}_{\mathsf{AIP}}(\widehat{\beta}^{\mathsf{G}})$	1.25	1.22	1.22	1.26	1.23	1.22
$\mathsf{SE}\{\widehat{\mathsf{V}}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathbb{E}\{\widehat{\sigma}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathsf{V}(\widehat{\beta}^{\mathsf{G}})$	1.18	1.19	1.19	1.16	1.18	1.18
CP (%)	95.2	96.0	92.6	94.0	95.4	94.4
S4	$\mathsf{V}(\beta_0)$	=	2.59			
$\widehat{\mathsf{V}}_{\mathsf{AIP}}(\widehat{\beta}^{\mathsf{G}})$	2.37	2.34	2.34	2.55	2.51	2.49
$\mathsf{SE}\{\widehat{\mathsf{V}}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathbb{E}\{\widehat{\sigma}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathsf{V}(\widehat{\beta}^{\mathsf{G}})$	2.32	2.32	2.33	2.41	2.43	2.44
CP (%)	77.6	66.2	55.2	94.6	92.0	90.0
S5	$\mathsf{V}(\beta_0)$	=	3.03			
$\widehat{\mathsf{V}}_{\mathsf{AIP}}(\widehat{\beta}^{\mathsf{G}})$	2.44	2.40	2.40	3.00	2.97	2.93
$\mathsf{SE}\{\widehat{\mathsf{V}}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathbb{E}\{\widehat{\sigma}_{\mathsf{AIP}}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathsf{V}(\widehat{\beta}^{\mathsf{G}})$	2.30	2.32	2.32	2.72	2.77	2.79
CP (%)	31.6	17.4	11.8	96.0	92.4	87.8

TABLE 3 Empirical results under the GEAR for Scenario 6 when $l = \{0, 0.4, 0.8\}$. Note the true value is 0.333.

	$l = 0$			$l = 0.4$			$l = 0.8$		
	$N_E = 200$	400	800	$N_E = 200$	400	800	$N_E = 200$	400	800
$\widehat{V}_{AIP}(\widehat{\beta}^G)$	0.546	0.494	0.470	0.505	0.472	0.434	0.470	0.434	0.401
$SE\{\widehat{V}_{AIP}\}$	0.154	0.113	0.091	0.156	0.113	0.086	0.156	0.118	0.088
$\mathbb{E}\{\widehat{\sigma}_{AIP}\}$	0.158	0.118	0.092	0.160	0.120	0.093	0.162	0.121	0.093
$V(\widehat{\beta}^G)$	0.265	0.276	0.284	0.285	0.296	0.298	0.293	0.306	0.314
CP (%)	73.8	72.8	69.4	82.4	81.8	81.8	86.6	86.6	90.8
RCD (%)	79.5	80.2	81.5	83.0	84.9	85.2	84.7	87.0	89.3
$\|\|\widehat{\beta}^G - \beta_0\|\|_2$	0.457	0.413	0.371	0.388	0.322	0.306	0.358	0.288	0.232

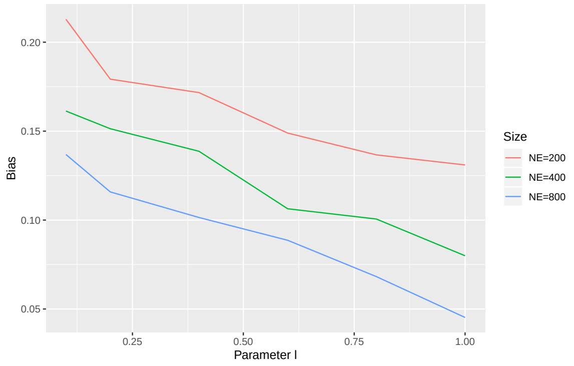

FIGURE 3 The trend of the bias of $V(\widehat{\beta}^G)$ under the GEAR over the parameter l.

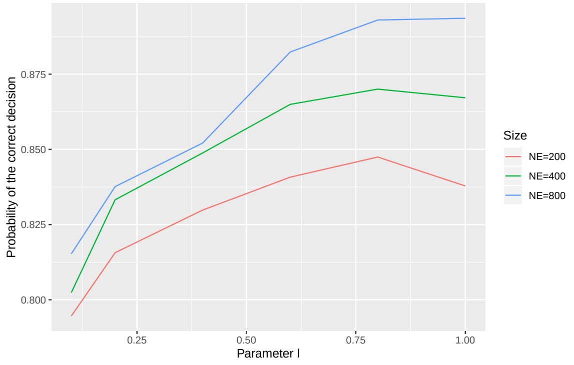

FIGURE 4 The trend of the average rate of the correct decision made by the GEAR over the parameter l.

