Personalized Two-sided Dose Interval

Anonymous authors
Paper under double-blind review

Abstract

In fields such as medicine and social sciences, the goal of treatment is often to maintain the outcome of interest within a desirable range rather than to optimize its value. To achieve this, it may be more practical to recommend a treatment dose interval rather than a single fixed level for a study unit. Since individuals may respond differently to the same treatment level, the recommended dose interval should be personalized based on their unique characteristics. Existing methods for one-sided dose intervals and iteratively constructed two-sided intervals provide useful foundations, but their theory does not directly address simultaneous estimation over unrestricted product function spaces. To address this gap, we propose a direct method for learning personalized two-sided dose intervals based on empirical risk minimization with a doubly-robust loss function that is well-defined over a tensor product function space. This formulation enables simultaneous estimation of the lower and upper bounds without constrained alternating updates. We establish statistical properties of the estimated dose interval in terms of excess risk by leveraging reproducing kernel Hilbert space theory. Our simulation study and a real-world application in warfarin dosing show that the proposed direct method compares favorably with competing indirect regression-based methods.

1 Introduction

Personalized treatment rules, also known as individualized treatment rules or treatment policies, are strategies for assigning treatments, such as medication or social policy, to individuals based on their characteristics. These strategies have gained attention in various fields, such as medicine, education, and social science, due to their superior performance compared to traditional "one-size-fits-all" treatment assignment approaches. For example, many studies have shown that personalized warfarin dosing strategies, which take into account patient characteristics, including pharmacogenetics, are more effective than standard dosing based on an empirical protocol (Anderson et al., 2007; Stergiopoulos & Brown, 2014). Most research on personalized treatment rules, including the warfarin example, aims to optimize the individual's outcome by producing a single recommended treatment level for each study unit under the rule (Laber & Zhao, 2015; Chen et al., 2016; Kallus & Zhou, 2018; Zhu et al., 2020; Zhou et al., 2021; Schulz & Moodie, 2021; Hua et al., 2022; Wang & Wang, 2023). However, single-level dose recommendations may be overly ambitious or inadequate, as the primary goal of treatment in many real-world applications is not to achieve a specific value, but rather to ensure that a study unit's outcome falls within a favorable range. For instance, warfarin should be prescribed to keep a patient's international normalized ratio, a measure of the time for the blood to clot, within a desired range, usually between 2 and 3 (January et al., 2014). Similarly, major medical associations recommend targeting appropriate ranges for chronic disease management metrics such as hemoglobin level and blood pressure (American Association of Clinical Endocrinologists and Others, 2019; Flack & Adekola, 2020).

In practice, the relationship between the amount of treatment and the likelihood of favorable outcomes, referred to as the dose-probability curve, can be biphasic, resulting in an inverted U-shaped curve across dose levels; this phenomenon is called hormesis (Mattson, 2008). These hormetic dosage responses are common in many real-world applications, including the above warfarin example (Blann et al., 2003), clinical trials (Calabrese, 2008), public health (Cook & Calabrese, 2006), biology, toxicology, and medicine (Calabrese

& Mattson, 2017). Under hormesis, to have a probability of favorable outcomes greater than a certain level, the treatment dose often falls within a two-sided interval.

As such, interval-based dosing, rooted in the concept of hormesis, is widely utilized in clinical practice. In many cases, it is not only preferred but, quite often, the only feasible option. This is especially evident in the following two key clinical settings. The first clinical setting involves subgroup-specific dose interval recommendations, which are commonly established from a policy or guideline-making perspective. For example, Ageno et al. (2012) and Witt et al. (2016) recommend an initial warfarin dose ranging from 5 to 10 mg for most patients, while a dose of less than 5 mg is often recommended for elderly patients or those with poor nutrition, liver disease, congestive heart failure, or a high risk of bleeding. In prostate cancer radiotherapy, Li et al. (2021) recommended that the definitive treatment doses for prostate cancer can increase to 76-80 gray (Gy) when applying 3-dimensional conformal radiotherapy or intensity-modulated radiotherapy with conventional fractionated radiotherapy. Similar types of interval recommendations were given for radiotherapy of thyroid cancer and lung cancer (Luster et al., 2008; Saffarzadeh et al., 2021; Wadsley et al., 2023). In many scenarios of this setting, the absorbed or ingested dose is more important than the administered dose. The second clinical setting is chronic disease management, where blood sugar level and blood pressure level control are of great importance for elderly subjects. Since these metrics are influenced by various factors, such as lifestyle and dietary habits, a single-level dose recommendation is not feasible even with pharmaceutical interventions. For example, the actual blood sugar level for a given dose of a glucose-lowering drug can vary depending on the timing of administration, the patient's activity level that day, and dietary intake. As a result, the American Diabetes Association (ADA) and American Heart Association only recommend intervals for these two metrics. For example, the ADA recommends a fasting glucose level of 80-130 mg/dL for healthy elderly and 90-150 mg/dL for elderly with notable co-morbidities (American Diabetes Association, 2023). A similar guideline exists for blood pressure management (Qaseem et al., 2017). Moreover, identifying and estimating the optimal range of variables associated with preferred outcomes is of significant interest in fields beyond medicine, including education, nutrition, neuroscience, management, and social sciences (Zivin & Neidell, 2014; Conner et al., 2015; Northoff & Tumati, 2019; Albalate et al., 2023; Liu et al., 2023). These examples illustrate the practical application of interval-based decision rules in various settings.

Not surprisingly, determining the optimal personalized therapeutic dose range is challenging, as it requires learning both the lower and upper bounds of the unit-specific dose interval. This target is distinct from that of optimal single-level dose rules such as Chen et al. (2016). Even when the dose-probability curve is unimodal or inverted U-shaped, a single-dose rule aims to identify the dose maximizing a reward, whereas our target is the set of doses for which the probability of a favorable outcome exceeds a prespecified level. Likewise, existing methods for one-sided intervals, which rely on monotonicity of the dose-probability curve (Chen et al., 2022; Park et al., 2024), can be used to recover a two-sided interval only indirectly by estimating one bound while treating the other as fixed and alternating between the two updates; see Remark 1. Such an iterative construction can be computationally demanding, may be sensitive to initialization and convergence behavior, and is not naturally aligned with unrestricted product function spaces. In particular, theoretical tools for studying approximation error over product reproducing kernel Hilbert spaces (RKHSs) do not directly apply when the optimization domain is restricted through alternating monotonicity constraints.

This paper proposes a new approach to estimating personalized therapeutic dose ranges under the hormetic dose-probability curve. Our contributions are summarized as follows:

1. We propose a doubly-robust loss function that accommodates both monotonic and non-monotonic pairs[1]. The proposed loss assigns higher values to non-monotonic intervals than to monotonic intervals, while remaining robust to misspecification of either the dose-probability curve or the propensity score (Scharfstein et al., 1999; Lunceford & Davidian, 2004; Bang & Robins, 2005); see Section 3.2 for details.

2. Building on these properties, we formulate the empirical risk minimization problem over the full product function space in Section 3.3. This enables simultaneous estimation of the lower and upper

[1]A monotonic pair refers to a pair in which the lower bound is less than or equal to the upper bound for all covariate values, whereas a non-monotonic pair violates this relationship for some covariate values; see Section 2 for the formal definitions.

bounds, without relying on constrained alternating updates or iterative learning of one bound at a time. We also provide practical implementation details, including hyperparameter tuning and optimization strategies.

3. In Section 4, we establish an excess-risk bound that accounts for both nuisance estimation error and product-RKHS approximation error. In addition, we establish the convergence rate of the estimated dose bounds themselves; see Theorems 4.2 and 4.3.

4. In Sections 5 and 6, we use simulations and a data application to show that the resulting direct estimator compares favorably with the indirect approach discussed in Section 3.1.

Although methods exist for designing personalized policies that involve interval-based quantities, our work offers a concrete and distinct contribution. For instance, the method proposed by Cai et al. (2023) can also generate two-sided dose interval recommendations, but its goal is not to identify the personalized therapeutic dose range. In simple terms, their target estimand is defined as follows. First, the dose range is divided into user-specified sub-intervals, which are often infinitesimally small. The optimal dose range is then identified as the sub-interval that maximizes the dose-probability curve when treatment is assigned within it. Their approach is motivated by detecting multiple change points in the dose-probability curve to identify a dose range that maximizes the outcome, which contrasts with our motivation and goal—identifying a treatment range that leads to a preferred outcome. Additionally, several works on Interval Estimation have emerged in the context of multi-stage contextual bandit and reinforcement learning problems (Lai, 1987; Kaelbling, 1993; Strehl & Littman, 2005; 2008; Karampatziakis et al., 2020; Hong et al., 2022). While these methods differ in procedural details, the central idea of Interval Estimation is to identify the policy that maximizes the upper bound of the outcome's uncertainty interval. In these approaches, the term interval refers to the uncertainty in the outcome estimate, rather than to the policy itself. In contrast, our framework defines intervals directly in terms of policies rather than outcome estimates; see Section 2 for a formal definition of our interval-based policy. Therefore, unlike the Interval Estimation approach, our framework focuses on policies that are explicitly defined as intervals.

2 Preliminary

Let the subscript i denote the ith study unit. For each unit $i \in \{1, \ldots, N\}$, we observe $\boldsymbol{O}_i = (Y_i, A_i, \boldsymbol{X}_i) \in \mathcal{O}$ which is an independent and identically distributed (i.i.d.) realization from a distribution $P_{\boldsymbol{O}}$. Here, $\boldsymbol{X}_i$ is a d-dimensional pretreatment covariate with support $\mathcal{X} \subseteq \mathbb{R}^d$, A_i is the treatment dose taking its value in the interval $[0, 1]$, and $Y_i \in \mathbb{R}$ is the observed outcome/reward. For simplicity, we assume that the treatment is transformed so that its range is the unit interval. To define the dose assignment rule, we use the potential outcome (Rubin, 1974). Specifically, let $Y_i^{(a)}$ denote the potential outcome/reward when the treatment dose is $A_i = a$. In what follows, we suppress the subscript i unless necessary.

Let $\mathcal{F}$ be a decision space, a set of dose assignment rules, specified by an investigator, where $f \in \mathcal{F}$ is a dose assignment rule that maps $\boldsymbol{X}$, the characteristics of a study unit, to $f(\boldsymbol{X}) \in [0, 1]$, a dose level. In the context of learning two-sided dose intervals, we define $\mathcal{F}^{\otimes 2} = \big\{(f_L, f_U) \,|\, f_L, f_U \in \mathcal{F}\big\}$, a set of pairs of dose levels, as a decision space. Let $\mathcal{F}_{\text{mnt}} = \{(f_L, f_U) \in \mathcal{F}^{\otimes 2} \mid f_L(\boldsymbol{x}) \leq f_U(\boldsymbol{x}) \text{ for all } \boldsymbol{x}\}$ be the collection of pairs of functions such that the lower bound is smaller than or equal to the upper bound for all $\boldsymbol{x}$. Let $\mathcal{F}_{\text{nmnt}} = \{(f_L, f_U) \in \mathcal{F}^{\otimes 2} \mid f_L(\boldsymbol{x}) > f_U(\boldsymbol{x}) \text{ for some } \boldsymbol{x}\}$ be the collection of pairs of functions such that the lower bound is larger than the upper bound for some $\boldsymbol{x}$.

