Classifying Treatment Responders: Bounds and Algorithms

Anpeng Wu*
Zhejiang University & MBZUAI
Hangzhou, China & Abu Dhabi, UAE
anpwu@zju.edu.cn

Haoxuan Li*
Peking University & MBZUAI
Beijing, China & Abu Dhabi, UAE
hxli@stu.pku.edu.cn

Chunyuan Zheng*
Peking University & MBZUAI
Beijing, China & Abu Dhabi, UAE
cyzheng@stu.pku.edu.cn

Kun Kuang†
Zhejiang University
Hangzhou, China
kunkuang@zju.edu.cn

Kun Zhang†
CMU & MBZUAI
Pittsburgh, USA & Abu Dhabi, UAE
kunz1@cmu.edu

Abstract

Treatment responders are individuals whose outcomes would change from negative to positive if treated, and learning a classifier to predict responders would help causal decision-making in real applications. Although many treatment effect estimation methods have been proposed to identify treatment responders, there are fundamental differences between treatment effect estimation and treatment responder classification, including: (1) accurate causal effect estimation is not necessary for optimal intervention decisions; (2) methods for accurate causal effect estimation do not directly optimize classification loss; (3) treatment responder classification requires identifying joint potential outcomes, while treatment effect estimation focuses on marginal distributions. To fill this gap, we tackle the treatment responder classification problem without assuming monotonicity. We derive sharp bounds of the probability that an individual is a responder and determine a sharp upper bound on the weighted classification risk to measure the worst classification performance. Based on these findings, we further propose a **C**lassifying **T**reatment **R**esponder **L**earning (**CTRL**) algorithm to accurately identify the treatment responders, and theoretically demonstrate the superiority of jointly learning over two-stage learning. Extensive experiments on semi-synthetic and real-world datasets show that our method better predicts treatment responders and adaptively trades off false-positives and false-negatives with varying weight coefficients.

CCS Concepts

• **Computing methodologies** → **Machine learning approaches**.

Keywords

Treatment Responder, Causal Effect Estimation, Causal Decision-Making, Causal Classification, Sharp Bound

*Anpeng Wu, Haoxuan Li, and Chunyuan Zheng contributed equally to this research.
†Corresponding authors.

KDD '25, August 3–7, 2025, Toronto, ON, Canada

ACM ISBN 979-8-4007-1245-6/25/08
https://doi.org/10.1145/3690624.3709191

ACM Reference Format:
Anpeng Wu, Haoxuan Li, Chunyuan Zheng, Kun Kuang, and Kun Zhang. 2025. Classifying Treatment Responders: Bounds and Algorithms. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1 (KDD '25), August 3–7, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3690624.3709191

1 Introduction

Estimating treatment effects from observational data plays a crucial role in a wide range of areas such as healthcare [12], marketing [16], and reinforcement learning [32]. The key challenge of this task arises because, for any individual in the collected data, we can only observe the factual outcome associated with the actually assigned treatment, while the counterfactual outcomes associated with other treatment arms are unavailable, which is also known as the fundamental problem of causal inference [15, 25]. To accurately estimate counterfactual results, benefiting from recent advances in deep learning, representation learning methods are proposed to learn a balanced representation of covariates to remove confounding bias [11, 17, 30, 37], demonstrating superior and promising prediction performance.

Despite many causal effect estimation methods have been proposed, methods designed directly for causal classification[1] have rarely been discussed. In particular, when the estimated causal effects are utilized for downstream decision-making tasks, the population will be classified into two subgroups, containing individuals who will apply the treatment and those who will not, respectively. In such cases, as shown in Table 1, researchers aim to identify individuals whose outcomes would change from negative to positive if treated, i.e., Treatment Responder Classifier. This is fundamentally different from the causal effects estimation task in the following three aspects: (1) **Accurate causal effect estimation is not necessary to make optimal intervention decisions**, and more accurate estimates of causal effects may not imply better decision-makings [9]. (2) **Causal effect estimation methods do not directly optimize the classification loss**, which may fail in producing the Bayes-optimal classifiers if the causal effect estimates incorporate errors [18]. (3) **Causal classification consists of identifying treatment responders**, those individuals whose outcome would change from negative to positive if they were treated [8], which requires the estimation of *joint* potential outcomes of an individual. Nonetheless, the conditional average treatment effect

[1]Causal classification aims to minimize the classification risk between the optimal treatment regime (OTR) and its predicted value.

(CATE) only describes the treatment effects on sub-populations. As a coarser-grained estimand identified via the marginal distribution of potential outcomes, CATE cannot identify treatment responders.

To tackle the above issues, we aim to accurately identify these treatment responders from a causal classification perspective. First, for a classifier that uses covariates to predict whether an individual is a responder or not, we define a weighted misclassification loss, where weights are used to trade-off between the false positives and false negatives. When estimating the probability that an individual is a responder given the covariates, we relax the monotonicity assumption made in previous work [18] that the individual treatment effect is nonnegative, and this leads to the parameter of interest being partially identifiable in our work, even when the strong unconfoundedness assumption holds. Therefore, we theoretically derive the sharp bounds (also known as the tightest possible bounds) of the probability that an individual is a responder given the covariates. Based on this, we derive a sharp upper bound on the weighted classification risk to measure the worst classification performance of the classifier. In particular, the sharp upper bound can be decomposed into two terms: the first term (called surrogate risk) relies on both the causal classifier and the potential outcome regressors, while the second term (called weighted risk) relies only on the potential outcome regressors. We show that the classification risk of the classifier learned by minimizing the proposed sharp upper bound converges to the optimal classification risk.

By utilizing the above theoretical results, we further propose a **C**lassifying **T**reatment **R**esponder **L**earning algorithm, namely **CTRL**, to accurately identify the treatment responders. Specifically, CTRL extends previous representation learning methods to include both counterfactual regression networks and responder classification network, which are built on a shared representation network to alleviate the confounding effect. In addition, we theoretically demonstrate the superiority of jointly learning counterfactual regression networks and responder classification network for minimizing the classification risk compared to the two-stage learning, which first pre-trains the counterfactual regression networks, then trains the responder classification network. We also empirically demonstrate that our method can better predict the treatment responder as well as adaptively trade-off the false-positives and false-negatives on both the synthetic and real-world datasets.

Our main contributions are summarized as follows:

- We address the treatment responder classification problem without assuming the monotonicity assumption, and derive the sharp bounds of the probability that an individual is a responder.
- We derive a sharp upper bound on the weighted classification risk to measure the worst classification performance and prove that the classification risk of the classifier learned by minimizing the proposed sharp upper bound converges to the optimal classification risk via generalization theory.
- We further propose a joint learning algorithm for classifying the treatment responders, named CTLR, and theoretically demonstrate the superiority of jointly learning over two-stage learning.
- We conduct extensive experiments on two real-world datasets and one semi-synthetic dataset, and the results demonstrate the superiority of our joint learning approach in classifying the treatment responders under varying misclassification weights.

2 Related Work

Causal Effect Estimation. In conventional causal methods, most previous works focus on predicting counterfactual outcomes and then making a comparison to identify the optimal treatment. To mitigate confounding bias arising from imbalanced data, researchers have developed a variety of methods, including those based on propensity scores, tree models, representation balancing, and generative models. Propensity score-based methods estimate the treatment probability conditional on covariates to adjust for confounding by matching[6], stratification[14], or doubly-robust learning[3]. Tree-based methods [36] usually build numerous causal trees to estimate heterogeneous treatment effects. Recently, Johansson et al. [17], Shalit et al. [30] proposed learning balanced representation to minimize confounding bias. Building upon this, Hassanpour and Greiner [11] proposed disentangled representations for counterFactual regression, while Zhong et al. [42] extended these methods to cover entire space counterfactual regression, aiming to enhance the precision and applicability of causal inference techniques. Further, Wang et al. [37] introduced an innovative approach to optimal transport in causality, adding a relaxed mass-preserving regularizer to refine the handling of mini-batch sampling effects. Generative methods include CEVAE [23] using variational autoencoders for hidden confounders, TEDVAE [41] inferring and disentangling latent variables, and GANITE [40] directly generating counterfactual outcomes. Although these methods have achieved significant success in estimating treatment effects, recent works have highlighted that accurate causal effect estimation does not necessarily translate into precise causal decision-making [9]. Unlike the above studies focusing on the causal effect estimation, our work aims to classifying treatment responders under partial identification.

Causal Classification. Causal classification aims to find the optimal treatment level (or action) that maximizes the expected value of the outcome of interest given the covariates [10], and is crucial in areas such as healthcare [12], marketing [16], and reinforcement learning [32]. For classifying treatment responders, Kallus [18] assumed the causal effect monotonicity to identify the responder probability. Fernández-Loría and Loría [10] proposed causal scoring, as a framework for treatment effect estimation, effect ordering, and effect classification. Fernández-Loría and Provost [8] present a theoretical analysis comparing treatment effect estimation and outcome prediction when addressing causal classification. Another task closely related to treatment responder classification in this paper is the contextual multiarmed bandit (MAB), which consists of estimating the arm (treatment level) with the largest reward (potential outcome) given some context (feature vector). As one of the online learning setting, the policy repeatedly observes a context, takes an action, and then observes a reward only for the chosen action [22]. Many methods have been proposed to maximize the cumulative reward, including regression based [5], reweighted based [20, 34, 35], and doubly robust methods [2, 7, 33, 38]. However, an important distinction with respect to our setting is that the goal in bandit problems is to learn a classification model while actively making treatment assignment decisions for incoming subjects. In this paper, we extend Kallus [18] to consider treatment responder classification without the monotonicity assumption, leading to partial identified causal parameters of interest.

Table 1: The units are divided into four types according to $Y(-1)$ and $Y(+1)$, named "responder", "type-1 non-responder", "type-2 non-responder", and "type-3 non-responder". Real Treatment Effect (RTE) is defined as $A = (Y(+1) - Y(-1))/2$.

Unit type	$Y(-1)$	$Y(+1)$	R (OTR)	τ (CATE)
Responder	**−1**	**+1**	**+1**	**+2**
Type-1 non-responder	−1	−1	−1	0
Type-2 non-responder	+1	+1	−1	0
Type-3 non-responder	+1	−1	−1	−2

In this paper, we extend the work of Kallus [18] to treatment responder classification without the monotonicity assumption, deriving a sharp upper bound on classification risk and proposing a joint learning algorithm to minimize it. Our CTRL method differs from previous approaches in several key ways: Fernández-Loría and Provost [8] analyzed the bias-variance tradeoff in causal classification but noted direct outcome prediction cannot identify responders; Fernández-Loría and Provost [9] distinguished between causal decision-making (CDM) and causal effect estimation (CEE), without providing a unified framework for CDM; Fernández-Loría and Loría [10] introduced causal scoring but did not address misclassification risks or treatment responder classification; and Kallus [19] focused on average hinge effect estimation, rather than CDM. In contrast, our method directly optimizes classification risk through joint learning, allowing tasks to exchange information and improve representations. We show in Proposition 5.1 that CTRL outperforms two-phase learning, and with infinite data, it reduces misclassification risk by pushing the classifier away from decision boundaries.