TABLE 4 Empirical results under the GEAR when (A4) is violated.

	S1 with $C^Y = M^{(1)} + 0.8M^{(2)}$			S2 with $C^Y = M^{(1)} + 0.8M^{(2)}$		
	N_E =200	400	800	N_E =200	400	800
$V(\beta_0)$	0.87			0.20		
$\widehat{V}_{AIP}(\widehat{\beta}^G)$	0.71	0.71	0.71	0.19	0.19	0.18
$SE\{\widehat{V}_{AIP}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$\mathbb{E}\{\widehat{\sigma}_{AIP}\}$	0.02	0.01	0.01	0.02	0.01	0.01
$V(\widehat{\beta}^G)$	0.68	0.68	0.68	0.14	0.15	0.15
CP (%)	88.2	81.8	62.8	94.4	94.2	93.0
RCD (%)	94.5	95.2	95.5	93.9	94.8	95.2

TABLE 5 Comparison results for ACTG 175 data.

	Linear	**B-spline**
$\widehat{V}_{AIP}(0)$	327.8	325.7
$\widehat{V}_{AIP}(1)$	333.6	328.4
$\widehat{V}_{AIP}(\widehat{\beta}^G)$ [SD]	351.4 [10.1]	344.2 [9.9]
95% CI for $\widehat{V}_{AIP}(\widehat{\beta}^G)$	(331.7, 371.1)	(324.7, 363.8)
Assign to 'ZDV+ddC'	145	189
Assign to 'ZDV+ddI'	231	187

How to cite this article: Cai H, Lu W, and Song R (2021), GEAR: On Optimal Decision Making with Auxiliary Data, *Stat.*, *2021;xx:x–x*.

APPENDIX

A INVERSE PROPENSITY-SCORE WEIGHTED ESTIMATOR

In this appendix section, we provide the inverse propensity-score weighted (IPW) value estimator and its related theories as a middle step. All the proofs are provided in the supplementary article.

A.1 IPW Estimator for the Long-term Outcome

According to Lemma 1 and the law of large number, the value function $V(\beta)$ can be consistently estimated by

$$V_n(\beta) = \frac{1}{N_E}\sum_{i=1}^{N_E} \frac{\mathbb{I}\{A_{E,i} = d(X_{E,i};\beta)\}\mu_U(M_{E,i}, X_{E,i})}{A_{E,i}\pi(X_{E,i}) + (1 - A_{E,i})\{1 - \pi(X_{E,i})\}}.$$

We posit parametric models for $\pi(x) \equiv \pi(x;\gamma)$ and $\mu_U(m, x) \equiv \mu_U(m, x; \lambda)$ with the true model parameter γ and λ, respectively. Then the above $V_n(\beta)$ can be rewritten as the model-based form,

$$V_n^\star(\beta) = \frac{1}{N_E}\sum_{i=1}^{N_E} \frac{\mathbb{I}\{A_{E,i} = d(X_{E,i};\beta)\}\mu_U(M_{E,i}, X_{E,i};\lambda)}{A_{E,i}\pi(X_{E,i};\gamma) + (1 - A_{E,i})\{1 - \pi(X_{E,i};\gamma)\}},$$

where $\pi(x;\gamma)$ can be estimated in the experimental sample, denoted as $\pi(x;\widehat{\gamma})$, and $\mu_U(m, x; \lambda)$ can be estimated in the auxiliary sample, denoted as $\mu_U(m, x; \widehat{\lambda})$. Then, by replacing the implicit functions in $V_n^\star(\beta)$ with their parametric estimators, it is straightforward to give the following IPW estimator for the value function $\mathsf{V}(\beta)$,

$$\widehat{\mathsf{V}}(\beta) = \frac{1}{\mathsf{N_E}} \sum_{\mathsf{i}=1}^{\mathsf{N_E}} \frac{\mathbb{I}\{\mathsf{A_{E,i}} = \mathsf{d}(\mathsf{X_{E,i}}, \beta)\}\mu_{\mathsf{U}}(\mathsf{M_{E,i}}, \mathsf{X_{E,i}}; \widehat{\lambda})}{\mathsf{A_{E,i}}\pi(\mathsf{X_{E,i}}; \widehat{\gamma}) + (1 - \mathsf{A_{E,i}})\{1 - \pi(\mathsf{X_{E,i}}; \widehat{\gamma})\}}. \tag{A1}$$

Define $\widehat{\beta} = \arg\max_{\beta} \widehat{\mathsf{V}}(\beta)$ with subject to $||\beta||_2 = 1$ for identifiability purpose, with the corresponding estimated value function $\widehat{\mathsf{V}}(\widehat{\beta})$.