The interval rules can be more useful than the single-level dose assignment rules when one or more of the following scenarios happen: (i) assigning any value in an interval yields almost the same outcome as assigning the optimal treatment rule; (ii) finding the optimal treatment rule is inefficient; (iii) an investigator has subject matter knowledge about the shape of the dose-probability curve. In particular, we consider a $(\alpha, \mathcal{T})$-probability dose interval (PDI) introduced in Li (2018) and Chen et al. (2022):

$$\forall a \in [f_L(\boldsymbol{X}; \alpha, \mathcal{T}), f_U(\boldsymbol{X}; \alpha, \mathcal{T})] \in \mathcal{F}_{\text{mnt}} \;\Rightarrow\; \Pr\big\{Y^{(a)} \in \mathcal{T} \,|\, \boldsymbol{X}\big\} \geq \alpha \,. \tag{1}$$

Hereafter, we omit $(\alpha, \mathcal{T})$ in f_L and f_U for notational brevity. In words, if a study unit with characteristic $\boldsymbol{X}$ is treated with a dose level in $[f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]$, then the study unit's outcome lies in the desired range $\mathcal{T} \subseteq \mathbb{R}$

with probability at least α. Here, α plays a role analogous to that of a confidence level in hypothesis testing because the corresponding PDI controls the type I error rate at $1-\alpha$. As a result, the choice of α depends on the desired confidence level for the scientific question at hand. In order to ensure meaningful PDIs, it is important not to set α too high, such as $\alpha = 0.99$, as this may result in no PDI satisfying the condition, especially when $\Pr\{Y^{(a)} \in \mathcal{T} \,|\, \boldsymbol{X}\}$ is small. The selection of α depends on the application and should ideally be guided by domain expertise. If such expertise is not available, we suggest choosing an $\alpha > 0.5$ or using estimated outcome regression to help determine α. We remark that α and $\mathcal{T}$ may depend on $\boldsymbol{X}$, i.e., the probability threshold and the desired range can also be personalized. In addition, let $R = \mathbb{1}(Y \in \mathcal{T})$ be the indicator of whether an individual's outcome belongs to the desired range $\mathcal{T}$, which is referred to as the discretized outcome. The optimal PDI, denoted by $[f_L^*, f_U^*]$, is the PDI that has the longest length and belongs to $\mathcal{F}_{\text{mnt}}$. Formally, for a fixed $(\alpha, \mathcal{T})$, f_L^* and f_U^* satisfy $[f_L(\boldsymbol{X}), f_U(\boldsymbol{X})] \subseteq [f_L^*(\boldsymbol{X}), f_U^*(\boldsymbol{X})]$ for any $\boldsymbol{X}$ and any f_L and f_U satisfying (1).

Lastly, we introduce additional notation used throughout. For random variables U, V, and W, let $U \perp\!\!\!\perp V \,|\, W$ denote the conditional independence of U and V given W. We use O_P and o_P to denote stochastic boundedness and convergence in probability. For two non-negative sequences a_N and b_N, let $a_N \precsim b_N$ denote $a_N \leq C \cdot b_N$ for some constant $C > 0$, and let $a_N \asymp b_N$ denote $a_N \precsim b_N$ and $b_N \precsim a_N$. Let $\mathbf{1}_N$ be the length N vector of ones. Lastly, we denote $(x)_+ = \max(x, 0)$ and $(x)_- = \max(-x, 0)$, satisfying the decomposition $x = (x)_+ - (x)_-$.

Next, in order to establish identification of the PDI, we make the following assumptions.

Assumption 1 (Stable Unit Value Treatment Assumption; Rubin (1978)) *$Y = Y^{(A)}$ almost surely;*

Assumption 2 (Unconfoundedness) *We have $Y^{(a)} \perp\!\!\!\perp A \,|\, \boldsymbol{X}$ for any $a \in [0, 1]$;*

Assumption 3 (Positivity) *The generalized propensity score $e^*(A \,|\, \boldsymbol{X}) = P(A \,|\, \boldsymbol{X})$ (Imbens, 2000) is bounded below by a constant $c_e > 0$, i.e., $e^*(A \,|\, \boldsymbol{X}) > c_e$ for all $(A, \boldsymbol{X}) \in [0, 1] \otimes \mathcal{X}$.*

We refer readers to Hirano & Imbens (2004) and Hernán & Robins (2020) for substantive discussions on these assumptions. Under these assumptions, the probability of obtaining favorable outcomes given covariates, i.e., $\Pr\{Y^{(a)} \in \mathcal{T} \,|\, \boldsymbol{X}\}$, can be identified from the observed data as $\mu^*(a, \boldsymbol{X}) = \Pr(R = 1 \,|\, A = a, \boldsymbol{X})$, which is the dose-probability curve of R for a unit with characteristic $\boldsymbol{X}$.

To guarantee the existence of f_L^* and f_U^*, we make the following assumptions on μ^*:

Assumption 4 (Hormesis) *For any $\boldsymbol{X} \in \mathcal{X}$, there exists a constant $c_\mu > 0$ such that $\mu^*(0, \boldsymbol{X}) < \alpha - c_\mu$, $\mu^*(1, \boldsymbol{X}) < \alpha - c_\mu$, and $\mu^*(a_{\boldsymbol{X}}, \boldsymbol{X}) > \alpha + c_\mu$ for some $a_{\boldsymbol{X}} \in (0, 1)$;*

Assumption 5 (Smoothness and Connected PDI) *For any $\boldsymbol{X} \in \mathcal{X}$, $\mu^*(a, \boldsymbol{X})$ is Lipschitz continuous with respect to a and a level set $\{a \,|\, \mu^*(a, \boldsymbol{X}) \geq \alpha\}$ is a closed interval.*

Assumption 4 states that moderate dose levels are more likely to result in a favorable outcome compared to low and high doses, also known as hormesis (Mattson, 2008). This phenomenon can be found in the inverted U-shaped dose-probability curve; see Section 6 for details on warfarin and Calabrese (2008) for additional examples of drugs that have hormetic dose-probability curves. If Assumption 4 is violated at $\boldsymbol{X} = \boldsymbol{x}^\dagger$, it may imply that (i) study units having their covariates as $\boldsymbol{x}^\dagger$ always have favorable or unfavorable outcomes regardless of the treatment dose level, i.e., $\mu^*(a, \boldsymbol{x}^\dagger) > \alpha$ or $\mu^*(a, \boldsymbol{x}^\dagger) < \alpha$ for all a, and/or that (ii) treatment is not hormetic, say $\mu^*(a, \boldsymbol{x}^\dagger)$ increases monotonically with a. The first part of Assumption 5 states that μ^* is smooth with respect to dose, while the second part states that the therapeutic dose range is a connected interval. Assumptions 4 and 5 guarantee the existence of the optimal PDI, but there are other sufficient conditions; see Section A.1 of the Appendix for details.

3 Methodology

3.1 Indirect Method

We present a simple two-step approach to estimate the optimal PDI, often referred to as the indirect method (Moodie et al., 2012; Chakraborty & Moodie, 2013). In the first step, we estimate the dose-probability curve $\widehat{\mu}$ based on a posited outcome model. For instance, one can fit a logistic regression model of R on $(A, A^2, \boldsymbol{X})$. Alternatively, flexible machine learning classifiers can be used, such as random forest (Breiman, 2001), gradient boosting (Friedman, 2001), neural net (Ripley, 1994), support vector machines (Steinwart & Christmann, 2008), and many others; see Hastie et al. (2009) for additional examples. In the second step, we find the range $[\widehat{f}_L(\boldsymbol{X}), \widehat{f}_U(\boldsymbol{X})]$ that satisfies the PDI criterion in (1), which can be done by a grid search over the unit square $[0,1]^2$.

Although the indirect method is simple to implement, it can have notable drawbacks in terms of finite-sample performance as demonstrated in the simulation study and application in Sections 5 and 6. Additionally, the estimated PDI using indirect methods may not be available in a closed form unless $\widehat{\mu}$ has a very simple form. Furthermore, since the indirect rule only uses the dose-probability curve, it is more sensitive to model misspecification than methods using both the dose-probability curve and the generalized propensity score. This motivates our direct approach presented in the following sections.

3.2 Construction of the Loss and Risk Functions

We propose a method that directly estimates the optimal PDI from empirical risk minimization (ERM). An essential part of the proposed direct method is to design a risk function that is minimized at (f_L^*, f_U^*). If the risk function does not have (f_L^*, f_U^*) as a minimizer, the estimator obtained from the ERM does not converge to (f_L^*, f_U^*) in probability. We first start by reviewing the loss function in Chen et al. (2022) that has an inverse probability-weighted (IPW) form defined over $(f_L, f_U) \in \mathcal{F}_{\text{mnt}}$:

$$\mathcal{L}_{\text{IPW}}(\boldsymbol{O}, f_L, f_U; e) = \frac{\alpha(1-R)\mathbb{1}\big\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\} + (1-\alpha)R\mathbb{1}\big\{A \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}}{e(A \,|\, \boldsymbol{X})} \,.$$

The loss function is determined by the following four arguments. The first argument, $\boldsymbol{O}$, is the observed data. The second and third arguments, (f_L, f_U), form a PDI candidate with the monotonicity condition, i.e., $f_L(\boldsymbol{x}) \leq f_U(\boldsymbol{x})$. The last argument is a user-specified propensity score model. The IPW loss function above penalizes a study unit for either of the following cases: (i) the outcome does not belong to the desired range, i.e., $R = 0$, even though the dose level belongs to the PDI candidate, or (ii) the outcome belongs to the desired range, i.e., $R = 1$, even though the dose level lies outside of the PDI candidate. These errors can be seen as false positives and false negatives, respectively. The coefficients α and $1-\alpha$ assign different weights to these two errors. When the propensity score model is correctly specified, the IPW loss function can be used to identify (f_L^*, f_U^*) as $\mathrm{E}\big\{\mathcal{L}_{\text{IPW}}(\boldsymbol{O}; f_L, f_U; e^*)\big\}$ is minimized at (f_L^*, f_U^*).