3 Problem Setup

In this paper, we consider the case of binary treatment. Suppose a simple random sampling of n units from a super population $\mathbb{P}$, for each unit i, the covariate and the assigned treatment are denoted as $X_i \in \mathbb{R}^m$ and $T_i \in \{-1, +1\}$, respectively, where $T_i = +1$ means receiving treatment, while $T_i = -1$ means not receiving treatment. Let $Y_i \in \{-1, +1\}$ be the corresponding binary outcome. Without loss of generality, we assume that the larger outcome is preferable. To study the treatment responder classification, we adopt the potential outcome framework [26, 28] in causal inference. Specifically, let $Y_i(-1)$ and $Y_i(+1)$ be the outcome of unit i had this unit receive treatment $T_i = -1$ and $T_i = +1$, respectively. Since each unit can be only assigned with one treatment, we always observe the corresponding outcome be either $Y_i(-1)$ or $Y_i(+1)$, but not both, which is also known as the fundamental problem of causal inference [15, 25].

We assume that the observation for unit i is $Y_i = Y_i(T_i)$. In other words, the observed outcome is the potential outcome corresponding to the assigned treatment, which also known as the consistency assumption in the causal literature. We assume that the stable unit treatment value assumption (STUVA) assumption holds, i.e., there should not be alternative forms of the treatment and interference between units. Furthermore, we follow [18] and [8] to assume that the strong ignorability assumption holds, i.e., $Y(-1), Y(+1) \perp\!\!\!\perp T \mid X$ and let $\eta < \mathbb{P}(T_i = +1 | X_i = x) < 1 - \eta$, where η is a constant between 0 and 1/2. The key notations used in this paper are summarized in Table 4.

As shown in Table 1, we define real treatment effect (RTE) as $A = (Y(+1) - Y(-1))/2$ and divide units into four types according to the joint values of $Y(-1)$ and $Y(+1)$, in which the responders refer to those individuals for whom treatment would have a positive effect. In addition to the responders, we have individuals who would have a negative outcome even with treatment (type-1 non-responders), individuals who would have a positive outcome with or without treatment (type-2 non-responders), and individuals for whom treatment would have a negative effect (type-3 non-responders). Denote the indicator of responder as

$$R = \begin{cases} +1 & Y(+1) > Y(-1) \quad \text{(responder)} \\ -1 & \text{Otherwise} \quad \text{(non-responder)} \end{cases}$$

and the probability of a responder given the covariates as

$$\rho(X) = \mathbb{P}(R = +1 \mid X),$$

we now aim at learning a classifier $f : \mathbb{R}^p \to \{-1, +1\}$ to predict R from X. For $\theta \in [0, 1]$, we define the weighted misclassification loss of the treatment responder classifier as

$$\begin{aligned} L_\theta(f) &= \theta \cdot \mathbb{P}(\text{false positive}) + (1 - \theta) \cdot \mathbb{P}(\text{false negative}) \\ &= \theta \cdot \mathbb{P}(f(X) = +1, R = -1) + (1 - \theta) \cdot \mathbb{P}(f(X) = -1, R = +1). \end{aligned}$$

4 Sharp Bounds on Classification Risk

To facilitate understanding of the gap between treatment effect estimation and treatment responder classification, we begin with the discussion on the relation between responder probability $\rho(X)$ and the conditional average treatment effect (CATE) defined as

$$\tau(X) = \mathbb{E}[Y(+1) - Y(-1) \mid X],$$

which measures the the difference in the conditional mean outcomes between treatments given covariates.

Lemma 4.1 (Relation to CATE). $\rho(X) = \frac{\tau(X)}{2} + \mathbb{P}(A = -1 \mid X)$.

For identifying $\rho(X)$, Kallus [18] assumes the monotonicity assumption that $Y(+1) \geq Y(-1)$, *i.e.*, type-3 non-responders do not exist. Then we have $\rho(X) = \frac{\tau(X)}{2}$ as a special case because of $\mathbb{P}(A = -1 \mid X) = 0$. Below, we aim to identify the treatment responders without the monotonicity assumption, and one of the critical challenges is, even with the strong unconfoundedness assumption, we are still unable to directly identify the responder probability $\rho(X)$ from observational data, which motivates us to derive its sharp bounds (also known as the tightest possible bounds).

Lemma 4.2 (Sharp Bounds on Responders). *The sharp lower bound of responder probability $\rho(X)$ is*

$$\rho(X) \geq \max\left\{\frac{\tau(X)}{2}, 0\right\};$$

The sharp upper bound of responder probability $\rho(X)$ is

$$\begin{aligned} \rho(X) &\leq \min\{1 - \mathbb{P}(Y(-1) = +1 \mid X), \mathbb{P}(Y(+1) = +1 \mid X)\} \\ &= \min\{1 - \mathbb{P}(Y = +1 \mid X, T = -1), \mathbb{P}(Y = +1 \mid X, T = +1)\}. \end{aligned}$$

Lemma 4.3 (Classification Risk). *The classification risk is*

$$L_\theta(f) = \underbrace{\frac{1}{2}\mathbb{E}[f(X)(\theta - \rho(X))]}_{\text{surrogate risk } L'_\theta(f,\rho)} + \underbrace{\frac{1}{2}\mathbb{E}[\theta + (1-2\theta)\rho(X)]}_{\text{weighted risk } L'_\theta(\rho)}.$$

Proof Sketch. By the law of iterated expectations, we have

$$\begin{aligned}
&L_\theta(f) = \theta \cdot \mathbb{P}(\text{false positive}) + (1-\theta)\cdot \mathbb{P}(\text{false negative})\\
&= \theta \cdot \mathbb{E}[\mathbb{I}[f(X)=+1]\mathbb{I}[R=-1]] + (1-\theta)\cdot \mathbb{E}[\mathbb{I}[f(X)=-1]\mathbb{I}[R=+1]]\\
&= \frac{\theta}{2}\cdot \mathbb{E}[(1+f(X))(1-\rho(X))] + \frac{1-\theta}{2}\cdot \mathbb{E}[(1-f(X))\rho(X)]\\
&= \frac{1}{2}\cdot \mathbb{E}[f(X)(\theta(1-\rho(X)) - (1-\theta)\rho(X))]\\
&+ \frac{1}{2}\cdot \mathbb{E}[\theta(1-\rho(X)) + (1-\theta)\rho(X))]\\
&= \frac{1}{2}\mathbb{E}[f(X)(\theta-\rho(X))] + \frac{1}{2}\mathbb{E}[\theta + (1-2\theta)\rho(X)],
\end{aligned}$$

which leads to the conclusion in Lemma 4.3. □

Lemma 4.3 indicates that the classification risk can be decomposed into two terms: the first term (called surrogate risk) relies on both the causal classifier and the potential outcome regressors, while the second term (called weighted risk) relies only on the potential outcome regressors. From the first term, given the true conditional probability $\rho(X)$, the Bayes-optimal classifier is $f^*_\theta(X) = \mathbb{I}[\rho(X) - \theta]$. Notably, without the monotonicity assumption of the treatment effect, the responder probability $\rho(X)$ is unidentifiable.

By combining Lemmas 4.2 and 4.3, we obtain the sharp upper bounds on the classification risk in below.

Theorem 4.4 (Sharp Bounds on Classification Risk). *When $\theta \in [0, 1/2]$, we have the classification risk upper bounded by*

$$L_\theta(f) \le \underbrace{\frac{1}{2}\mathbb{E}[f(X)(\theta - \max\left\{\frac{\tau(X)}{2}, 0\right\})]}_{\text{surrogate risk } L'_\theta(f,\tau)}$$
$$+ \underbrace{\frac{1}{2}\mathbb{E}[\theta + (1-2\theta)\min\{1-\mathbb{P}(Y(-1)=+1\mid X), \mathbb{P}(Y(+1)=+1\mid X)\}]}_{\text{weighted risk } L'_{\theta\le 1/2}(\mu_{-1},\mu_{+1})};$$

When $\theta \in [1/2, 1]$, we have the classification risk upper bounded by

$$L_\theta(f) \le \underbrace{\frac{1}{2}\mathbb{E}[f(X)(\theta - \max\left\{\frac{\tau(X)}{2}, 0\right\})]}_{\text{surrogate risk } L'_\theta(f,\tau)}$$
$$+ \underbrace{\frac{1}{2}\mathbb{E}[\theta + (1-2\theta)\max\left\{\frac{\tau(X)}{2}, 0\right\}]}_{\text{weighted risk } L'_{\theta>1/2}(\mu_{-1},\mu_{+1})}$$

The above sharp upper bounds on the weighted classification risk provide an effective measurement of the worst classification performance of predicting the treatment responders, then the treatment responder classifier is trained by minimizing the sharp upper bounds. Finally, we derive the generalization bound as follows.

Definition 4.5 (Empirical Rademacher complexity). Let $\mathcal{F}$ be a family of functions mapping from x to $\{-1,+1\}$ and $S = (x_1,\ldots,x_n)$ a fixed sample of size n with elements in x. Then, the empirical Rademacher complexity of $\mathcal{F}$ with respect to the sample S is

$$\mathfrak{R}(\mathcal{F}) = \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(x_i)\right],$$

where $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_n)^\top$, with σ_is independent uniform random variables taking values in $\{-1,+1\}$. The random variables σ_i are called Rademacher variables.

Theorem 4.6 (Generalization Bound). *With probability at least $1-\delta$, we have*

$$L_\theta(\hat{f}) \le \min_{f\in\mathcal{F}} L_\theta(f) + 2\max\{\theta, \theta - \max_X\{\tau(X)\}\}\mathfrak{R}(\mathcal{F})$$
$$+ 5\max\{\theta, \theta - \max_X\{\tau(X)\}\}\sqrt{\frac{2\ln(8/\delta)}{n}}.$$

Remarkably, the last two terms will vanish as the sample size tends to infinity, which shows that the classification risk of the classifier learned by minimizing the proposed sharp upper bound converges to the optimal classification risk. All proofs of theorems and lemmas are deferred to Appendix C.