A.2 Theoretical Results of the IPW Estimator

First, we establish some theoretical results for the IPW estimator as a middle step to prove the results for the AIPW estimator. Here, we use $\phi_{\mathsf{X}}(\mathsf{X})$ and $\phi_{\mathsf{M}}(\mathsf{M})$ to represent appropriate basis functions for X and M, respectively. The following theorem gives the consistency result of our IPW estimator for the value function to the true.

Theorem A.1. (Consistency) When (A1)-(A9) and (A11) hold, given $\forall\beta$, we have

$$\widehat{\mathsf{V}}(\beta) = \mathsf{V}(\beta) + \mathsf{o_p}(1).$$

Next, we establish the asymptotic normality of $\sqrt{\mathsf{N_E}}\{\widehat{\mathsf{V}}(\widehat{\beta}) - \mathsf{V}(\beta_0)\}$ through the following lemma that states the estimator $\widehat{\beta}$ has a cubic rate towards the true β_0.

Lemma 4. Under (A1)-(A11), we have

$$\mathsf{N_E^{1/3}}||\widehat{\beta} - \beta_0||_2 = \mathsf{O_p}(1), \tag{A2}$$

where $||\cdot||_2$ is the $\mathsf{L_2}$ norm.

We next show the asymptotic distribution of $\widehat{\mathsf{V}}(\widehat{\beta})$ as follows.

Theorem A.2. (Asymptotic Distribution) When (A1)-(A11) are satisfied, we have

$$\sqrt{\mathsf{N_E}}\{\widehat{\mathsf{V}}(\widehat{\beta}) - \mathsf{V}(\beta_0)\} \xrightarrow{\mathcal{D}} \mathsf{N}(0, \sigma^2_{\mathsf{IPW}}), \tag{A3}$$

where $\sigma^2_{\mathsf{IPW}} = \mathsf{t}\sigma^2_{\mathsf{U}} + \sigma^2_{\mathsf{E,I}}$, and $\sigma^2_{\mathsf{U}} = \mathbb{E}[\{\xi_{\mathsf{i}}^{(\mathsf{U})}\}^2]$ and $\sigma^2_{\mathsf{E,I}} = \mathbb{E}[\{\xi_{\mathsf{i}}^{(\mathsf{E,I})}\}^2]$.

Here, $\xi_{\mathsf{i}}^{(\mathsf{U})} \equiv \mathsf{G}_2^\top \mathsf{H}_2^{-1} \begin{bmatrix} \phi_{\mathsf{X}}(\mathsf{X_{U,i}}) \\ \phi_{\mathsf{M}}(\mathsf{M_{U,i}}) \end{bmatrix} \{\mathsf{Y_{U,i}} - \mu_{\mathsf{U}}(\mathsf{M_{U,i}}, \mathsf{X_{U,i}}; \lambda)\}$ is the I.I.D. term in the auxiliary sample, and $\xi_{\mathsf{i}}^{(\mathsf{E,I})} \equiv \mathsf{G}_1^\top \mathsf{H}_1^{-1}\phi_{\mathsf{X}}(\mathsf{X_{E,i}})\{\mathsf{A_{E,i}} - \pi(\mathsf{X_{E,i}}; \gamma)\} + \{\mathsf{A_{E,i}}\pi(\mathsf{X_{E,i}}; \gamma) + (1 - \mathsf{A_{E,i}})\{1 - \pi(\mathsf{X_{E,i}}; \gamma)\}\}^{-1}\mathbb{I}\{\mathsf{A_{E,i}} = \mathsf{d}(\mathsf{X_{E,i}}; \beta_0)\}\mu_{\mathsf{U}}(\mathsf{M_{E,i}}, \mathsf{X_{E,i}}; \lambda) - \mathsf{V}(\beta_0)$ is the I.I.D. term in the experimental sample.