One drawback of $\mathcal{L}_{\text{IPW}}$ is its sensitivity to misspecification of the propensity score model. Therefore, we can consider the following augmented IPW (AIPW) loss function over $(f_L, f_U) \in \mathcal{F}_{\text{mnt}}$ that is robust to the propensity score model misspecification:

$$\begin{aligned}
\mathcal{L}_{\text{AIPW}}(\boldsymbol{O}, f_L, f_U; \mu, e) = {} & \alpha \frac{\{\mu(A, \boldsymbol{X}) - R\}\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}}{e(A \,|\, \boldsymbol{X})} \\
& + \alpha \int \{1 - \mu(a, \boldsymbol{X})\}\mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \\
& + (1-\alpha)\frac{\{R - \mu(A, \boldsymbol{X})\}\mathbb{1}\{A \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}}{e(A \,|\, \boldsymbol{X})} \\
& + (1-\alpha) \int \mu(a, \boldsymbol{X})\mathbb{1}\{a \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \,.
\end{aligned}$$

Compared to $\mathcal{L}_{\text{IPW}}$, $\mathcal{L}_{\text{AIPW}}$ has one more argument μ which is a user-specified model for the dose-probability curve. The terms weighted by α and those weighted by $1-\alpha$ can be seen as functions that penalize false

positive and false negative errors, respectively. These terms resemble the efficient influence function of the average treatment effect (Hahn, 1998), and, not surprisingly, it has the doubly-robust property in the following manner: the AIPW loss function is minimized at the optimal PDI so long as μ or e, but not necessarily both, is correctly specified; see Lemma 3.1 for a related discussion.

The AIPW loss function $\mathcal{L}_{\text{AIPW}}$ has a limitation in that it is defined solely over monotonic intervals $\mathcal{F}_{\text{mnt}}$, i.e., PDI candidates in $\mathcal{F}_{\text{nmnt}}$ are excluded from the loss function. Consequently, if $\mathcal{L}_{\text{AIPW}}$ is used, the optimization domain, denoted by $\mathcal{H}_{\text{mnt}}$, must adhere to the monotonicity requirement between the lower and upper bound candidates, which implies that $\mathcal{H}_{\text{mnt}}$ should be a subset of $\mathcal{F}_{\text{mnt}}$. Moreover, $\mathcal{H}_{\text{mnt}}$ should be rich enough to accurately approximate any function in $\mathcal{F}_{\text{mnt}}$. These two observations indicate that an ideal choice for $\mathcal{H}_{\text{mnt}}$ would be a rich subset of $\mathcal{F}_{\text{mnt}}$. However, achieving both of these goals simultaneously is a significant challenge for practitioners for the following reasons. First, to maintain the monotonicity between the lower and upper bounds, iterative procedures are typically utilized in practice. By doing so, $\mathcal{H}_{\text{mnt}}$ may be restricted in a specific way, casting doubt on its suitability as an approximation subspace of $\mathcal{F}_{\text{mnt}}$. On the other hand, naively choosing a rich product function space, such as a product reproducing kernel Hilbert space (RKHS), for $\mathcal{H}_{\text{mnt}}$ can create difficulties in the optimization procedure because the AIPW loss function is not compatible with non-monotonic intervals.

The limitation of the AIPW loss function originates from its constrained domain. One way to resolve the limitation is to design a loss function that is well-defined over $\mathcal{F}^{\otimes 2}$ and is minimized at the optimal PDI (f_L^*, f_U^*). We can then estimate (f_L^*, f_U^*) over a product function space that approximates $\mathcal{F}^{\otimes 2}$ without the need to restrict the optimization domain. Our goal is to construct a simple extension of the AIPW loss from $\mathcal{F}_{\text{mnt}}$ to $\mathcal{F}^{\otimes 2}$ while preserving the same minimizer over monotonic candidates. To this end, we propose a new doubly-robust loss function $\mathcal{L} : \mathcal{O} \otimes \mathcal{F}^{\otimes 2} \to \mathbb{R}$, which is defined below:

$$\mathcal{L}(\boldsymbol{O}, f_L, f_U; \mu, e) = \begin{cases} \mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu, e) & \text{if } f_L(\boldsymbol{X}) \le f_U(\boldsymbol{X}) \\ C_{\mathcal{L}} & \text{if } f_L(\boldsymbol{X}) > f_U(\boldsymbol{X}) \end{cases}, \tag{2}$$

$$\mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu, e) = \begin{bmatrix} \{\mu(A, \boldsymbol{X}) - R\} \mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\} / e(A \,|\, \boldsymbol{X}) \\ + \int \{\alpha - \mu(a, \boldsymbol{X})\} \mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\} \, da \end{bmatrix},$$

where $C_{\mathcal{L}}$ is a sufficiently large constant such as $C_{\mathcal{L}} = 2 \sup_{\boldsymbol{O}, f_L, f_U} \left| \mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu, e) \right|$. In words, a common penalty $C_{\mathcal{L}}$ to every non-monotonic pair, chosen to be sufficiently large. Therefore, this construction guarantees that the penalty dominates the monotonic part of the loss and therefore preserves the desired ordering between admissible and non-admissible pairs. Additionally, from simple algebra, we find that the difference between $\mathcal{L}_{\text{AIPW}}$ and $\mathcal{L}^{(1)}$ does not depend on (f_L, f_U), indicating that the minimizers of $\mathcal{L}_{\text{AIPW}}$ and $\mathcal{L}^{(1)}$ are the same; see Section A.2 of the Appendix for details. In most observational studies, the true nuisance components (μ^*, e^*) are unknown and must be estimated from the observed data; see Section 3.3 for details.

The loss function $\mathcal{L}$ has advantageous properties, which are summarized in the following lemma.

Lemma 3.1 *Suppose that Assumptions 1-5 hold. Let $\mathcal{R} : \mathcal{F}^{\otimes 2} \to \mathbb{R}$ be the risk function associated with $\mathcal{L}$, i.e., $\mathcal{R}(f_L, f_U; \mu, e) = E\{\mathcal{L}(\boldsymbol{O}, f_L, f_U; \mu, e)\}$. Then, the optimal PDI is the minimizer of the risk function with the true nuisance functions, i.e.,*

$$(f_L^*, f_U^*) \in \underset{(f_L, f_U) \in \mathcal{F}^{\otimes 2}}{\arg\min} \mathcal{R}(f_L, f_U; \mu^*, e^*) \, . \tag{3}$$

Additionally, (f_L^*, f_U^*) is the minimizer of (3) so long as either μ^* or e^* is correctly specified, i.e., (3) holds for (μ', e^*) and (μ^*, e') for any μ' and e'.

Lemma 3.1 states that the optimal PDI is achieved by minimizing the risk function when both nuisance components are correctly specified. This means that the loss function is Fisher consistent in detecting the optimal PDI. Additionally, if either the dose-probability curve or the propensity score is correctly specified, the loss function remains Fisher consistent. This property is referred to as a doubly-robust property (Scharfstein et al., 1999; Lunceford & Davidian, 2004; Bang & Robins, 2005) and is further discussed in Theorem 4.2.

3.3 Empirical Risk Minimization Via the Difference of Convex Functions Algorithm

Despite its theoretical validity, the loss function $\mathcal{L}$ in (2) itself is difficult to use in ERM because it is not convex or smooth as induced by the indicator functions $\mathbb{1}\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\}$ and $\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}$. To address the non-smoothness, we design a surrogate loss function $\mathcal{L}_{\text{sur}}$ of the loss function $\mathcal{L}$ by replacing the indicator functions with truncated hinge-type functions as follows:

$$\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \mu, e) = \begin{cases} \mathcal{L}^{(1)}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \mu, e) & \text{if } f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X}) \\ \mathcal{L}^{(2)}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \mu, e) & \text{if } f_U(\boldsymbol{X}) < f_L(\boldsymbol{X}) < f_U(\boldsymbol{X}) + \epsilon \\ C_{\mathcal{L}} & \text{if } f_U(\boldsymbol{X}) + \epsilon \leq f_L(\boldsymbol{X}) \end{cases},$$

$$\mathcal{L}^{(1)}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \mu, e) = \begin{bmatrix} \big\{\mu(A, \boldsymbol{X}) - R\big\}\Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)/e(A \,|\, \boldsymbol{X}) \\ + \int \big\{\alpha - \mu(a, \boldsymbol{X})\big\}\mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \end{bmatrix},$$

$$\mathcal{L}^{(2)}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \mu, e) = \Phi_\epsilon(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}))\left[0.5\left\{ \begin{matrix} \mathcal{L}^{(1)}_{\text{sur}}(\boldsymbol{O}, f_L, f_L; \mu, e) \\ + \mathcal{L}^{(1)}_{\text{sur}}(\boldsymbol{O}, f_U, f_U; \mu, e) \end{matrix} \right\} - C_{\mathcal{L}}\right] + C_{\mathcal{L}},$$

where Ψ_ϵ and Φ_ϵ are truncated hinge-type functions:

$$\Psi_\epsilon(\ell, t, u) = \begin{cases} \frac{t-\ell+\epsilon}{\epsilon} & \text{if } t \in [\ell - \epsilon, \ell] \\ 1 & \text{if } t \in [\ell, u] \\ \frac{u+\epsilon-t}{\epsilon} & \text{if } t \in [u, u + \epsilon] \\ 0 & \text{otherwise} \end{cases}, \qquad \Phi_\epsilon(\ell, u) = \begin{cases} 0 & \text{if } u - \ell \in (-\infty, -\epsilon] \\ \frac{u-\ell+\epsilon}{\epsilon} & \text{if } u - \ell \in [-\epsilon, 0] \\ 1 & \text{if } u - \ell \in [0, \infty) \end{cases}.$$

Here, $\epsilon > 0$ is a bandwidth parameter where Ψ_ϵ and Φ_ϵ converge to the indicator functions as ϵ goes to zero; an appropriate choice of ϵ is discussed in Section 4. As depicted in Figure 1, Ψ_ϵ is a continuous approximation of $\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}$, whereas Φ_ϵ is a continuous approximation of $\mathbb{1}\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\}$. These continuous approximations enable the use of computationally efficient algorithms, such as gradient-based iterative methods, one of which we implement later in this section. Additionally, the truncated hinge-type functions used in $\mathcal{L}_{\text{sur}}$ ensure that the loss function is bounded, making it robust to noisy training data (Wu & Liu, 2007; Chen et al., 2016).

Using $\mathcal{L}_{\text{sur}}$, we obtain a PDI estimator from the ERM with the estimated nuisance components, i.e.,

$$(\widehat{f}_L, \widehat{f}_U) = \underset{(f_L, f_U) \in \mathcal{H}^{\otimes 2}}{\arg\min} \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{\text{sur}}\big(\boldsymbol{O}_i, f_L, f_U; \widehat{\mu}, \widehat{e}\big) + \lambda \big\| f_L \big\|^2_{\mathcal{H}} + \lambda \big\| f_U \big\|^2_{\mathcal{H}} \right\}. \tag{4}$$

Here, $\mathcal{H}$ is a user-specified function space, such as linear functions, splines, RKHS, or deep neural networks, and $\lambda > 0$ is a regularization parameter. While other choices are possible in principle, we focus on the case where $\mathcal{H}$ is an RKHS for the remainder of the paper. Specifically, let $\mathcal{H}$ be an RKHS with the Gaussian kernel function $k(\boldsymbol{x}, \boldsymbol{x}') = \exp\{-\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2/\gamma^2\}$ where $\gamma > 0$ is a bandwidth parameter. We remark that other universal kernels can be used with minor modifications. The RKHS is widely recognized for its approximation properties. For example, the Gaussian RKHS is dense in $L_2(P)$, the space of square-integrable functions, whenever P is a finite measure (Steinwart & Christmann, 2008, Theorem 4.63). Moreover, the RKHS is computationally efficient for our problem, as its parameters can be obtained efficiently via an iterative convex optimization procedure. Further details are provided in the remainder of this section.