5 CTRL: Classifying Treatment Responder Learning Algorithm

Guided by the above preliminary theorems, we propose a Classifying Treatment Responder Learning algorithm, namely CTRL, to perform counterfactual regression and causal decision-making. CTRL seeks to improve how we identify individuals who will respond positively to treatments, enhancing the precision and effectiveness of causal interventions. Specifically, the overall architecture of CTRL consists of the following components:

- Shared representation network $\Phi(X)$ for learning balanced representation that benefits multiple tasks simultaneously;
- Counterfactual regression networks $h_{-1}(\Phi(X))$ and $h_{+1}(\Phi(X))$ for estimating $\mathbb{P}(Y(-1)=+1\mid X)$ and $\mathbb{P}(Y(+1)=+1\mid X)$;
- Responder Classifier $f(X) = \text{Sign}(h_f(\Phi(X)))$ for distinguishing whether an individual is a responder.

Below, we first show that the classification risk for a responder classifier trained via joint learning is lower than that trained via two-phase learning. Then, we explain how to optimize $h_{-1}(\Phi(X))$, $h_{+1}(\Phi(X))$ and $h_f(\Phi(X))$ by minimizing the proposed losses.

5.1 Joint Learning for Lower Classification Risk

The objective of CTRL is to minimize the classification risk $L_\theta(f)$ of the responder classifier $f(X)$ with the weight θ. Remarkably, in the proposed CTRL, we prefer "joint learning" of τ and f, rather than "two-phase learning" (first τ, then f). Here, $\tau(X) = h_{+1}(\Phi(X)) - h_{-1}(\Phi(X))$ represents the treatment effect, and the classification risk of the responder classifier is reduced through joint learning compared to two-phase learning.

Two-phase learning: first, learn $\hat{\tau}$ by the existing CATE estimation method; then, learn f by minimizing the surrogate risk (1st)

$$\hat{f}^{TP} = \arg\min_{f\in\mathcal{F}}\mathbb{E}[f(X)(\theta - \max\left\{\frac{\tau(X)}{2}, 0\right\})],$$

Algorithm 1: Alternating Training for Joint Learning.

Input: Factual samples $(x_1, t_1, y_1), \ldots, (x_n, t_n, y_n)$, $\lambda > 0$, classification network $h_{\mathbf{U}}$ with initial weights $\mathbf{U}$, outcome network $h_{\mathbf{V}}$ with initial weights $\mathbf{V}$, representation network $\Phi_{\mathbf{W}}$ with initial weights $\mathbf{W}$.

while *stopping criteria is not satisfied* **do**
  for *number of steps for training classification network* **do**
    Sample a mini-batch $\{i_1, i_2, \ldots, i_m\} \subset \{1, 2, \ldots, n\}$;
    $[\mathbf{U}, \mathbf{W}] \leftarrow [\mathbf{U}, \mathbf{W}] - \eta \nabla_{\mathbf{U},\mathbf{W}} \mathcal{L}_\theta(h_{\mathbf{U}}, h_{\mathbf{V}}, \Phi_{\mathbf{W}})$;
  end
  for *number of steps for training outcome network* **do**
    Sample a mini-batch $\{j_1, j_2, \ldots, j_m\} \subset \{1, 2, \ldots, n\}$;
    $[\mathbf{V}, \mathbf{W}] \leftarrow [\mathbf{V}, \mathbf{W}] - \lambda\eta \nabla_{\mathbf{V},\mathbf{W}} \mathcal{L}_{CFR}(h_{\mathbf{V}}, \Phi_{\mathbf{W}})$;
    if $\theta \leq 1/2$ **then**
      $[\mathbf{V}, \mathbf{W}] \leftarrow [\mathbf{V}, \mathbf{W}] - \eta \nabla_{\mathbf{V},\mathbf{W}} \mathcal{L}_{\theta \leq 1/2}(h_{\mathbf{V}}, \Phi_{\mathbf{W}})$;
    else
      $[\mathbf{V}, \mathbf{W}] \leftarrow [\mathbf{V}, \mathbf{W}] - \eta \nabla_{\mathbf{V},\mathbf{W}} \mathcal{L}_{\theta > 1/2}(h_{\mathbf{V}}, \Phi_{\mathbf{W}})$;
    end
  end
end

Joint learning: learn f by minimizing both of the risks

$$(\hat{f}^{JL}, \hat{\mu}_{-1}^{JL}, \hat{\mu}_{+1}^{JL}) = \arg \min_{f,\mu_{-1},\mu_{+1}} \mathbb{E}[f(X)(\theta - \max\left\{\frac{\tau(X)}{2}, 0\right\})] + \mathbb{I}(\theta > 1/2) L'_{\theta>1/2}(\mu_{-1}, \mu_{+1}) + \mathbb{I}(\theta \leq 1/2) L'_{\theta\leq 1/2}(\mu_{-1}, \mu_{+1}).$$

Proposition 5.1. *The classification risk of the responder classifier trained via joint learning is lower than that via two-phase learning*

$$L_\theta(\hat{f}^{JL}) \leq L_\theta(\hat{f}^{TP}).$$

Proposition 5.1 highlights the superiority of joint learning over two-phase learning in minimizing the classification risk of the responder classifier within the CTRL framework. By integrating the estimation of τ and the optimization of f into a cohesive learning process, joint learning not only simplifies the methodology but also improves the efficacy of the classification risk reduction.

5.2 Optimizing with the Proposed Loss Function

Based on the above theoretical findings, CTRL introduces three distinct loss functions to alternately optimize $h_{-1}(\Phi(X))$, $h_{+1}(\Phi(X))$, and $h_f(\Phi(X))$ using joint learning.

(I) Classifier learning loss function $\mathcal{L}_\theta(h_f, h_{-1}, h_{+1}, \Phi)$ is

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i)(\theta - \max\{h_{+1}(\Phi(x_i)) - h_{-1}(\Phi(x_i)), 0\}).$$

By minimizing the term $\mathcal{L}_\theta(h_f, h_{-1}, h_{+1}, \Phi)$, our model ensures $f(x_i)$ becomes an accurate classifier. When $\theta > h_{+1}(\Phi(x_i)) - h_{-1}(\Phi(x_i))$, indicating the risk coefficient exceeds the probability of benefit, this term will push $f(x_i)$ towards 0; conversely, it will push it towards 1. Additionally, by alternately optimizing $h_f(\Phi(x_i))$ and $\tau = h_{+1}(\Phi(x_i)) - h_{-1}(\Phi(x_i))$, this approach helps us distance τ from the decision boundary θ.

(II) Classifier learning loss function $\mathcal{L}_{\theta>1/2}(h_{-1}, h_{+1}, \Phi)$ is

$$\frac{1}{n}\sum_{i=1}^{n}(1 - 2\theta)\max\{h_{+1}(\Phi(x_i)) - h_{-1}(\Phi(x_i)), 0\}.$$

When $\theta > 1/2$, minimizing the term $\mathcal{L}_{\theta>1/2}(h_{-1}, h_{+1}, \Phi)$ indicates a reduction in the probability of adopting treatment due to excessive risk. The higher the risk coefficient θ, the more we aim to lower the probability of treatment adoption, suggesting $\hat{\tau}(x_i) = h_{+1}(\Phi(x_i)) - h_{-1}(\Phi(x_i))$ should only be minimized when necessary.

Similarly, classifier learning loss function $\mathcal{L}_{\theta\leq 1/2}(h_{-1}, h_{+1}, \Phi)$ is

$$\frac{1}{n}\sum_{i=1}^{n}(1 - 2\theta)\min\{1 - h_{-1}(\Phi(x_i)), h_{+1}(\Phi(x_i))\}.$$

When $\theta \leq 1/2$, minimizing the term $\mathcal{L}_{\theta\leq 1/2}(h_{-1}, h_{+1}, \Phi)$ means either minimizing $h_{+1}(\Phi(x_i))$ or maximizing $h_{-1}(\Phi(x_i))$, which also reduces $\hat{\tau} = h_{+1}(\Phi(x_i)) - h_{-1}(\Phi(x_i))$, thereby decreasing the probability of treatment adoption. As θ increases, the penalty becomes more large, leading to a probability reduction in the adoption of treatment.

(III) Counterfactual regression with balancing representation is

$$\mathcal{L}_{CFR}(h_{-1}, h_{+1}, \Phi) = \frac{1}{n}\sum_{i=1}^{n} w_i \cdot \mathcal{L}(h(\Phi(x_i), t_i), y_i) + \Omega(h) + \alpha \cdot \mathrm{IPM}_G(\{\Phi(x_i)\}_{i:t_i=0}, \{\Phi(x_i)\}_{i:t_i=1})$$

$$\text{with } w_i = \frac{t_i}{2u} + \frac{1 - t_i}{2(1 - u)}, \quad \text{where } u = \frac{1}{n}\sum_{i=1}^{n} t_i,$$

and Ω is a model complexity term. In the CTRL framework, counterfactual regression can be flexibly exploited from conventional causal methods, and we take the widely used counterfactual regression (CFR) [30], which uses Integral Probability Metric IPM_G to measure the discrepancy between the representation distributions of the treated group and the control group. Consistent with CFR, in this paper, we adopt the Wasserstein measure as IPM_G. By minimizing the term IPM_G, we obtain a shared representation that balances the confounding effect. Subsequently, minimizing $\mathcal{L}_{CFR}(h_{-1}, h_{+1}, \Phi)$ aids in learning counterfactual estimators $h_{-1}(\Phi(X))$ and $h_{+1}(\Phi(X))$ for estimating counterfactual outcomes $Y(-1)$ and $Y(+1)$.

Overall, the full objective loss function would be:

$$\mathcal{L}_\theta^{JL}(h_f, h_{-1}, h_{+1}, \Phi) = \mathcal{L}_\theta(h_f, h_{-1}, h_{+1}, \Phi) + \lambda\mathcal{L}_{CFR}(h_{-1}, h_{+1}, \Phi) + \mathbb{I}(\theta \leq 1/2)\mathcal{L}_{\theta\leq 1/2}(h_{-1}, h_{+1}, \Phi) + \mathbb{I}(\theta > 1/2)\mathcal{L}_{\theta>1/2}(h_{-1}, h_{+1}, \Phi),$$

where λ is a hyper-parameter for trade-off the classification risk and the accuracy of treatment effect estimation. Then, we adopt a joint learning strategy to alternatively optimize the $f(X) = h_f(\Phi(X))$ and $\hat{\tau}(X)$. The details are shown in Algorithm 1. We generalize our CTRL to non-binary treatments in Appendix D.

In the CTRL framework, we use CFRNet as the regression network, with two fully-connected (FC) layers using ELU activation for representation learning, two FC layers for predicting control and treated outcomes, and two Tanh-activated layers for classification, with softmax output. Each layer has 16 hidden units. The model is trained for 2000 iterations with a learning rate of 0.001, using the Adam optimizer and a batch size of 1000. The λ is set to 1. The codes are available at: https://github.com/anpwu/CTRL.