Under this specification, the ERM (4) is defined over the tensor product RKHS $\mathcal{H}^{\otimes 2}$. From the representer theorem (Kimeldorf & Wahba, 1970; Schölkopf et al., 2001), the solution to the ERM, denoted by $(\widehat{f}_L, \widehat{f}_U)$, is represented as linear combinations of the kernel-transformed covariates, i.e.,

$$\widehat{f}_L(\boldsymbol{x}) = \widehat{\xi}_{L,0} + \sum_{i=1}^{N} \widehat{\xi}_{L,i} k(\boldsymbol{x}, \boldsymbol{X}_i), \qquad \widehat{f}_U(\boldsymbol{x}) = \widehat{\xi}_{U,0} + \sum_{i=1}^{N} \widehat{\xi}_{U,i} k(\boldsymbol{x}, \boldsymbol{X}_i).$$

The coefficients $\widehat{\boldsymbol{\xi}}_L = (\widehat{\xi}_{L,0}, \widehat{\xi}_{L,1}, \ldots, \widehat{\xi}_{L,N})^\intercal$ and $\widehat{\boldsymbol{\xi}}_U = (\widehat{\xi}_{U,0}, \widehat{\xi}_{U,1}, \ldots, \widehat{\xi}_{U,N})^\intercal$ are obtained from an ERM over an Euclidean space below:

$$\big(\widehat{\boldsymbol{\xi}}_L, \widehat{\boldsymbol{\xi}}_U\big) = \underset{\boldsymbol{\xi}_L, \boldsymbol{\xi}_U}{\arg\min}\, Q(\boldsymbol{\xi}_L, \boldsymbol{\xi}_U),$$

where

$$Q(\boldsymbol{\xi}_L, \boldsymbol{\xi}_U) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_{\mathrm{sur}}\big(\boldsymbol{O}_i, \xi_{L,0} + \boldsymbol{k}_i^\intercal \boldsymbol{\xi}_{L,(-0)}, \xi_{U,0} + \boldsymbol{k}_i^\intercal \boldsymbol{\xi}_{U,(-0)}; \widehat{\mu}, \widehat{e}\big) + \lambda\big\{\boldsymbol{\xi}_{L,(-0)}^\intercal K \boldsymbol{\xi}_{L,(-0)} + \boldsymbol{\xi}_{U,(-0)}^\intercal K \boldsymbol{\xi}_{U,(-0)}\big\} \ . \tag{5}$$

Here, $K = [k(\boldsymbol{X}_i, \boldsymbol{X}_j)]_{i,j} \in \mathbb{R}^{N\times N}$ is the gram matrix, $\boldsymbol{k}_i \in \mathbb{R}^N$ is the ith column of K, $\boldsymbol{\xi}_{L,(-0)} = (\xi_{L,1}, \ldots, \xi_{L,N})^\intercal$, and $\boldsymbol{\xi}_{U,(-0)} = (\xi_{U,1}, \ldots, \xi_{U,N})^\intercal$.

Although (5) provides a characterization of the PDI estimator, solving it remains challenging due to the nonconvexity of the surrogate loss function. While other nonconvex optimization methods can be applied, we use the Difference of Convex Functions algorithm (DC algorithm; Le Thi & Pham Dinh, 1997) due to its ready adaptation to our problem. Briefly, the DC algorithm is implemented as follows. First, we decompose the surrogate loss function $\mathcal{L}_{\mathrm{sur}}$ into a difference between two convex functions, say $\mathcal{L}_{\mathrm{sur}} = \mathcal{L}_{\mathrm{sur},+} - \mathcal{L}_{\mathrm{sur},-}$. We then iteratively solve a sequence of convex optimization problems using $\mathcal{L}_{\mathrm{sur},+}$, $\mathcal{L}_{\mathrm{sur},-}$, and specified initial values. Section 5.4 of Le Thi & Pham Dinh (2018), a recent review of the DC algorithm, emphasizes that the stability of the algorithm depends on (DC-1): selecting an appropriate DC decomposition and (DC-2): implementing a strategy to calculate good initial values. Although there is no unified approach to addressing these two factors, we outline how they are handled in our problem below. The proposed specifications demonstrate reasonable performance, as shown in the simulation study in Section 5.

(DC-1): *Decomposition of the Surrogate Loss Function*

First, we consider the convex decompositions of Ψ_ϵ and Φ_ϵ:

$$\Psi_{\epsilon,+}(\ell,t,u) = \begin{cases} \frac{-t+\ell}{\epsilon} & \text{if } t \le \ell - \epsilon \\ 1 & \text{if } \ell - \epsilon < t < u + \epsilon \\ \frac{t-u}{\epsilon} & \text{if } u + \epsilon \le t \\ 0 & \text{otherwise} \end{cases}, \qquad \Psi_{\epsilon,-}(\ell,t,u) = \begin{cases} \frac{-t+\ell}{\epsilon} & \text{if } t \le \ell \\ 0 & \text{if } \ell < t < u \\ \frac{t-u}{\epsilon} & \text{if } u \le t \\ 0 & \text{otherwise} \end{cases},$$

$$\Phi_{\epsilon,+}(\ell,u) = \begin{cases} \frac{\ell-u}{\epsilon} & \text{if } u - \ell < -\epsilon \\ 1 & \text{if } -\epsilon \le u - \ell \end{cases}, \qquad \Phi_{\epsilon,-}(\ell,u) = \begin{cases} \frac{\ell-u}{\epsilon} & \text{if } u - \ell < 0 \\ 0 & \text{if } 0 \le u - \ell \end{cases}.$$

Note that $\Psi_\epsilon = \Psi_{\epsilon,+} - \Psi_{\epsilon,-}$ and $\Phi_\epsilon = \Phi_{\epsilon,+} - \Phi_{\epsilon,-}$; see Figure 1 for graphical illustrations of Ψ_ϵ and Φ_ϵ and surrogates of these functions.

Let $\mu_+(a, \boldsymbol{X})$ and $\mu_-(a, \boldsymbol{X})$ be non-negative, non-decreasing, Lipschitz continuous functions in a satisfying $\mu(a, \boldsymbol{X}) = \mu_+(a, \boldsymbol{X}) - \mu_-(a, \boldsymbol{X})$. Since $\mu(a, \boldsymbol{X})$ is Lipschitz continuous, we can easily find such μ_+ and μ_-. For $0 \le \ell \le u \le 1$, we additionally define $G_+(\ell, u, \boldsymbol{X})$ and $G_-(\ell, u, \boldsymbol{X})$ as follows:

$$G_+(\ell,u,\boldsymbol{X}) = \int_0^\ell \mu_+(a,\boldsymbol{X})\,da + \int_0^u \mu_-(a,\boldsymbol{X})\,da + \alpha u \ ,$$
$$G_-(\ell,u,\boldsymbol{X}) = \int_0^u \mu_+(a,\boldsymbol{X})\,da + \int_0^\ell \mu_-(a,\boldsymbol{X})\,da + \alpha \ell \ .$$

Using these functions, we define $\mathcal{L}_{\mathrm{sur},+}$ and $\mathcal{L}_{\mathrm{sur},-}$ as follows:

- If $f_L(X) \le f_U(X)$:

$$\mathcal{L}_{\mathrm{sur},+}\big(\boldsymbol{O}, f_L, f_U; \mu, e\big) = \frac{\Phi_{\epsilon,+}(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}))}{e(A \,|\, \boldsymbol{X})} \begin{bmatrix} \big\{\mu(A,\boldsymbol{X}) - R\big\}_+ \big[\Psi_{\epsilon,+}\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)\big] \\ -\big\{\mu(A,\boldsymbol{X}) - R\big\}_- \big[\Psi_{\epsilon,-}\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)\big] \end{bmatrix} + G_+\big(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}), \boldsymbol{X}\big) + C_{\mathcal{L}} \ ,$$

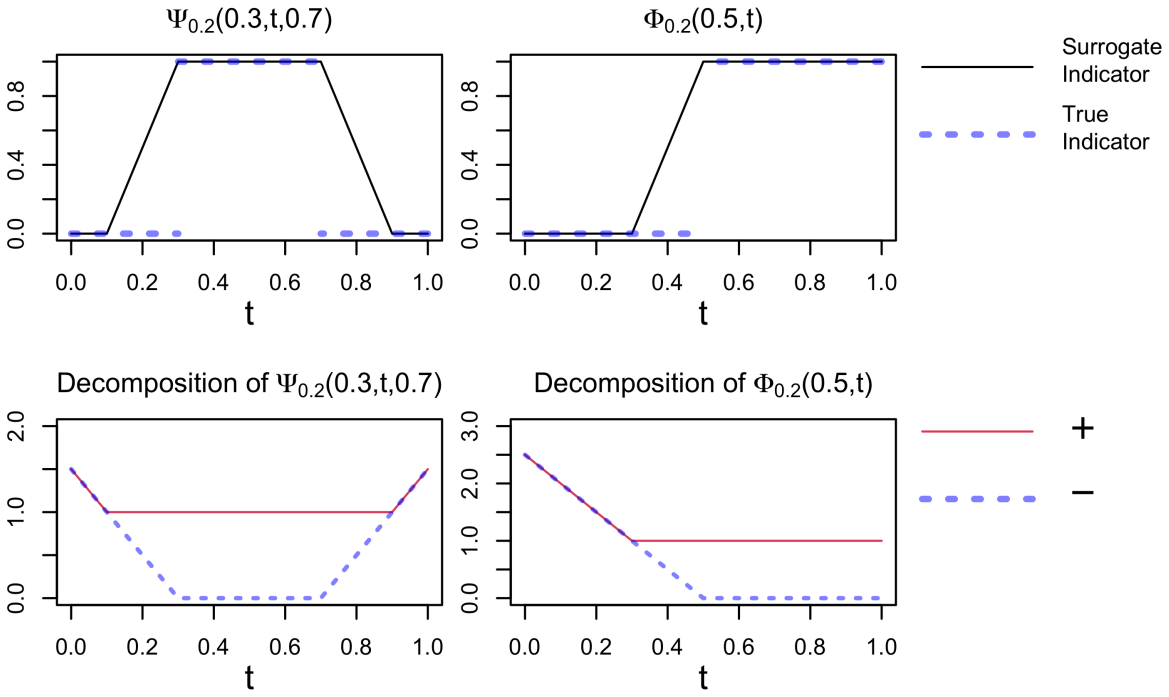

Figure 1: Graphical illustrations of Ψ_ϵ and Φ_ϵ with $\epsilon = 0.2$, and their convex decomposition.

and

$$\begin{aligned}
&\mathcal{L}_{\mathrm{sur},-}(\boldsymbol{O}, f_L, f_U; \mu, e) \\
&= \frac{\Phi_{\epsilon,+}(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}))}{e(A \,|\, \boldsymbol{X})} \begin{bmatrix} \{\mu(A,\boldsymbol{X}) - R\}_+ [\Psi_{\epsilon,-}(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X}))] \\ -\{\mu(A,\boldsymbol{X}) - R\}_- [\Psi_{\epsilon,+}(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X}))] \end{bmatrix} \\
&\quad + G_-(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}), \boldsymbol{X}) + C_{\mathcal{L}} \,.
\end{aligned}$$

- If $f_L(X) > f_U(X)$:

$$\begin{aligned}
&\mathcal{L}_{\mathrm{sur},+}(\boldsymbol{O}, f_L, f_U; \mu, e) \\
&= \frac{\Phi_{\epsilon,+}(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}))}{2e(A \,|\, \boldsymbol{X})} \begin{bmatrix} \{\mu(A,\boldsymbol{X}) - R\}_+ [\Psi_{\epsilon,+}(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X}))] \\ +\{\mu(A,\boldsymbol{X}) - R\}_+ [\Psi_{\epsilon,+}(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X}))] \\ -\{\mu(A,\boldsymbol{X}) - R\}_- [\Psi_{\epsilon,-}(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X}))] \\ -\{\mu(A,\boldsymbol{X}) - R\}_- [\Psi_{\epsilon,-}(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X}))] \end{bmatrix} \\
&\quad + \frac{\Phi_{\epsilon,-}(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}))}{2e(A \,|\, \boldsymbol{X})} \begin{bmatrix} \{\mu(A,\boldsymbol{X}) - R\}_+ [\Psi_{\epsilon,-}(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X}))] \\ +\{\mu(A,\boldsymbol{X}) - R\}_+ [\Psi_{\epsilon,-}(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X}))] \\ -\{\mu(A,\boldsymbol{X}) - R\}_- [\Psi_{\epsilon,+}(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X}))] \\ -\{\mu(A,\boldsymbol{X}) - R\}_- [\Psi_{\epsilon,+}(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X}))] \\ +2C_{\mathcal{L}} e(A \,|\, \boldsymbol{X}) \end{bmatrix} \\
&\quad + C_{\mathcal{L}} + C_{\mathrm{cvx}} \{f_L(\boldsymbol{X}) - f_U(\boldsymbol{X})\}^2 / \epsilon \,,
\end{aligned}$$

and

$$
\begin{aligned}
&\mathcal{L}_{\mathrm{sur},-}\big(\boldsymbol{O}, f_L, f_U; \mu, e\big) \\
&= \frac{\Phi_{\epsilon,+}(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}))}{2e(A \,|\, \boldsymbol{X})} \left[\begin{array}{l} \big\{\mu(A,\boldsymbol{X}) - R\big\}_+ \big[\Psi_{\epsilon,-}\big(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X})\big)\big] \\ +\big\{\mu(A,\boldsymbol{X}) - R\big\}_+ \big[\Psi_{\epsilon,-}\big(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)\big] \\ -\big\{\mu(A,\boldsymbol{X}) - R\big\}_- \big[\Psi_{\epsilon,+}\big(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X})\big)\big] \\ -\big\{\mu(A,\boldsymbol{X}) - R\big\}_- \big[\Psi_{\epsilon,+}\big(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)\big] \\ +2C_{\mathcal{L}} e(A \,|\, \boldsymbol{X}) \end{array} \right] \\
&\quad + \frac{\Phi_{\epsilon,-}(f_L(\boldsymbol{X}), f_U(\boldsymbol{X}))}{2e(A \,|\, \boldsymbol{X})} \left[\begin{array}{l} \big\{\mu(A,\boldsymbol{X}) - R\big\}_+ \big[\Psi_{\epsilon,+}\big(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X})\big)\big] \\ +\big\{\mu(A,\boldsymbol{X}) - R\big\}_+ \big[\Psi_{\epsilon,+}\big(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)\big] \\ -\big\{\mu(A,\boldsymbol{X}) - R\big\}_- \big[\Psi_{\epsilon,-}\big(f_L(\boldsymbol{X}), A, f_L(\boldsymbol{X})\big)\big] \\ -\big\{\mu(A,\boldsymbol{X}) - R\big\}_- \big[\Psi_{\epsilon,-}\big(f_U(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)\big] \end{array} \right] \\
&\quad + C_{\mathrm{cvx}} \{f_L(\boldsymbol{X}) - f_U(\boldsymbol{X})\}^2 / \epsilon \ .
\end{aligned}
$$

Here, C_{cvx} is a constant that guarantees the convexity. Specifically, if C_{cvx} is large enough, $\mathcal{L}_{\mathrm{sur},+}(\boldsymbol{O}, \ell, u; \mu, e)$ and $\mathcal{L}_{\mathrm{sur},-}(\boldsymbol{O}, \ell, u; \mu, e)$ are convex in (ℓ, u). Therefore, we recommend choosing a large value for C_{cvx}, say $C_{\mathrm{cvx}} = 10^4 C_{\mathcal{L}}$.

Using these functions, we finally define the following convex functions Q_+ and Q_-:

$$
\begin{aligned}
Q_+(\boldsymbol{\xi}_L, \boldsymbol{\xi}_U) &= \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{\mathrm{sur},+}\big(\boldsymbol{O}_i, \xi_{L,0} + \boldsymbol{k}_i^\intercal \boldsymbol{\xi}_{L,(-0)}, \xi_{U,0} + \boldsymbol{k}_i^\intercal \boldsymbol{\xi}_{U,(-0)}; \widehat{\mu}, \widehat{e}\big) \\
&\qquad + \lambda \big\{ \boldsymbol{\xi}_{L,(-0)}^\intercal K \boldsymbol{\xi}_{L,(-0)} + \boldsymbol{\xi}_{U,(-0)}^\intercal K \boldsymbol{\xi}_{U,(-0)} \big\} \ , \\
Q_-(\boldsymbol{\xi}_L, \boldsymbol{\xi}_U) &= \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{\mathrm{sur},-}\big(\boldsymbol{O}_i, \xi_{L,0} + \boldsymbol{k}_i^\intercal \boldsymbol{\xi}_{L,(-0)}, \xi_{U,0} + \boldsymbol{k}_i^\intercal \boldsymbol{\xi}_{U,(-0)}; \widehat{\mu}, \widehat{e}\big) \ .
\end{aligned} \tag{6}
$$

From straightforward algebra, one can show that Q_+ and Q_- are convex decompositions of Q in (5), i.e., Q_+ and Q_- are convex and satisfy $Q = Q_+ - Q_-$.

(DC-2): *Computation of Initial Values*

The solution to the DC algorithm depends on a $(2N+2)$-dimensional vector $(\boldsymbol{\xi}_L^{(0)}, \boldsymbol{\xi}_U^{(0)})$, which can be quite large even for moderate-sized datasets. Consequently, greedy algorithms are not appropriate for searching initial points when the dataset has a moderate to large number of observations. Therefore, we propose an alternative strategy as follows. We focus on the case of $\boldsymbol{\xi}_L^{(0)}$ below, as $\boldsymbol{\xi}_U^{(0)}$ can be selected using a similar procedure. Given the indirect rules ℓ_i for $i = 1, \ldots, N$, we consider the following internal division points with the internal division ratio parameter $\rho \in [0, 1]$:

$$
\ell_i^{(\rho)} = \rho \ell_i + (1 - \rho) \overline{\ell} \ , \quad i = 1, \ldots, N \ , \qquad\qquad \overline{\ell} = \frac{1}{N} \sum_{i=1}^{N} \ell_i \ . \tag{7}
$$

In other words, the internal division points $\boldsymbol{\ell}^{(\rho)} = (\ell_1^{(\rho)}, \ldots, \ell_N^{(\rho)})^\intercal$ are homogeneous and are close to the average of the indirect rules at small ρ. When $\rho = 0$, the internal division points for the lower bound are chosen as the average of the indirect rules for all i, i.e., $\boldsymbol{\ell}^{(\rho)} = \overline{\ell} \mathbf{1}_N$. On the contrary, when $\rho = 1$, the internal division points are chosen as the indirect rules themselves, i.e., $\boldsymbol{\ell}^{(\rho)} = (\ell_1, \ldots, \ell_N)^\intercal$. For the given internal division points $\boldsymbol{\ell}^{(\rho)}$ and the gram matrix K, we find the coefficient vector $\boldsymbol{\xi}_L^{(0)}$ satisfying $\boldsymbol{\ell}^{(\rho)} = \xi_{L,0}^{(0)} \mathbf{1}_N + K \boldsymbol{\xi}_{L,(-0)}^{(0)}$. That is, we find the coefficients that are associated with the internal division points. For example, these coefficients can be calculated as follows:

$$
\xi_{L,0}^{(0)} = \frac{1}{N} \sum_{i=1}^{N} \ell_i^{(\rho)} \ , \qquad\qquad \boldsymbol{\xi}_{L,(-0)}^{(0)} = K^{-1} \big\{ \boldsymbol{\ell}^{(\rho)} - \xi_{L,0}^{(0)} \mathbf{1}_N \big\} \ .
$$

With this choice, we have $\xi_{L,0}^{(0)} = \overline{\ell}$ for all ρ. Additionally, $\boldsymbol{\xi}_{L,(-0)}^{(0)}$ tends to be small and homogeneous for small values of ρ, and large and heterogeneous for large values of ρ, reflecting the properties of $\boldsymbol{\ell}^{(\rho)}$.

Given the convex decomposition and the initial value, the DC algorithm can be implemented as in Algorithm 1:

Algorithm 1. DC Algorithm

Require: Initial coefficients $(\boldsymbol{\xi}_L^{(0)}, \boldsymbol{\xi}_U^{(0)})$

Let Q_+ and Q_- be the convex functions satisfying $Q = Q_+ - Q_-$, i.e., (6).

Update $\boldsymbol{\xi}_L^{(t)}$ and $\boldsymbol{\xi}_U^{(t)}$ based on the following optimization until convergence:

$$\begin{pmatrix} \boldsymbol{\xi}_L^{(t+1)} \\ \boldsymbol{\xi}_U^{(t+1)} \end{pmatrix} = \underset{\boldsymbol{\xi}_L, \boldsymbol{\xi}_U}{\arg\min} \left[Q_+(\boldsymbol{\xi}_L, \boldsymbol{\xi}_U) - \left\{ \nabla Q_-(\boldsymbol{\xi}_L^{(t)}, \boldsymbol{\xi}_U^{(t)}) \right\}^\intercal \begin{pmatrix} \boldsymbol{\xi}_L \\ \boldsymbol{\xi}_U \end{pmatrix} \right] .$$

return Converged $\boldsymbol{\xi}_L^{(t)}$ and $\boldsymbol{\xi}_U^{(t)}$

We discuss how to select hyperparameters for implementing the optimization problem (5), namely the kernel bandwidth parameter γ, the regularization parameter λ, and the internal division ratio parameter ρ. For γ, we recommend using the median heuristic (Garreau et al., 2017), while for λ and ρ, we suggest using cross-validation. The details of these selection methods are provided in Section A.3 of the Appendix.