Table 2: Results (mean±std) of FPR, FNR, Risk and Rel. Err. on Toy Dataset with $\theta = 0.4$

	Within-samples on Toy with $\theta = 0.4$				Out-of-samples on Toy with $\theta = 0.4$			
	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓
S-learner	0.226 ± 0.023	0.248 ± 0.013	0.239 ± 0.004	104.3%	0.226 ± 0.027	0.250 ± 0.018	0.240 ± 0.001	101.7%
T-learner	0.137 ± 0.018	0.111 ± 0.019	0.121 ± 0.008	3.4%	0.188 ± 0.022	0.163 ± 0.018	0.173 ± 0.005	45.4%
X-learner	0.149 ± 0.017	0.124 ± 0.013	0.134 ± 0.004	14.5%	0.183 ± 0.020	0.160 ± 0.016	0.169 ± 0.005	42.0%
R-learner	0.141 ± 0.057	0.186 ± 0.078	0.168 ± 0.026	43.6%	0.162 ± 0.066	0.216 ± 0.076	0.195 ± 0.021	63.9%
DR-learner	0.142 ± 0.010	0.110 ± 0.010	0.122 ± 0.004	4.3%	0.179 ± 0.020	0.168 ± 0.019	0.172 ± 0.004	44.5%
CF	0.191 ± 0.044	0.270 ± 0.023	0.238 ± 0.005	103.4%	0.191 ± 0.046	0.273 ± 0.032	0.240 ± 0.002	101.7%
DRORF	0.209 ± 0.019	0.259 ± 0.014	0.239 ± 0.004	104.3%	0.204 ± 0.021	0.264 ± 0.014	0.240 ± 0.002	101.7%
DMLORF	0.225 ± 0.023	0.249 ± 0.014	0.239 ± 0.004	104.3%	0.215 ± 0.026	0.257 ± 0.018	0.240 ± 0.002	101.7%
CFRNet	0.159 ± 0.025	0.144 ± 0.026	0.150 ± 0.016	28.2%	0.171 ± 0.027	0.154 ± 0.028	0.161 ± 0.014	35.2%
DeRCFR	0.170 ± 0.070	0.118 ± 0.059	0.139 ± 0.014	18.8%	0.181 ± 0.072	0.126 ± 0.061	0.148 ± 0.015	24.4%
ESCFR	0.418 ± 0.042	0.034 ± 0.009	0.188 ± 0.015	60.7%	0.428 ± 0.035	0.041 ± 0.010	0.196 ± 0.012	64.7%
CEVAE	0.346 ± 0.034	0.071 ± 0.015	0.181 ± 0.006	54.7%	0.347 ± 0.046	0.080 ± 0.020	0.187 ± 0.007	57.1%
DragonNet	0.148 ± 0.025	0.128 ± 0.017	0.136 ± 0.005	16.2%	0.160 ± 0.018	0.138 ± 0.018	0.147 ± 0.005	23.5%
DESCN	0.275 ± 0.076	0.054 ± 0.023	0.142 ± 0.019	21.4%	0.318 ± 0.053	0.060 ± 0.020	0.164 ± 0.014	37.8%
w/o $L'_\theta(f,\tau)$	0.164 ± 0.030	0.117 ± 0.019	0.136 ± 0.011	16.2%	0.181 ± 0.029	0.138 ± 0.021	0.155 ± 0.006	30.3%
w/o $L'_\theta(\mu_{-1},\mu_{+1})$	0.150 ± 0.023	0.137 ± 0.026	0.142 ± 0.016	21.4%	0.159 ± 0.025	0.145 ± 0.028	0.151 ± 0.016	26.9%
CTRL	0.143 ± 0.017	0.123 ± 0.007	0.131 ± 0.007	12.0%	0.156 ± 0.018	0.126 ± 0.010	0.138 ± 0.004	16.0%
Oracle	0.164 ± 0.008	0.086 ± 0.007	0.117 ± 0.004	-	0.167 ± 0.005	0.087 ± 0.004	0.119 ± 0.003	-

6 Empirical Experiments

6.1 Experiment Setup

Baselines. We evaluate the proposed **CTRL** method under the causal classification task. We compare our method with meta-learners (**S-Learner**, **T-Learner**, **X-Learner**, **R-Learner**, and **DR-Learner**), and the following methods: (1) Causal Forest methods, i.e., **CF** [1], **DRORF** [27] and **DMLORF** [4], which use decision trees to estimate treatment effects from observational data; (2) **CFRNet** [17, 30], **DeRCFR** [39], and **ESCFR** [37], which learns balanced representations for counterfactual regression; (3) **CEVAE** [24], which employs variational inference to generate potential outcomes; (4) **DragonNet** [31], which uses an adaptive neural network learn propensity scores and counterfactual outcomes in an end-to-end way; (5) **DESCN** [42] uses deep networks to model treatment effects in the entire population space. Meanwhile, we consider two ablation versions of the CTRL method as baselines: (1) a two-phase learning paradigm that first pre-trains $\tau(X)$ using CFR loss and then trains the classifier $f(\cdot)$ using $L'_\theta(f,\tau)$, denoted as **w/o** $L'_\theta(\mu_{-1},\mu_{+1})$; (2) Classifying using $\hat{\tau}$ learned by $L'_\theta(\mu_{-1},\mu_{+1})$ and CFR loss, denoted as **w/o** $L'_\theta(f,\tau)$. Additionally, we consider the method that uses the ground truth probability of being a responder $\mathbb{P}(R=1)$ to make classification as **Oracle**.

Evaluation Metrics. We utilize three metrics for evaluation. First, we use the False Positive Rate (FPR) to reflect the proportion of negative instances incorrectly predicted as positive, and the False Negative Rate (FNR) to indicate the proportion of positive instances which the model fails to identify. Additionally, we calculate the weighted risk based on the weight parameter θ, i.e., $\mathbf{Risk}(\boldsymbol{\theta}) = \mathrm{FPR}*\theta+\mathrm{FNR}*(1-\theta)$ to evaluate the overall quality of a classification method and report the risk relative error (Rel. Err.) to evaluate the difference between the method and the **Oracle** one. The Rel. Err. is calculate as follows: $|\mathrm{Risk}_{\mathrm{method}} - \mathrm{Risk}_{\mathrm{oracle}}|/\mathrm{Risk}_{\mathrm{oracle}}$.

6.2 Synthetic Experiment

Simulation Procedure. In the simulations (Toys), we first generate d_X-covariates from $X \sim \mathcal{N}(0_{d_x}, \Sigma_{d_x})$, where 0_{d_x} represents m_X-dimensional zero vectors, and Σ_{d_x} is a covariance matrix with diagonal elements equal to 1 and off-diagonal elements (covariances) equal to 0.3. The observed treatments are sampled from a Bernoulli distribution where with probability $\mathbb{P}(T \mid X)$, $T = +1$, and with probability $1 - \mathbb{P}(T \mid X)$, $T = -1$, where $\mathbb{P}(T \mid X) = \mathrm{sigmoid}(\sum_{i=1}^{d_x} \psi_i^T [X_i + (X_i + X_{i-1})^2/d_x])$ with $\psi_i^T \sim \mathrm{Unif}(-1, 1)$. Then, we generate the potential outcomes $Y(-1)$ and $Y(+1)$ from the joint probability $\mathbb{P}(Y(-1) = -1, Y(+1) = +1) = p^1(X) * p^2(X)$, $\mathbb{P}(Y(-1) = +1, Y(+1) = -1) = p^1(X) * (1 - p^2(X))$, $\mathbb{P}(Y(-1) = +1, Y(+1) = +1) = (1 - p^1(X)) * p^3(X)$, $\mathbb{P}(Y(-1) = -1, Y(+1) = -1) = (1-p^1(X))*(1-p^3(X))$, where $p^j(X) = \mathrm{sigmoid}(\sum_{i=1}^{d_x} \phi_i^j [X_i + (X_i + X_{i-1})^2/d_x] + a^j)$ with $\phi_i^j \sim \mathrm{Unif}(-1, 1)$ and default $a^1 = 0.5$, $a^2 = 2.4$, $a^3 = 0.0$. Notably, we can change Type-3 non-responder ratios $\mathbb{P}(Y(-1) = +1, Y(+1) = -1)$ by varying $a^2 \in \{0.0, 0.8, 1.5, 2.4, 3.5\}$. In this paper, we use $a^2 = 2.4$ to generate 10% Type-3 non-responder. Therefore, observed outcomes correspond to $Y(T)$ for the observed T, and the responders are from $\mathbb{P}(R = 1) = p^1(X) * p^2(X)$. Then, we generate 2000/500/5000 samples as training, validation, and testing data. We conduct 10 replications and report the mean and standard deviation of results.

Performance Comparison. To verify the validity of our method, we conducted experiments for weight $\theta \in \{0.4, 0.5, 0.6\}$ and the results are shown in Tables 2 and 5. First, our method stably outperforms CFRNet in both within-sample and out-of-sample scenarios, which proves the superiority of our method. Second, we find that the two-stage learning method w/o $L'_\theta(\mu_{-1},\mu_{+1})$ stably outperforms CFRNet. The w/o $L'_\theta(\mu_{-1},\mu_{+1})$ method first uses CFRNet loss to pretrain the causal effect estimation model to get the same $\hat{\tau}$ as CFRNet and frozen it, and then use $L'_\theta(f,\tau)$ loss to train a classifier $f(\cdot)$

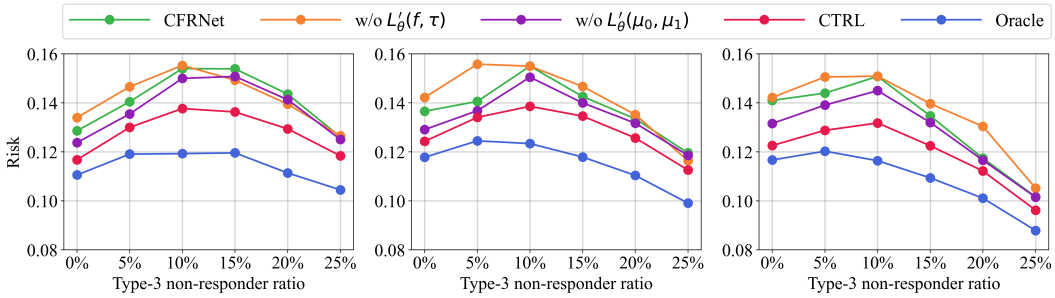

Figure 1: The Risk performance under varying degree of the non-monotonicity with $\theta = 0.4, 0.5$ and 0.6, respectively.

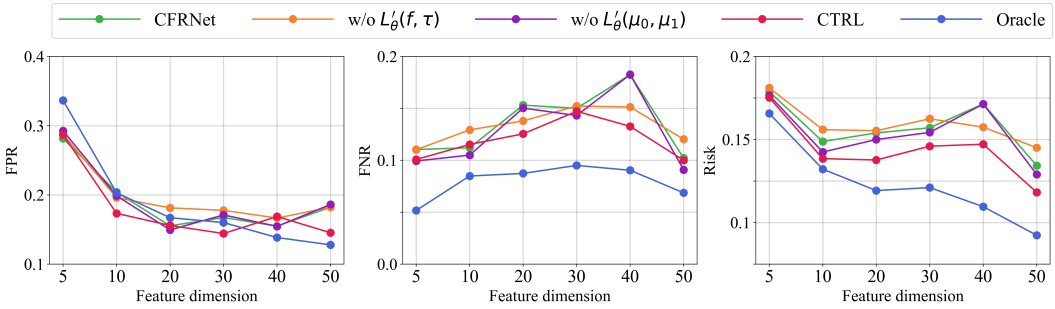

Figure 2: The classification performance with varying feature dimension under the weight $\theta = 0.4$.

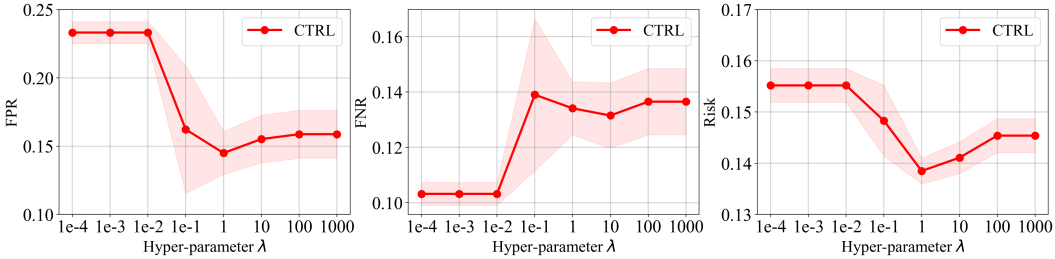

Figure 3: The classification performance under varying hyper-parameter λ with weight $\theta = 0.4$.

to make classification. The competitive performance can be attributed to the validity of the proposed $L'_\theta(f,\tau)$ loss. Meanwhile, w/o $L'_\theta(f,\tau)$ also stably outperforms CFRNet in both scenario. which is a two-stage learning using CFR loss and D loss for causal effect estimation model training and then using $\hat{\tau}$ for classification, which empirically provides that $L'_\theta(\mu_{-1},\mu_{+1})$ loss helps in causal classification task. See Appendix for more results with varying θ.

In-Depth Analysis. To verify the effectiveness of our method, we conduct experiments with different degrees of non-monotonicity, sample sizes and data complexity with different weight θ on synthetic data. Figure 1 shows the Risk performance under varying degree of the non-monotonicity. First, our method is the closest one to oracle in terms of Risk, which further verifies the effectiveness of our method. Second, as the type-3 non-responder ratio increases, the Risk of all methods has a tendency to increase first and then decrease, which is because a small amount of type-3 non-responder will make the causal effect estimation becoming smaller, leading to some positive sample on the classification boundary cannot be recognized. Meanwhile, when there exists a large amount of type-3 non-responder, since the causal effect estimate is already small at this time, some negative samples will be difficult to be misclassified as positive sample, resulting in a smaller FNR and a lower Risk. Additionally, we find that as weight θ increases, i.e., the weight of FNR increases, Risk decreases more significantly in the presence of a large number of type-3 non-responder, which further validates the rationality of this phenomenon. Figure 2 shows the classification performance with varying dimensions. Our method stably outperforms the baselines, which shows the robustness of our method. See Appendix for the results across varying sample sizes.

Sensitivity Analysis. The hyperparameter λ before the CFR loss $\mathcal{L}_{CFR}$ plays a very important role in the training phase. In order to explore under which lambda our method has the best classification performance, we constructed sensitivity analysis experiments and the results are shown in Figure 3. As λ increases, the Risk of our method first decreases and then increases. This is because when λ is small, the causal effect estimation $\hat{\tau}$ is not learned accurately due to the low weight of the CFR loss, which makes the coefficients of f in $L'_\theta(f,\tau)$ inaccurate, leading to learning a sub-optimal classifier.

Table 3: Results (mean±std) of FPR, FNR, Risk and Rel. Err. on IHDP/Jobs Dataset with $\theta = 0.4$

	Within-samples on IHDP with $\theta = 0.4$				**Out-of-samples on IHDP with $\theta = 0.4$**			
	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓
S_learner	0.282 ± 0.023	0.219 ± 0.022	0.244 ± 0.004	130%	0.292 ± 0.051	0.210 ± 0.049	0.243 ± 0.010	111%
T_learner	0.151 ± 0.045	0.084 ± 0.024	0.111 ± 0.017	4.7%	0.195 ± 0.062	0.117 ± 0.048	0.148 ± 0.030	28.7%
X_learner	0.169 ± 0.046	0.101 ± 0.024	0.128 ± 0.014	20.8%	0.204 ± 0.066	0.117 ± 0.041	0.152 ± 0.023	32.2%
R_learner	0.255 ± 0.041	0.095 ± 0.041	0.159 ± 0.020	50.0%	0.299 ± 0.082	0.108 ± 0.026	0.184 ± 0.029	60.0%
DR_learner	0.167 ± 0.049	0.092 ± 0.019	0.122 ± 0.018	15.1%	0.212 ± 0.074	0.122 ± 0.039	0.158 ± 0.024	37.4%
CF	0.156 ± 0.273	0.262 ± 0.144	0.219 ± 0.024	106.6%	0.151 ± 0.267	0.237 ± 0.129	0.203 ± 0.038	76.5%
DRORF	0.276 ± 0.031	0.223 ± 0.028	0.244 ± 0.004	130.2%	0.287 ± 0.041	0.212 ± 0.042	0.242 ± 0.011	110.4%
DMLORF	0.290 ± 0.031	0.211 ± 0.028	0.243 ± 0.004	129.2%	0.306 ± 0.049	0.196 ± 0.048	0.240 ± 0.010	108.7%
CFRNet	0.198 ± 0.041	0.098 ± 0.032	0.138 ± 0.017	30.1%	0.243 ± 0.056	0.123 ± 0.027	0.171 ± 0.026	48.6%
DeRCFR	0.393 ± 0.163	0.029 ± 0.038	0.175 ± 0.043	65.1%	0.284 ± 0.223	0.113 ± 0.154	0.182 ± 0.046	58.3%
ESCFR	0.327 ± 0.071	0.024 ± 0.021	0.145 ± 0.017	36.8%	0.368 ± 0.116	0.037 ± 0.034	0.170 ± 0.031	47.8%
CEVAE	0.375 ± 0.142	0.109 ± 0.118	0.215 ± 0.016	102.8%	0.257 ± 0.184	0.180 ± 0.165	0.211 ± 0.034	83.5%
DragonNet	0.205 ± 0.067	0.087 ± 0.029	0.134 ± 0.018	26.4%	0.216 ± 0.079	0.096 ± 0.051	0.144 ± 0.020	25.2%
DESCN	0.311 ± 0.073	0.044 ± 0.032	0.151 ± 0.018	42.5%	0.299 ± 0.107	0.075 ± 0.058	0.165 ± 0.023	43.5%
w/o $L'_\theta(f,\tau)$	0.171 ± 0.048	0.099 ± 0.028	0.128 ± 0.018	20.8%	0.230 ± 0.055	0.137 ± 0.060	0.174 ± 0.026	51.3%
w/o $L'_\theta(\mu_{-1},\mu_{+1})$	0.186 ± 0.039	0.089 ± 0.032	0.128 ± 0.017	20.8%	0.226 ± 0.066	0.107 ± 0.031	0.154 ± 0.034	33.9%
CTRL	0.213 ± 0.083	0.070 ± 0.040	0.127 ± 0.016	19.8%	0.241 ± 0.090	0.065 ± 0.057	0.135 ± 0.020	17.4%
Oracle	0.104 ± 0.009	0.119 ± 0.010	0.106 ± 0.004	-	0.119 ± 0.031	0.122 ± 0.019	0.115 ± 0.017	-
	Within-samples on Jobs with 0.4				**Out-of-samples on Jobs with 0.4**			
	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓
S_learner	0.296 ± 0.015	0.168 ± 0.034	0.220 ± 0.021	60.0%	0.295 ± 0.023	0.171 ± 0.031	0.221 ± 0.018	123%
T_learner	0.175 ± 0.033	0.068 ± 0.027	0.111 ± 0.021	5.7%	0.180 ± 0.041	0.076 ± 0.026	0.118 ± 0.028	19.2%
X_learner	0.176 ± 0.033	0.068 ± 0.027	0.112 ± 0.023	6.7%	0.182 ± 0.040	0.073 ± 0.031	0.116 ± 0.029	17.2%
R_learner	0.248 ± 0.062	0.053 ± 0.016	0.131 ± 0.022	24.7%	0.262 ± 0.057	0.067 ± 0.017	0.144 ± 0.023	45.5%
DR_learner	0.164 ± 0.026	0.079 ± 0.024	0.113 ± 0.019	7.6%	0.194 ± 0.034	0.117 ± 0.037	0.148 ± 0.030	49.5%
CF	0.129 ± 0.273	0.270 ± 0.145	0.214 ± 0.028	103.8%	0.147 ± 0.311	0.271 ± 0.148	0.222 ± 0.045	124.2%
DRORF	0.296 ± 0.017	0.167 ± 0.036	0.218 ± 0.020	107.6%	0.295 ± 0.021	0.169 ± 0.032	0.219 ± 0.018	121.2%
DMLORF	0.297 ± 0.015	0.165 ± 0.035	0.218 ± 0.020	107.6%	0.298 ± 0.022	0.166 ± 0.032	0.219 ± 0.018	121.2%
CFRNet	0.196 ± 0.043	0.072 ± 0.022	0.121 ± 0.023	15.2%	0.192 ± 0.055	0.072 ± 0.024	0.120 ± 0.028	21.2%
DeRCFR	0.228 ± 0.108	0.050 ± 0.037	0.121 ± 0.033	15.2%	0.230 ± 0.132	0.053 ± 0.041	0.124 ± 0.043	25.3%
ESCFR	0.266 ± 0.104	0.038 ± 0.025	0.129 ± 0.035	22.9%	0.282 ± 0.098	0.034 ± 0.018	0.133 ± 0.034	34.3%
CEVAE	0.227 ± 0.143	0.131 ± 0.199	0.170 ± 0.075	61.9%	0.231 ± 0.146	0.136 ± 0.211	0.174 ± 0.081	75.8%
DragonNet	0.182 ± 0.045	0.064 ± 0.032	0.111 ± 0.022	5.7%	0.183 ± 0.036	0.064 ± 0.033	0.112 ± 0.025	13.1%
DESCN	0.356 ± 0.075	0.009 ± 0.007	0.148 ± 0.029	41.0%	0.354 ± 0.081	0.006 ± 0.005	0.145 ± 0.032	46.5%
w/o $L'_\theta(f,\tau)$	0.189 ± 0.038	0.070 ± 0.019	0.117 ± 0.021	11.4%	0.180 ± 0.052	0.080 ± 0.029	0.120 ± 0.026	21.2%
w/o $L'_\theta(\mu_{-1},\mu_{+1})$	0.182 ± 0.048	0.066 ± 0.023	0.112 ± 0.024	6.7%	0.178 ± 0.054	0.059 ± 0.026	0.110 ± 0.026	11.1%
CTRL	0.196 ± 0.060	0.055 ± 0.030	0.111 ± 0.026	5.7%	0.188 ± 0.073	0.053 ± 0.036	0.107 ± 0.029	8.1%
Oracle	0.172 ± 0.045	0.060 ± 0.021	0.105 ± 0.022	-	0.168 ± 0.054	0.052 ± 0.025	0.099 ± 0.024	-