Lastly, we consider some remedies to address the invalid intervals; notably, the PDI estimate can be invalid when (i) the bound estimate does not lie over a proper dose range or (ii) the upper bound estimate is not larger than the lower bound estimate, i.e., non-monotonic bounds. For the first violation, we simply winsorize the estimates escaping the dose range; we remark that the winsorization does not affect its statistical property; see Theorem 4.2 for details. To avoid the second violation, in the ERM, we may introduce an additional regularization term penalizing when violations happen. For instance, in the optimization problem in (5), we may use the following $\widetilde{Q}_\kappa$ having an additional regularization term:

$$\widetilde{Q}_\kappa(\boldsymbol{\xi}_L, \boldsymbol{\xi}_U) = Q(\boldsymbol{\xi}_L, \boldsymbol{\xi}_U) + \kappa \sum_{i=1}^{N} (\xi_{L,i} - \xi_{U,i})_+ , \quad \kappa \geq 0 . \tag{8}$$

The new regularization term assigns a larger loss function value if the lower bound coefficient is larger than the upper bound coefficient. This helps to ensure that the PDI estimate stays within the specified dose range. Again, the regularization parameter κ can be determined through cross-validation. Despite these measures, there may still be cases where violations occur in the test data. In such instances, we resort to the pointwise dose rule by taking the average of $\widehat{f}_L$ and $\widehat{f}_U$. However, our simulations and data analysis indicate that the direct method yields very few violations, even without incorporating the coefficient-regularization term in (8); see the simulations and data applications in Sections 5 and 6 for empirical evidence. Therefore, we take $\kappa = 0$ in the subsequent simulation studies and real-world data analyses.

Remark 1 *Instead of learning f_L and f_U jointly, one can use an iterative procedure to optimize the lower or upper bound function at a time while the other is treated as fixed. In particular, the procedure can be summarized as follows: (i) given an estimator of the lower bound at iteration t, denoted by $\widehat{f}_L^{(t)}$, the upper bound is only estimated, denoted by $\widehat{f}_U^{(t+1)}$, with the restriction $\widehat{f}_L^{(t)} \leq \widehat{f}_U^{(t+1)}$; (ii) given the new upper bound estimator $\widehat{f}_U^{(t+1)}$, the lower bound is estimated, denoted by $\widehat{f}_L^{(t+1)}$, with the restriction $\widehat{f}_L^{(t+1)} \leq \widehat{f}_U^{(t+1)}$; and (iii) repeat (i) and (ii) until both bound estimators converge. The procedure can be viewed as an extension of the blockwise optimization method, also known as the nonlinear Gauss-Seidel method; see Wright (2015) for details. However, we argue that the iterative procedure is suboptimal for several reasons. First, there is no guarantee of its convergence and it faces a theoretical issue due to the restrictions on the PDI estimators, i.e., $\widehat{f}_L^{(t)} \leq \widehat{f}_U^{(t+1)}$ and $\widehat{f}_L^{(t+1)} \leq \widehat{f}_U^{(t+1)}$. These constraints restrict the range of the PDI estimators in a complex way, making the optimization domain smaller than the unrestricted product RKHS $\mathcal{H}^{\otimes 2}$. Consequently, the approximation theory established under the unrestricted RKHS may not be applicable to the estimator obtained from this iterative procedure and the restricted RKHS. On the other hand, the ERM in (5) simultaneously uses*

both lower and upper bound estimators without any restrictions over the candidates, making the computation simpler and the RKHS theory applicable to our setting at the expense of more complicated, yet closed-form surrogate loss function.

Remark 2 *One may restrict the relationship between f_L^* and f_U^* so that the width of the PDI is constant for all $\boldsymbol{X}$, i.e., $f_U^*(\boldsymbol{X}) = f_L^*(\boldsymbol{X}) + w$ for a non-negative constant $w \geq 0$. The estimators of the PDI are represented as $\widehat{f}_L(\boldsymbol{x}) = \widehat{\xi}_0 + \sum_{i=1}^N \widehat{\xi}_i k(\boldsymbol{x}, \boldsymbol{X}_i)$ and $\widehat{f}_U(\boldsymbol{x}) = \widehat{\xi}_0 + \widehat{w} + \sum_{i=1}^N \widehat{\xi}_i k(\boldsymbol{x}, \boldsymbol{X}_i)$ where the coefficients $\widehat{\xi}_0$, $\widehat{\boldsymbol{\xi}}_{(-0)} = (\widehat{\xi}_1, \ldots, \widehat{\xi}_N)^\intercal$, and $\widehat{w} \geq 0$ are obtained from the following ERM:*

$$\left(\widehat{\xi}_0, \widehat{\boldsymbol{\xi}}_{(-0)}, \widehat{w}\right) = \underset{\xi_0, \boldsymbol{\xi}_{(-0)}, w}{\arg\min} \left\{ \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{sur}\left(\boldsymbol{O}_i, \xi_0 + k_i^\intercal \boldsymbol{\xi}_{(-0)}, \xi_0 + w + k_i^\intercal \boldsymbol{\xi}_{(-0)}; \widehat{\mu}, \widehat{e}\right) + \lambda \boldsymbol{\xi}_{(-0)}^\intercal K \boldsymbol{\xi}_{(-0)} \right\}.$$

Although the constant-width PDI is easier to interpret and implement than the proposed method, it fails to account for variations in PDI width that may depend on $\boldsymbol{X}$. In the context of the warfarin dataset analyzed in Section 6, the estimated PDI exhibited significant variations with respect to age and gender in the warfarin dataset; see Table 3 for details. This suggests that the constant-width PDI model may be unreasonable for numerous real-world applications, including the warfarin dataset.

3.4 Estimation of the Nuisance Components

We conclude the section by presenting some methods for estimating nuisance functions μ^* and e^*. First, as assumed in Assumption 4, the dose-probability curve μ^* has a hormetic form, so it is desirable to obtain an estimator of μ^* that maintains the same form. Some examples of estimation techniques are (semi-)parametric models with a quadratic term of the dose level, and carefully designed nonparametric methods such as multidimensional monotone Bayesian additive regression trees (Chipman et al., 2022). Second, the generalized propensity score e^* is a conditional density due to the continuous nature of A, which is often regarded as a statistically more difficult object than the propensity score under discrete treatment. Again, one may use (semi-)parametric models to estimate the conditional density of A given $\boldsymbol{X}$ or nonparametric methods (Li & Racine, 2008; Zhu et al., 2015; Schuster et al., 2020).

The nuisance functions can also be estimated nonparametrically, but this requires a more careful estimation procedure. Suppose nonparametric methods are naively used to estimate the nuisance functions by utilizing the entire dataset, denoted as $\mathcal{D} = \{1, \ldots, N\}$, to (i) estimate the nuisance functions and (ii) estimate the PDI by solving the ERM, using the nuisance functions estimated in step (i). In this setup, each observation is used twice—once in step (i) and again in step (ii). Unfortunately, this "double-dipping" approach can lead to inconsistent PDI estimates unless the nuisance functions are estimated within function spaces of well-controlled complexity, such as the Donsker class or the Vapnik-Chervonenkis class.

To address this issue, we recommend using the cross-fitting procedure (Schick, 1986; Chernozhukov et al., 2018), which mitigates the risk of non-diminishing bias in PDI estimators. The cross-fitting procedure consists of the following steps: First, split the dataset $\mathcal{D}$ into non-overlapping split samples $\{\mathcal{D}_1, \ldots, \mathcal{D}_S\}$; second, use a subset of data to estimate the nuisance functions; and third, use the remaining subsample to solve the ERM for estimating the PDI. Algorithm 2 below summarizes the cross-fitting procedure. While any finite value of S can be used in principle, one can select $S = 2$ for simplicity and computational efficiency by following Bach et al. (2024).

Algorithm 2. Cross-fitting

Require: Split the data into equally sized S folds, say $\mathcal{D}_1, \ldots, \mathcal{D}_S$

for $s = 1, \ldots, S$ **do**

Obtain $(\widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})$, estimators of the nuisance functions μ^* and e^* using $\mathcal{D}_s^c$.

Choose hyperparameters from the cross-validation procedure (see Algorithm 3 of the Appendix) using $\mathcal{D}_s$ where the loss function is evaluated with $(\widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})$.

Obtain a PDI estimator $(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)})$ using the hyperparameters obtained in the previous step and $\mathcal{D}_s$.

end for

return Aggregated PDI estimator $\widehat{f}_L = S^{-1} \sum_{s=1}^S \widehat{f}_L^{(s)}$ and $\widehat{f}_U = S^{-1} \sum_{s=1}^S \widehat{f}_U^{(s)}$.

The cross-fitting procedure enables the characterization of the theoretical properties of the estimated PDI (e.g., excess risk) by ensuring that the nuisance functions are estimated on a separate portion of the data from where the PDI is evaluated. This approach mitigates the "double-dipping" issue by eliminating the dependence between the nuisance function estimation and PDI estimation steps; see the next section for further details. However, in practice, nonparametrically estimated nuisance functions can be unstable due to limited sample size and may yield little or no improvement compared to well-designed parametric methods. For this reason, in both the simulation study and data applications, we used parametrically estimated nuisance functions without the cross-fitting procedure, resulting in well-performing PDI estimates.

4 Theoretical Properties

We examine theoretical properties of the PDI estimator obtained from the ERM with cross-fitting. Let $\mathrm{E}^{\mathcal{D}}(\cdot)$ be the expectation operator that regards the sample $\mathcal{D}$ as given. Note that, for a generic function estimated from $\mathcal{D}_s^c$, denoted by $\widehat{g}^{(-s)}$, we have $\mathrm{E}^{\mathcal{D}}\{\widehat{g}^{(-s)}(\boldsymbol{O})\} = \int \widehat{g}^{(-s)}(\boldsymbol{o})\, dP(\boldsymbol{o})$. Similarly, we denote $\mathcal{R}^{\mathcal{D}}(f_L, f_U; \mu^{(-s)}, e^{(-s)}) = \mathrm{E}^{\mathcal{D}}\{\mathcal{L}(\boldsymbol{O}, f_L, f_U; \mu^{(-s)}, e^{(-s)})\}$ and $\mathcal{R}^{\mathcal{D}}_{\mathrm{sur}}$ analogously.

First, we show that using the surrogate indicator function is a reasonable choice when ϵ is small under an additional condition. We begin by introducing regularity conditions.

(RC1) The estimated propensity score $\widehat{e}^{(-s)}$ is lower bounded by a constant $c_e > 0$, i.e., $\widehat{e}^{(-s)}(A \,|\, \boldsymbol{X}) > c_e$ for all $(A, \boldsymbol{X})$. Additionally, the estimated dose curve $\widehat{\mu}^{(-s)}(a, \boldsymbol{X})$ is bounded in the unit interval, i.e., $\widehat{\mu}^{(-s)}(a, \boldsymbol{X}) \in [0, 1]$.

(RC2) For any sequence $a_N = o(1)$, there exists a finite non-negative number b and b_0 satisfying $\mathrm{E}^{\mathcal{D}}\big[\mathbb{1}\{\widehat{f}_L^{(s)}(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X}) \in (0, a_N)\}\big] \leq a_N b_0 N^b$.