When λ is large, although an accurate causal effect estimation model can be learned, the low weight of $L'_\theta(f,\tau)$ prevents the classifier f from being adequately trained. To address this, we select the hyperparameter λ by minimizing the bound on the classification risk using validation data. In our experiments, we set $\lambda = 1$.

6.3 Real-World Experiment

Dataset and Prepossessing. Following previous studies [24, 30, 40], we conduct experiments on one semi-synthetic dataset, **IHDP**, and one real-world dataset, **Jobs**. The IHDP dataset [13] is based on the Infant Health and Development Program (IHDP), which is aim to examine the effects of specialist home visits on future cognitive test scores. The dataset includes 747 units and 25 covariates measuring aspects of children and their mothers. The **Jobs** dataset [21] is based on the National Supported Work program to examine the effects of job training on income and employment status after training, which includes 3,212 units and 17 covariates. We follow Shalit et al. [30] to split the data into training/validation/testing set with ratios 63/27/10 for both the **IHDP** and **Jobs** datasets. We repeat experiment 100 times for the **IHDP** and 10 times for **Jobs**.

We preprocessed the **IHDP** dataset as follows to obtain binary outcomes: we treat the original covariates and treatments as X and T, use the original potential outcomes as M_0 and M_1. Then, define $p_M^0 = \text{sigmoid}((M_0 - \mathbb{E}M_0) * 2)$ and $p_M^1 = \text{sigmoid}((M_1 - \mathbb{E}M_1) * 4)$, we sample $Y(-1)$ and $Y(+1)$ from joint probability $\mathbb{P}(Y(-1) = -1, Y(+1) = +1) = p_M^1$, $\mathbb{P}(Y(-1) = +1, Y(+1) = +1) = (1 - p_M^1) * p_M^0$, $\mathbb{P}(Y(-1) = -1, Y(+1) = -1) = (1 - p_M^1) * (1 - p_M^0)$. In addition, since the **Jobs** dataset contains only observed outcomes and not counterfactual outcomes, we only collect 17 covariates from the original **Jobs** dataset and then simulate the observed treatment

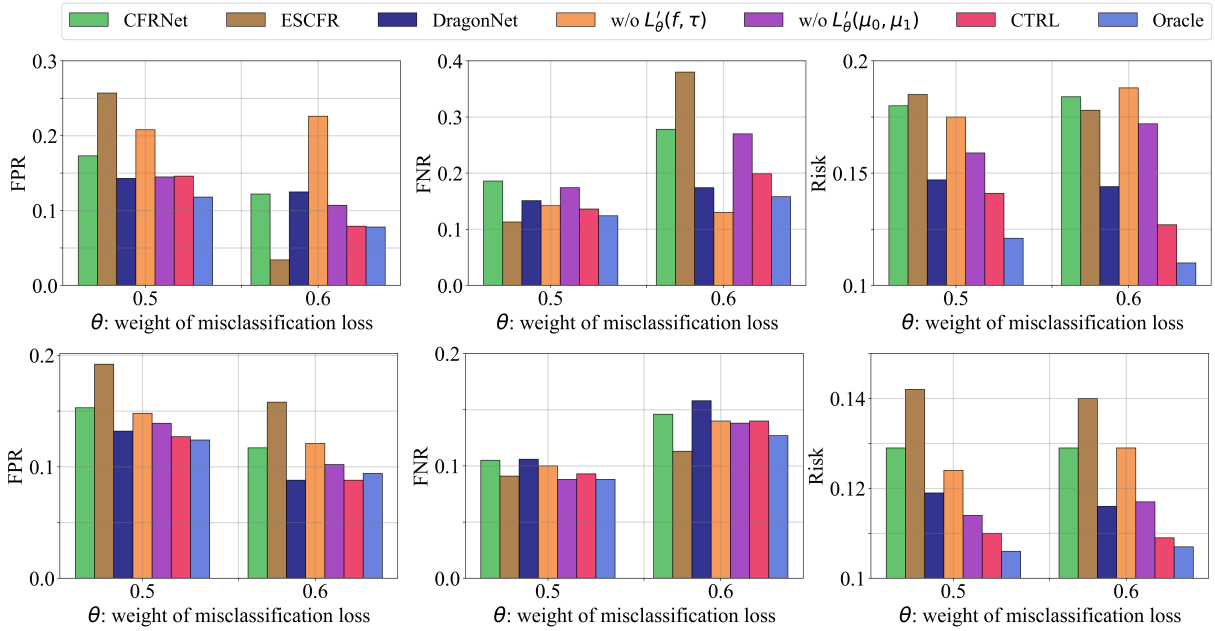

Figure 4: The out-of-sample results for IHDP dataset (above) and the Jobs dataset (below) with $\theta \in \{0.5, 0.6\}$.

and outcomes using the same approach as in the simulation, with $p^2(X) = 0$ kept constant (implying positive returns from training).

Performance Comparison. Tables 3 show the performance of our method and the baseline for causal classification task on the **IHDP** dataset and the **Jobs** dataset with $\theta = 0.4$. First, DragonNet achieves the most competitive performance across the baseline. Second, our method stably outperforms the baseline method in both within sample scenarios and out-of sample scenarios, which further validates the effectiveness of our method. Third, method w/o $L'_\theta(\mu_{-1}, \mu_{+1})$ and method w/o $L'_\theta(f, \tau)$ still stably outperform the CFRNet method, which validates the effectiveness of two proposed losses. The results presented in Tables 2 and 3 highlight the significant advantages of our proposed algorithm: (1) In the Toy data simulation with $\theta = 0.4$, as shown in Table 2, our CTRL method achieved reductions of 14.3% in Risk and 19.2% in Relative Error compared to the traditional causal effect estimator CFRNet, demonstrating superior out-of-sample performance. (2) In the IHDP data with $\theta = 0.4$, as shown in Table 3, our CTRL method reduced Risk by 6.2% and Relative Error by 7.8% compared to the best baseline, DragonNet, highlighting improved out-of-sample performance. (3) In the Jobs data with $\theta = 0.4$, as shown in Table 4, our CTRL method achieved reductions of 4.5% in Risk and 5.0% in Relative Error compared to DragonNet, showing significant out-of-sample performance improvements. These improvements are substantial in real-world applications, especially in high-risk projects such as government financial policy or healthcare policy. Figure 4 shows the performance of our method and the baseline method under different weights θ, and similar results can be found that the risk of our method remains the lowest and is still the closest to the Oracle.

6.4 Ablation Studies

We conducted comprehensive ablation studies to evaluate each component of the CTRL estimator on Toy, IHDP, and Jobs datasets:

- Theorem 4.4 shows the classification risk is bounded by the surrogate risk $L'_\theta(f, \tau)$ and weighted risk $L'_\theta(\mu_{-1}, \mu_{+1})$. We incorporate them and CFR loss into the objective function.
- In our experiments, CFRNet is the foundation model of CTRL. 'w/o $L'_\theta(f, \tau)$' denotes CTRL without surrogate risk, and 'w/o $L'_\theta(\mu_{-1}, \mu_{+1})$' denotes CTRL without weighted risk. These represent three ablation versions of our CTRL algorithm.

The results in Tables 2 & 3 and Figures 1, 2 & 4 demonstrate that each component significantly reduces the weighted risk.

7 Conclusion

This paper studies the treatment responder classification problem without assuming the monotonicity assumption. First, we derive the sharp bounds of the probability that an individual is a responder, as well as a sharp upper bound on the weighted classification risk to measure the worst classification performance. Next, we theoretically prove that the classification risk of the classifier learned by minimizing the proposed sharp upper bound converges to the optimal classification risk via generalization theory. Then, a joint learning algorithm for classifying the treatment responders is proposed, named CTLR, and demonstrate the superiority of jointly learning over two-stage learning. Finally, we conduct extensive experiments on two real-world datasets and one semi-synthetic dataset, and the results demonstrate the superiority of our joint learning approach in classifying the treatment responders.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (623B2002, 62376243, 62441605, 62037001), and the Starry Night Science Fund at Shanghai Institute for Advanced Study (Zhejiang University).

References

[1] Susan Athey and Stefan Wager. 2019. Estimating treatment effects with causal forests: An application. *Observational studies* 5, 2 (2019), 37–51.

[2] S. Athey and S. Wager. 2021. Policy Learning With Observational Data. *Econometrica* 89, 1 (2021), 133–161.

[3] H. Bang and J. M. Robins. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics* 61, 4 (2005), 962–73.

[4] Keith Battocchi, Eleanor Dillon, Maggie Hei, Greg Lewis, Paul Oka, Miruna Oprescu, and Vasilis Syrgkanis. 2019. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/py-why/EconML. Version 0.x.

[5] Dimitris Bertsimas, Nathan Kallus, Alexander M. Weinstein, and Ying Daisy Zhuo. 2016. Personalized Diabetes Management Using Electronic Medical Records. *Diabetes Care* 40, 2 (12 2016), 210–217.

[6] R. H. Dehejia and S. Wahba. 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* 84, 1 (2002), 151–161.

[7] Miroslav Dudík, John Langford, and Lihong Li. 2011. Doubly robust policy evaluation and learning. In *International Conference on Machine Learning*. PMLR, 1097–1104.

[8] Carlos Fernández-Loría and Foster Provost. 2022. Causal classification: Treatment effect estimation vs. outcome prediction. *The Journal of Machine Learning Research* 23, 1 (2022), 2573–2607.