Regularity Condition (RC1) states that the estimated propensity score and dose-probability curve satisfy the positivity and boundedness conditions, respectively. In practice, (RC1) can be empirically verified by investigating $\widehat{e}^{(-s)}$ and $\widehat{\mu}^{(-s)}$ over the observed samples, or it can be enforced by trimming or applying a link function. Of note, these conditions are commonly made for tractable inference of causal effects (e.g., Assumption 5.1 of Chernozhukov et al. (2018)).

Regularity Condition (RC2) means that the probability of $\widehat{f}_L^{(s)}$ and $\widehat{f}_U^{(s)}$ being closer than a small bandwidth a_N is proportional to the bandwidth with a factor $b_0 N^b$. Similar conditions have been used in the literature (e.g., (3.3) of Qian & Murphy (2011), (16) of Luedtke & van der Laan (2016), (A5) and (A6) of Shi et al. (2020), (A5) of Cai et al. (2023)). To better understand the condition, one can consider a derivative of the left-hand side with respect to a_N at 0, which reduces the condition that the derivative value is upper bounded by a polynomial function $b_0 N^b$. This implies that the probability density function of a random variable $\widehat{f}_L^{(s)} - \widehat{f}_U^{(s)}$, which can be seen as the degree of non-monotonicity, is effectively bounded by a polynomial function. A sufficient condition of (RC2) is that $\widehat{f}_L^{(s)}(\boldsymbol{X}) \leq \widehat{f}_U^{(s)}(\boldsymbol{X})$ uniformly holds over $\boldsymbol{X}$. This monotonicity condition can be empirically checked by investigating the empirical distribution of $\widehat{f}_L^{(s)}(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X})$ for the observed $\boldsymbol{X}$ values. If the monotonicity condition appears to be violated, potentially undermining the validity of (RC2), one can make (RC2) more plausible by using $\widetilde{Q}_\kappa$ in (8) with a large κ to enforce the constraint $\widehat{f}_L^{(s)} \leq \widehat{f}_U^{(s)}$.

The following lemma shows that using the surrogate indicator function is a reasonable choice when ϵ is small.

Lemma 4.1 *Let $(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)})$ be the PDI estimator obtained from the ERM (5) with the cross-fitting procedure. (Result 1) Suppose Assumptions 1-5 and Regularity Condition (RC1). Then, we have the following results for some constant c_0:*

$$\left|\mathcal{R}^{\mathcal{D}}_{sur}(f_L^*, f_U^*; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \leq c_0 \epsilon \ .$$

(Result 2) Suppose Assumptions 1-5 and Regularity Conditions (RC1)-(RC2) are satisfied. Then, we have the following results for some constant c_0:

$$\left|\mathcal{R}^{\mathcal{D}}_{sur}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \leq c_0 \epsilon N^b \ .$$

(Result 1) of Lemma 4.1 states that, by viewing the estimated nuisance components as fixed functions, the surrogate risk function evaluated at the optimal PDI is not far from the original risk function evaluated at the optimal PDI. (Result 2) of the Lemma states that a similar result is established for the PDI estimator obtained from the ERM under the additional condition (RC2).

Next, we present the result of the convergence of the estimated PDI in terms of the excess risk bound. In order to do so, we introduce additional regularity conditions.

(RC3) The support of $\boldsymbol{X}$ is compact, and the density of $\boldsymbol{X}$ is bounded above by a constant.

(RC4) Let $f_L^\dagger$ and $f_U^\dagger$ be the optimal PDI given $\widehat{\mu}^{(-s)}$ and $\widehat{e}^{(-s)}$, i.e.,

$$(f_L^\dagger, f_U^\dagger) \in \underset{(f_L, f_U) \in \mathcal{F}^{\otimes 2}}{\arg\min}\ \mathcal{R}^{\mathcal{D}}(f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) \ .$$

Then, the optimal PDI bounds (f_L^*, f_U^*) and $(f_L^\dagger, f_U^\dagger)$ belong to a Besov space with smoothness parameter $\beta > 0$, which is $\mathcal{B}^\beta_{1,\infty}(\mathbb{R}^d) = \{f \in L_\infty(\mathbb{R}^d) \,|\, \sup_{t>0} t^{-\beta}\{\omega_{r,L_1(\mathbb{R}^d)}(f,t)\} < \infty, r > \beta\}$ where ω_r is the modulus of continuity of order r.

(RC5) There exist positive constants C, r_μ, and r_e satisfying $\mathrm{E}^{\mathcal{D}}\left[\left\{\widehat{\mu}^{(-s)}(A, \boldsymbol{X}) - \mu^*(A, \boldsymbol{X})\right\}^2\right] \leq C \cdot N^{-2r_\mu}$, and $\mathrm{E}^{\mathcal{D}}\left[\left\{\widehat{e}^{(-s)}(A \,|\, \boldsymbol{X}) - e^*(A \,|\, \boldsymbol{X})\right\}^2\right] \leq C \cdot N^{-2r_e}$ with probability not less than $1 - \Delta_N$ where $\Delta_N \to 0$ as $N \to \infty$.

Regularity Condition (RC3) states that the covariate does not concentrate at a certain point. Regularity Condition (RC4) states that the optimal PDI corresponding to the true nuisance functions (μ^*, e^*) and the one corresponding to the estimated nuisance functions $(\widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})$ have smoothness of order β in a Besov space. Regularity Condition (RC5) implies that the estimated nuisance functions converge to their truth as N grows. Regularity Condition (RC5) can be seen as requirements on the smoothness of the nuisance functions μ^* and e^*. For instance, if μ^* belongs to the Hölder space with smoothness exponent δ_μ, the kernel density estimator for μ^* achieves a convergence rate of $O_P(N^{-2\delta_\mu/(2\delta_\mu + d')})$ in terms of mean squared error, where $d' = d+1$ is the dimension of $(A, \boldsymbol{X})$. This rate is minimax optimal, and similar rates can be attained by other nonparametric methods; for further details, see Stone (1980) and Chapter 1 of Tsybakov (2009). Therefore, (RC5) would hold for any r_μ smaller than $\delta_\mu/(2\delta_\mu + d')$. A similar interpretation applies to e^*.

Under these Regularity Conditions, the convergence of the PDI estimator can be characterized in terms of excess risk. Theorem 4.2 formally states the result.

Theorem 4.2 *Let $(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)})$ be the PDI estimator obtained from ERM with the cross-fitting procedure. Suppose that Assumptions 1-5 and Regularity Conditions (RC1)-(RC5) hold. Then, for all fixed $\epsilon > 0$, $d/(d+\tau) < p < 1$, $\tau > 0$, and $\lambda > 0$, we obtain the following result with probability $P_{\boldsymbol{O}}^N$ not less than $1 - 3e^{-\tau} - \Delta_N$:*

$$\begin{aligned}
&\mathcal{R}^{\mathcal{D}}\left(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*\right) - \mathcal{R}^{\mathcal{D}}\left(f_L^*, f_U^*; \mu^*, e^*\right) \\
&\leq c_1 \lambda \gamma^{-d} + c_2 \gamma^\beta + c_3 \left\{\gamma^{(1-p)(1+\epsilon)d} \lambda^p N\right\}^{-\frac{1}{2-p}} + c_4 N^{-1/2} \tau^{1/2} + c_5 N^{-1} \tau + c_6 \epsilon N^b + c_7 N^{-r_e - r_\mu}
\end{aligned}$$

and

$$\begin{aligned}
&\mathcal{R}^{\mathcal{D}}\left(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\right) - \mathcal{R}^{\mathcal{D}}\left(f_L^*, f_U^*; \mu^*, e^*\right) \\
&\leq c_1 \lambda \gamma^{-d} + c_2 \gamma^\beta + c_3 \left\{\gamma^{(1-p)(1+\epsilon)d} \lambda^p N\right\}^{-\frac{1}{2-p}} + c_4 N^{-1/2} \tau^{1/2} + c_5 N^{-1} \tau + c_6 \epsilon N^b + c_8 N^{-r_e - r_\mu}
\end{aligned}$$

where constants $c_1, \ldots, c_8$ do not depend on N.

Theorem 4.2 states that the estimated PDI with the cross-fitting procedure converges to the optimal PDI in terms of the excess risk bound. This convergence is subject to several regularity conditions, including the boundedness of the estimated nuisance components and the density of $\boldsymbol{X}$, the smoothness of the optimal PDI, the condition outlined in Lemma 4.1, and certain properties of the surrogate loss functions. If the hyperparameters are chosen with specific rates, specifically $\gamma \asymp N^{-1/(2\beta+d)}$, $\lambda \asymp N^{-(\beta+d)/(2\beta+d)}$, and $\epsilon \asymp N^{-\beta/(2\beta+d)-b}$ where b is the parameter satisfying (RC2) in Lemma 4.1, the excess risk bound has a rate of $O_P\big(N^{-r_e-r_\mu} + N^{-\beta/(2\beta+d)}\big)$, which has two leading terms. The first term is proportional to the product of the convergence rates of the nuisance components, which can achieve an $o_P(N^{-1/2})$ rate when both e^* and μ^* are estimated at $o_P(N^{-1/4})$ rates. These rates can be attained using various machine learning techniques, such as lasso (Belloni & Chernozhukov, 2011; 2013), random forests (Wager & Walther, 2016; Syrgkanis & Zampetakis, 2020), neural networks (Chen & White, 1999; Farrell et al., 2021), and boosting (Luo & Spindler, 2016), provided that e^* and μ^* exhibit sufficient smoothness. If at least one of the true nuisance functions were known, as in experimental settings where e^* was known, the first term would be zero. The second term is from the approximation error of using the ERM over the Gaussian RKHS. When the optimal PDI is sufficiently smooth compared to the dimension of the covariate, i.e., $\beta/d \simeq \infty$, the second term approaches $O_P(N^{-1/2})$.

While Theorem 4.2 characterizes the convergence of the PDI estimator in terms of excess risk, it does not ensure that the PDI estimator converges to the optimal PDI. To address this limitation, we directly analyze the convergence rate of the PDI estimator. We begin by making the following regularity condition to establish this convergence result:

(RC6) There exists a constant $\mathfrak{b} > 0$ and $\underline{L} > 0$ such that $f_U^*(\boldsymbol{X}) - f_L^*(\boldsymbol{X}) \geq 2\mathfrak{b}$, and

$$\left|\frac{\partial}{\partial a}\mu^*(a, \boldsymbol{X})\right| \geq \underline{L} \quad \text{for all } a \in [f_L^*(\boldsymbol{X}) - \mathfrak{b}, f_L^*(\boldsymbol{X}) + \mathfrak{b}] \cup [f_U^*(\boldsymbol{X}) - \mathfrak{b}, f_U^*(\boldsymbol{X}) + \mathfrak{b}] \ .$$

Regularity Condition (RC6) consists of two parts. The first part means that the optimal PDI has a positive length. The second part requires that the dose-response curve $\mu^*(\cdot, \boldsymbol{X})$ exhibits sufficient variation near the boundary points of the optimal PDI. Roughly speaking, the second part ensures that differences in the dose-response curve correspond to differences in the dose level, provided that the dose level is near the boundary points of the optimal PDI. This assumption is comparable to the conditions required in M-estimation, where the gradient of the moment equation must be non-zero near the true parameter value (e.g., Theorem 5.21 in van der Vaart (1998)).