[9] Carlos Fernández-Loría and Foster Provost. 2022. Causal decision making and causal effect estimation are not the same… and why it matters. *INFORMS Journal on Data Science* 1, 1 (2022), 4–16.

[10] Carlos Fernández-Loría and Jorge Loría. 2023. Causal Scoring: A Framework for Effect Estimation, Effect Ordering, and Effect Classification. arXiv:2206.12532

[11] Negar Hassanpour and Russell Greiner. 2020. Learning Disentangled Representations for CounterFactual Regression. In *International Conference on Learning Representations*.

[12] M.A. Hernán and J. M. Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC.

[13] Jennifer L Hill. 2011. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics* 20, 1 (2011), 217–240.

[14] K. Hirano, G. W. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica: Journal of the econometric society* 71, 4 (2003), 1161–1189.

[15] Paul W. Holland. 1986. Statistics and Causal Inference. *J. Amer. Statist. Assoc.* 81 (1986), 945–960.

[16] G. W. Imbens and D. B. Rubin. 2015. *Causal Inference For Statistics Social and Biomedical Science*. Cambridge University Press.

[17] Fredrik Johansson, Uri Shalit, and David Sontag. 2016. Learning representations for counterfactual inference. In *International conference on machine learning*. PMLR, 3020–3029.

[18] Nathan Kallus. 2019. Classifying treatment responders under causal effect monotonicity. In *International Conference on Machine Learning*. PMLR, 3201–3210.

[19] Nathan Kallus. 2022. What's the harm? sharp bounds on the fraction negatively affected by treatment. *Advances in Neural Information Processing Systems* 35 (2022), 15996–16009.

[20] T. Kitagawa and A. Tetenov. 2018. Who should be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica* 86, 2 (2018).

[21] Robert J LaLonde. 1986. Evaluating the econometric evaluations of training programs with experimental data. *The American economic review* (1986), 604–620.

[22] John Langford and Tong Zhang. 2007. The epoch-greedy algorithm for contextual multi-armed bandits. *Advances in neural information processing systems* 20, 1 (2007), 96–1.

[23] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. 2017. Causal effect inference with deep latent-variable models. *Advances in Neural Information Processing Systems* 30 (2017).

[24] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. 2017. Causal effect inference with deep latent-variable models. *Advances in Neural Information Processing Systems* 30 (2017).

[25] Stephen L. Morgan and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (second ed.). Cambridge University Press.

[26] Jerzy Splawa Neyman. 1990. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* 5 (1990), 465–472.

[27] Miruna Oprescu, Vasilis Syrgkanis, and Zhiwei Steven Wu. 2019. Orthogonal random forest for causal inference. In *International Conference on Machine Learning*. PMLR, 4932–4941.

[28] D. B. Rubin. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational psychology* 66 (1974), 688–701.

[29] Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

[30] Uri Shalit, Fredrik D Johansson, and David Sontag. 2017. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*. PMLR, 3076–3085.

[31] Claudia Shi, David Blei, and Victor Veitch. 2019. Adapting neural networks for the estimation of treatment effects. *Advances in Neural Information Processing Systems* 32 (2019).

[32] Chengchun Shi, Xiaoyu Wang, Shikai Luo, Hongtu Zhu, Jieping Ye, and Rui Song. 2023. Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework. *J. Amer. Statist. Assoc.* 118, 543 (2023), 2059–2071.

[33] Yi Su, Maria Dimakopoulou, Akshay Krishnamurthy, and Miroslav Dudík. 2020. Doubly robust off-policy evaluation with shrinkage. In *International Conference on Machine Learning*. PMLR, 9167–9176.

[34] Adith Swaminathan and Thorsten Joachims. 2015. Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research* 16, 1 (2015), 1731–1755.

[35] Adith Swaminathan and Thorsten Joachims. 2015. The self-normalized estimator for counterfactual learning. *Advances in neural information processing systems* 28 (2015).

[36] Stefan Wager and Susan Athey. 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *J. Amer. Statist. Assoc.* 113, 523 (2018), 1228–1242.

[37] Hao Wang, Jiajun Fan, Zhichao Chen, Haoxuan Li, Weiming Liu, Tianqiao Liu, Quanyu Dai, Yichao Wang, Zhenhua Dong, and Ruiming Tang. 2023. Optimal Transport for Treatment Effect Estimation. *Advances in Neural Information Processing Systems* (2023).

[38] Yu-Xiang Wang, Alekh Agarwal, and Miroslav Dudık. 2017. Optimal and adaptive off-policy evaluation in contextual bandits. In *International Conference on Machine Learning*. PMLR, 3589–3597.

[39] Anpeng Wu, Junkun Yuan, Kun Kuang, Bo Li, Runze Wu, Qiang Zhu, Yueting Zhuang, and Fei Wu. 2022. Learning decomposed representations for treatment effect estimation. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2022), 4989–5001.

[40] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2018. GANITE: Estimation of Individualized Treatment Effects using Generative Adversarial Nets. In *International Conference on Learning Representations*.

[41] Weijia Zhang, Lin Liu, and Jiuyong Li. 2021. Treatment effect estimation with disentangled latent factors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35.

[42] Kailiang Zhong, Fengtong Xiao, Yan Ren, Yaorong Liang, Wenqing Yao, Xiaofeng Yang, and Ling Cen. 2022. DESCN: Deep Entire Space Cross Networks for Individual Treatment Effect Estimation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4612–4620.

Appendix

A Notations

In this section, we provide a summary of the key notation and their descriptions, as shown in Table 4.

Table 4: Key notations used in this paper.

Notation	Description
X	Observed Covariate
T	Observed Treatment
Y	Observed Outcome
$Y(-1)$	Potential Outcome under $T = -1$
$Y(+1)$	Potential Outcome under $T = +1$
θ	Risk Cost
R	$R = +1$: Responder; $R = -1$: Non-Responder.
$\rho(X) = \mathbb{P}(R = +1 \mid X)$	The Probability of Becoming Responder
$A = (Y(+1) - Y(-1))/2$	Real Treatment Effect
$\tau(X) = \mathbb{E}[Y(+1) - Y(-1) \mid X]$	Conditional Average Treatment Effect
$L_\theta(f)$	The Weighted Misclassification Loss
$L'_\theta(f, \tau)$	Surrogate Risk
$L'_\theta(\mu_{-1}, \mu_{+1})$	Weighted Risk

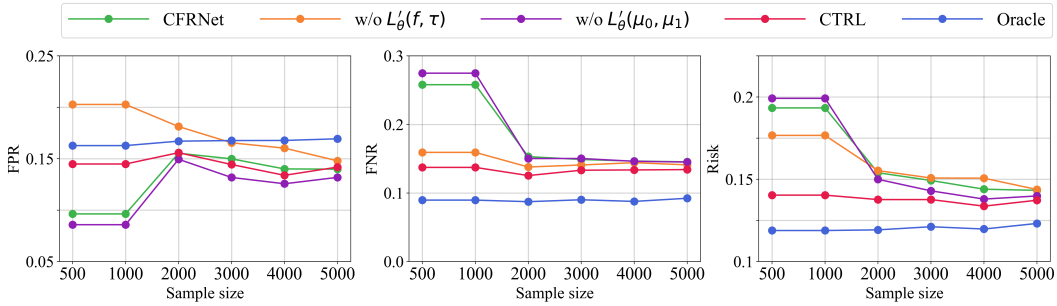

Figure 5: The classification performance with varying sample sizes on Toy dataset with $\theta = 0.4$.

Table 5: Results (mean±std) of FPR, FNR, Risk and Rel. Err. on Toy dataset with $\theta \in \{0.5, 0.6\}$.

	Out-of-samples on Toy with $\theta = 0.5$				Out-of-samples on Toy with $\theta = 0.6$			
	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓	FPR ↓	FNR ↓	Risk ↓	Rel. Err. ↓
S-learner	0.159 ± 0.025	0.295 ± 0.017	0.227 ± 0.004	84.5%	0.101 ± 0.022	0.334 ± 0.015	0.194 ± 0.007	67.2%
T-learner	0.145 ± 0.014	0.198 ± 0.014	0.171 ± 0.004	39.0%	0.107 ± 0.010	0.232 ± 0.013	0.157 ± 0.004	35.3%
X-learner	0.137 ± 0.017	0.195 ± 0.017	0.166 ± 0.005	35.0%	0.099 ± 0.014	0.231 ± 0.017	0.152 ± 0.006	31.0%
DR-learner	0.132 ± 0.017	0.210 ± 0.020	0.171 ± 0.004	39.0%	0.095 ± 0.013	0.252 ± 0.021	0.157 ± 0.004	35.3%
R-learner	0.085 ± 0.058	0.287 ± 0.080	0.186 ± 0.014	51.2%	0.034 ± 0.044	0.346 ± 0.067	0.159 ± 0.005	37.1%
CF	0.083 ± 0.021	0.346 ± 0.016	0.214 ± 0.004	74.0%	0.017 ± 0.010	0.390 ± 0.010	0.166 ± 0.003	43.1%
DRORF	0.133 ± 0.017	0.312 ± 0.011	0.222 ± 0.004	80.5%	0.079 ± 0.013	0.348 ± 0.009	0.187 ± 0.005	61.2%
DMLORF	0.146 ± 0.022	0.303 ± 0.015	0.225 ± 0.004	82.9%	0.090 ± 0.018	0.341 ± 0.012	0.190 ± 0.007	63.8%
CFRNet	0.120 ± 0.028	0.208 ± 0.028	0.164 ± 0.010	33.3%	0.070 ± 0.032	0.285 ± 0.063	0.156 ± 0.010	34.4%
DeRCFR	0.115 ± 0.048	0.185 ± 0.066	0.150 ± 0.016	21.9%	0.053 ± 0.026	0.264 ± 0.054	0.137 ± 0.010	18.1%
ESCFR	0.000 ± 0.000	0.401 ± 0.006	0.201 ± 0.003	63.4%	0.000 ± 0.000	0.401 ± 0.006	0.160 ± 0.002	37.9%
CEVAE	0.133 ± 0.141	0.261 ± 0.145	0.197 ± 0.005	60.1%	0.000 ± 0.000	0.401 ± 0.006	0.160 ± 0.002	37.9%
DragonNet	0.107 ± 0.020	0.189 ± 0.024	0.148 ± 0.006	20.3%	0.061 ± 0.014	0.251 ± 0.022	0.137 ± 0.005	18.1%
DESCN	0.159 ± 0.140	0.207 ± 0.163	0.183 ± 0.016	48.8%	0.000 ± 0.000	0.401 ± 0.006	0.160 ± 0.002	37.9%
w/o $L'_\theta(f,\tau)$	0.134 ± 0.028	0.176 ± 0.023	0.155 ± 0.004	26.0%	0.124 ± 0.025	0.191 ± 0.022	0.151 ± 0.008	30.2%
w/o $L'_\theta(\mu_{-1},\mu_{+1})$	0.102 ± 0.029	0.206 ± 0.025	0.154 ± 0.012	25.2%	0.041 ± 0.031	0.301 ± 0.073	0.145 ± 0.013	25.0%
CTRL	0.124 ± 0.009	0.154 ± 0.012	0.139 ± 0.004	13.0%	0.093 ± 0.017	0.191 ± 0.023	0.132 ± 0.004	13.8%
Oracle	0.114 ± 0.005	0.133 ± 0.003	0.123 ± 0.004	-	0.071 ± 0.004	0.185 ± 0.004	0.116 ± 0.003	-

B Supplementary Experiments

B.1 Scalability to Small Sample Sizes

Figure 5 illustrates the responder classification performance on the Toy dataset with varying sample sizes under the cost $\theta = 0.4$. The results demonstrate that CTRL achieves stable and reliable performance even with small sample sizes, highlighting its ability to effectively leverage limited data. Specifically, at sample sizes of 500 and 1000, CTRL exhibits significantly lower risk compared to CFRNet, showcasing its superiority and robustness.

B.2 Supplementary experiments on Toy dataset across $\theta \in \{0.5, 0.6\}$

We conduct additional experiments on the synthetic Toy dataset with cost weights $\theta \in \{0.5, 0.6\}$, as listed in Table 5. The results demonstrate that CTRL consistently outperforms baseline methods, achieving lower risk and relative error across both θ values. Notably, CTRL's performance is comparable to the Oracle model, underscoring its robustness, effectiveness, and ability to generalize well under varying cost weights.

B.3 Computational Complexity Analysis

The experimental results demonstrate that incorporating a classifier network and leveraging surrogate and weighted risks introduces a linear increase in computational complexity. As shown in Table 6, the CTRL model efficiently completes training on all three datasets—Toys, IHDP, and Jobs—within 30 seconds. This highlights its computational scalability while maintaining effective performance across diverse datasets.

C Theoretical Proof

C.1 Proof of Lemma 1

According to the definition of $\tau(X)$, we have:

$$\begin{aligned}\tau(X) &= \mathbb{E}(Y(+1) - Y(-1) \mid X) \\ &= \sum_{a,b\in\{-1,+1\}} \mathbb{P}(Y(-1) = a, Y(+1) = b \mid X) \cdot (b - a) \\ &= 2\mathbb{P}(Y(-1) = -1, Y(+1) = +1 \mid X) \\ &\quad - 2\mathbb{P}(Y(-1) = +1, Y(+1) = -1 \mid X) \\ &= 2\rho(X) - 2\mathbb{P}(A = -1 \mid X)\end{aligned}$$

Table 6: Training time(s) of various methods in a single execution on different datasets.

Model	Toys	IHDP	Jobs
CF	2.51	0.14	1.02
DRORF	102.46	35.14	113.07
DMLORF	72.63	18.2	71.69
CFRNet	3.21	2.84	3.20
DeRCFR	17.86	12.49	17.49
ESCFR	3.64	3.55	3.57
CEVAE	49.20	13.04	50.22
DragonNet	1.50	1.42	1.45
DESCN	2.06	2.04	1.98
CTRL	23.11	10.53	29.71

Rearranging the terms provides the desired formulation: $\rho(X) = \frac{\tau(X)}{2} + \mathbb{P}(A = -1 \mid X)$.

C.2 Proof of Lemma 2

Given $\rho(x) = \frac{\tau(X)}{2} + \mathbb{P}(A = -1 \mid X) \geq \frac{\tau(X)}{2}$ and $\rho(X) \geq 0$, we can establish the sharp lower bound: $\rho(X) \geq \max\{\frac{\tau(X)}{2}, 0\}$.
Given $P(Y(-1) = -1 \mid X) = \rho(X) + P(Y(-1) = Y(+1) = -1 \mid X) \geq \rho(X)$ and $P(Y(+1) = +1 \mid X) = \rho(X) + P(Y(-1) = Y(+1) = +1 \mid X) \geq \rho(X)$, we can establish the sharp upper bound:$\rho(x) \leq \min\{1 - \mathbb{P}(Y(-1) = +1 \mid X), \mathbb{P}(Y(+1) = +1 \mid X)\}$.

C.3 Proof of Lemma 3

By the law of iterated expectations, we have

$$
\begin{aligned}
&L_\theta(f) = \theta \cdot \mathbb{P}(\text{false positive}) + (1-\theta) \cdot \mathbb{P}(\text{false negative}) \\
&= \theta \cdot \mathbb{E}[\mathbb{I}[f(X) = +1]\mathbb{I}[R = -1]] + (1-\theta) \cdot \mathbb{E}[\mathbb{I}[f(X) = -1]\mathbb{I}[R = +1]] \\
&= \frac{\theta}{2} \cdot \mathbb{E}[(1 + f(X))(1 - \rho(X))] + \frac{1-\theta}{2} \cdot \mathbb{E}[(1 - f(X))\rho(X)] \\
&= \frac{1}{2} \cdot \mathbb{E}[f(X)(\theta(1-\rho(X)) - (1-\theta)\rho(X))] \\
&\quad + \frac{1}{2} \cdot \mathbb{E}[\theta(1-\rho(X)) + (1-\theta)\rho(X))] \\
&= \frac{1}{2}\mathbb{E}[f(X)(\theta - \rho(X))] + \frac{1}{2}\mathbb{E}[\theta + (1-2\theta)\rho(X)],
\end{aligned}
$$

which leads to the conclusion in Lemma 4.3.

C.4 Proof of Theorem 2

In our derivation, $L_\theta(f)$ is Lipschitz continuous with respect to $f(X)$ with Lipschitz constant $c = \max\{\theta, \theta - \max_X\{\tau(X)\}\}$. By Talagrand's lemma, the upper bound of Rademacher complexity is $\mathfrak{R}(L_\theta \circ \mathcal{F}) \leq c\mathfrak{R}(\mathcal{F})$. By Theorem 26.5(c) in Shalev-Shwartz and Ben-David [29], with probability at least $1 - \delta$, we have the result:

$$
\begin{aligned}
L_\theta(\hat{f}) \leq \min_{f \in \mathcal{F}} L_\theta(f) &+ 2\max\{\theta, \theta - \max_X\{\tau(X)\}\}\mathfrak{R}(\mathcal{F}) \\
&+ 5\max\{\theta, \theta - \max_X\{\tau(X)\}\}\sqrt{\frac{2\ln(8/\delta)}{n}}.
\end{aligned}
$$

C.5 Proof of Proposition 1

Two-phase learning: first, learn $\hat{\tau}$ by the existing CATE estimation method; then, learn f by minimizing the surrogate risk (1st)

$$
\hat{f}^{TP} = \arg\min_{f \in \mathcal{F}} \mathbb{E}[f(X)(\theta - \max\left\{\frac{\tau(X)}{2}, 0\right\})],
$$

Joint learning: learn f by minimizing both of the risks

$$
\begin{aligned}
&(\hat{f}^{JL}, \hat{\mu}^{JL}_{-1}, \hat{\mu}^{JL}_{+1}) = \arg\min_{f, \mu_{-1}, \mu_{+1}} \mathbb{E}[f(X)(\theta - \max\left\{\frac{\tau(X)}{2}, 0\right\})] \\
&+ \mathbb{I}(\theta > 1/2)L'_{\theta > 1/2}(\mu_{-1}, \mu_{+1}) + \mathbb{I}(\theta \leq 1/2)L'_{\theta \leq 1/2}(\mu_{-1}, \mu_{+1}).
\end{aligned}
$$

For brevity, we denote the classification risk in Joint learning as $L(\hat{f}, \hat{\mu}_{-1}, \hat{\mu}_{+1})$. The intuition of Proposition 1 is that for any given $\hat{\mu}_{-1}$ and $\hat{\mu}_{+1}$ (so as for given $\hat{\tau}$), we have $\min_{f, \mu_{-1}, \mu_{+1}} L(f, \mu_{-1}, \mu_{+1}) \leq \min_f L(f, \hat{\mu}_{-1}, \hat{\mu}_{+1})$, thus the classification risk of joint learning is not larger than that of two-phase learning.

D Generalize to Non-Binary Treatments

We now discuss how to generalize our results to non-binary treatments. For multi-value treatments T, we have the following observations in Table 7.

Table 7: The units are divided into $K + 1$ types according to $Y(T), T \in \{1, 2, \cdot, K\}$.

Unit Type	$Y(1)$	$Y(2)$	$\cdots$	$Y(K-1)$	$Y(K)$	A (RTE)
Type-1 responder	+1	+1	$\cdots$	+1	+1	1
Type-2 responder	0	+1	$\cdots$	+1	+1	2
·	·	·	$\cdots$	·	·	·
Type-K responder	0	0	$\cdots$	0	+1	K
Non-responder	0	0	$\cdots$	0	0	0

Consider a drug with dosages $\mathcal{T} = \{1, 2, \ldots, K\}$ and binary potential outcome $Y(t) \in \{0, 1\}$ for all $t \in \mathcal{T}$. We define an individual to be a type- k responder, if the individual can be cured given at least dosage k, i.e., $Y(t) = 0$ when $t \leq k-1$ and $Y(t) = 1$ when $t \geq k$ (here we implicitly assume the monotonicity assumption that $Y(t_1) \geq Y(t_2)$ for $t_1 \geq t_2$). Under the above monotonicity assumption, the probability of being all types treatment responders can be identified, i.e., $P(A = k \mid X) = P(Y(k) = 1 \mid X) - P(Y(k-1) = 1 \mid X)$ for all $k \geq 1$. Without the monotonicity assumption, we can similarly develop the Boole-Frechet-Hoeffding bounds on the joint probability of discrete events: $\max(0, P(A \text{ and } B \mid X) + P(C \mid X) - 1) \leq P(A \text{ and } B \text{ and } C \mid X) \leq \min(P(A \mid X), P(B \mid X), P(C \mid X))$, where the left-hand-side can be bounded by using the binary-case Boole-Frechet-Hoeffding bounds again.

Furthermore, for complex treatments, we can set a threshold C, e.g., $C = 0$, to convert real-valued outcomes into binary outcomes. Specifically, the outcomes greater than the threshold would be classified as +1, while those less than the threshold as -1. This allows us to reformulate the decision-making problem as a binary classification issue involving positive responders. Then we can use our CTRL to bound the probability of an individual is a positive responder, i.e., $P(Y(-1) < C, P(Y(+1) >= C))$ again.