Theorem 4.3 below establishes the convergence rate of the PDI estimator with respect to the $L_2(P)$-norm under the regularity conditions introduced above.

Theorem 4.3 *Suppose that the conditions of Theorem 4.2 hold, and Regularity Condition (RC6) also holds. Then, there exists a sequence δ_N such that $\delta_N \to 0$ as $N \to \infty$, and the following result holds with probability $P_{\boldsymbol{O}}^N$ not less than $1 - 3e^{-\tau} - \Delta_N - \delta_N$:*
For all fixed $\epsilon > 0$, $d/(d+\tau) < p < 1$, $\tau > 0$, and $\lambda > 0$, we have

$$\begin{aligned}
&E^{\mathcal{D}}\left[\left\{\widehat{f}_L^{(s)}(\boldsymbol{X}) - f_L^*(\boldsymbol{X})\right\}^2 + \left\{\widehat{f}_U^{(s)}(\boldsymbol{X}) - f_U^*(\boldsymbol{X})\right\}^2\right] \\
&\leq c_1'\lambda\gamma^{-d} + c_2'\gamma^{\beta} + c_3'\left\{\gamma^{(1-p)(1+\epsilon)d}\lambda^p N\right\}^{-\frac{1}{2-p}} + c_4' N^{-1/2}\tau^{1/2} + c_5' N^{-1}\tau + c_6'\epsilon N^b + c_7' N^{-r_e-r_\mu}
\end{aligned}$$

where constants $c_1', \ldots, c_7'$ do not depend on N.

As in Theorem 4.2, the convergence rate established in Theorem 4.3 follows the same rate, differing only in the constant terms. Consequently, the interpretation of Theorem 4.2 similarly applies to the convergence rate of the PDI estimator. In particular, if $\gamma \asymp N^{-1/(2\beta+d)}$, $\lambda \asymp N^{-(\beta+d)/(2\beta+d)}$, and $\epsilon \asymp N^{-\beta/(2\beta+d)-b}$, the convergence rate of the PDI estimator has a rate of $O_P\big(N^{-r_e-r_\mu} + N^{-\beta/(2\beta+d)}\big)$.

5 Simulation

We conducted a simulation study to assess the finite-sample performance of the proposed method based on the following data generating process with $N \in \{500, 1000, 1500, 2000\}$ observations. Each observation consisted of a 10-dimensional covariates $\boldsymbol{X}$, a dose level A, and a discretized outcome R. We generated the first four covariates, (X_1, X_2, X_3, X_4), from Unif$(0, 1)$, the next three, (X_5, X_6, X_7), from $N(0, 1)$, and the remaining three, (X_8, X_9, X_{10}), from Ber(0.5), with each covariate generated independently. We then generated A from $N\big(\sum_{j=1}^{10} X_j/15 + 0.2, 0.1^2\big)$, and R from Ber$(\nu(A, \boldsymbol{X}))$ where $\nu(a, \boldsymbol{x}) = \mathrm{expit}(-(300 + 30\sum_{j=1}^{10} x_j)(a - 0.45)^2 + 0.1\sum_{j=1}^{10} x_j + 2.5)$. We considered the range of the probability level α in $\{0.6, 0.65, 0.7\}$.

Using the generated data, we first estimated the PDI based on the approach proposed in Sections 3.2 and 3.3. Specifically, we estimated the propensity score and the dose-probability curve using parametric models that are widely used in practice. For the propensity score, we fitted a linear regression model of A regressed on $\boldsymbol{X}$ positing a normal error and obtained the density estimate based on the regression model, i.e., $\widehat{e}(A \,|\, \boldsymbol{X}) = \phi(A; \widehat{\eta}_A(\boldsymbol{X}), \widehat{\sigma}_A^2)$ where $\phi(\cdot; \eta, \sigma^2)$ is the density of $N(\eta, \sigma^2)$, and $\widehat{\eta}_A(\boldsymbol{X})$ and $\widehat{\sigma}_A^2$ are the estimated mean function and variance obtained from the regression model, respectively. We obtained the estimated dose-probability curve $\widehat{\mu}(A, \boldsymbol{X})$ based on a logistic regression model of R regressed on $(A, A^2, \boldsymbol{X})$. We remark that this parametric model is misspecified because the coefficient of A^2 in the logistic regression model is assumed to be constant, while it varies with $\boldsymbol{X}$ in the true model. However, we intentionally considered this misspecified model to reflect its widespread use in practice.

For the surrogate loss function bandwidth parameter, we chose $\epsilon = 10^{-3}$. The kernel bandwidth parameter γ was determined using the median heuristic described in Appendix A.3. The monotonicity-inducing regularization parameter κ in (8) was selected as $\kappa = 0$, which reduced to the optimization problem (5); we remark that even though the monotonicity-inducing term is not introduced, the proportion of resulting in non-monotonic bounds was very small; see the results below for details. The regularization parameter λ and the internal division ratio parameter ρ for initial points were selected through 5-fold cross-validation from the sets $\lambda \in \{2^{-4}, 2^0, 2^4\}$ and $\rho \in \{0, 0.5, 1\}$, respectively. The described PDI estimator is referred to as (Direct).

For comparison, we additionally considered the following two estimators. First, we obtained the indirect PDI estimator by following Section 3.1 based on the dose-probability curve estimator used in (Direct); this indirect estimator is referred to as (Ind-Para). Furthermore, we obtained an additional indirect PDI estimator where the dose-probability curve was estimated by an ensemble learner of random forest and gradient boosting (Friedman, 2001); this PDI estimator is referred to as (Ind-Ens). For indirect methods, we performed grid search to find an interval that made the estimated outcome regression greater than or equal to the probability level α, i.e., $[\widehat{f}_L(\boldsymbol{X}), \widehat{f}_U(\boldsymbol{X})] = \{a \,|\, \widehat{\mu}(a, \boldsymbol{X}) \geq \alpha\}$. If such an interval did not exist, we took the dose level maximizing the estimated dose-probability curve as the indirect rule recommendation, i.e., $\widehat{f}_L(\boldsymbol{X}) = \widehat{f}_U(\boldsymbol{X}) = \arg\max_a \widehat{\mu}(a, \boldsymbol{X})$. Lastly, for reference, we considered the optimal PDI, which is referred to as (True). We evaluated the performance of these PDI estimators using a test dataset $\mathcal{D}_{\text{test}}$ having $N \in \{500, 1000, 1500, 2000\}$ observations generated from the same distribution as the training data. We repeated the simulation 1000 times for each N.

To evaluate the performance of a PDI estimator $(\widehat{f}_L, \widehat{f}_U)$, we considered the following criteria. First, we obtained the empirical excess risk (EER) over the test set, which is defined by:

$$\mathrm{EER} = \frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{i \in \mathcal{D}_{\text{test}}} \left\{ \mathcal{L}(\boldsymbol{O}_i, \widehat{f}_L, \widehat{f}_U; \mu^*, e^*) - \mathcal{L}(\boldsymbol{O}_i, f_L^*, f_U^*; \mu^*, e^*) \right\} .$$

Note that the EER is an empirical analogue of the excess risk in Theorem 4.2. Of note, the EER of the optimal PDI serves as the reference value of zero. Second, we calculated the proportion of cases in which a method does not produce a valid PDI interval. Here, an invalid PDI means that at least one of the following occurs: the estimated lower bound exceeds the estimated upper bound, or one of the estimated bounds falls outside the admissible dose range and therefore requires post-hoc correction. Third, we focused on the mean

absolute error (MAE) and the root mean squared error (RMSE), which are defined as follows:

$$\text{MAE} = \frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{i \in \mathcal{D}_{\text{test}}} \left\{ |f_L^*(\boldsymbol{X}_i) - \widehat{f}_L(\boldsymbol{X}_i)| + |f_U^*(\boldsymbol{X}_i) - \widehat{f}_U(\boldsymbol{X}_i)| \right\} ,$$

$$\text{RMSE} = \left[\frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{i \in \mathcal{D}_{\text{test}}} \left[\{f_L^*(\boldsymbol{X}_i) - \widehat{f}_L(\boldsymbol{X}_i)\}^2 + \{f_U^*(\boldsymbol{X}_i) - \widehat{f}_U(\boldsymbol{X}_i)\}^2 \right] \right]^{1/2} .$$

For these four evaluation metrics (EER, proportion of invalid PDI, MAE, and RMSE), a lower value indicates better performance.

Lastly, we also evaluated the performance of the PDI estimators in terms of widely used classification performance measures. Specifically, for each PDI estimator, we defined the contingency table counts, true positives, true negatives, false positives, and false negatives as follows:

$$\text{TP} = \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{1}\left\{ R_i = 1, A_i \in [\widehat{f}_L(\boldsymbol{X}_i), \widehat{f}_U(\boldsymbol{X}_i)] \right\} , \quad \text{TN} = \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{1}\left\{ R_i = 0, A_i \notin [\widehat{f}_L(\boldsymbol{X}_i), \widehat{f}_U(\boldsymbol{X}_i)] \right\} ,$$

$$\text{FP} = \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{1}\left\{ R_i = 0, A_i \in [\widehat{f}_L(\boldsymbol{X}_i), \widehat{f}_U(\boldsymbol{X}_i)] \right\} , \quad \text{FN} = \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{1}\left\{ R_i = 1, A_i \notin [\widehat{f}_L(\boldsymbol{X}_i), \widehat{f}_U(\boldsymbol{X}_i)] \right\} .$$

Using these contingency table counts, we calculated accuracy, F1 score, Matthews correlation coefficient (MCC) (Matthews, 1975), recall, and Cohen's kappa (Cohen, 1960), which are defined as follows:

$$\begin{aligned}
\text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} , \\
\text{F1} &= \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} , \\
\text{MCC} &= \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})\}^{1/2}} , \\
\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} , \\
\text{Cohen's kappa} &= \frac{2 \times (\text{TP} \times \text{TN} - \text{FN} \times \text{FP})}{(\text{TP} + \text{FP}) \times (\text{FP} + \text{TN}) + (\text{TP} + \text{FN}) \times (\text{FN} + \text{TN})} .
\end{aligned}$$

A larger value of these classification measures indicates better performance with fewer type I and/or II errors.

The results of the simulation are summarized in Table 1. First, the EER of (Direct) is the lowest among the competing estimators, likely due to its construction based on empirical risk minimization. Moreover, it decreases as the sample size N increases, consistent with Theorem 4.2. Second, we focus on the proportion of invalid PDI estimates. (Direct) generates invalid PDI intervals only at negligible proportions, while the two indirect methods produce invalid intervals more often. Among these, (Ind-Ens) is particularly susceptible to this issue. As α increases, the proportion of invalid intervals rises because higher α values produce narrower PDIs and therefore leave less room for estimation error in the lower and upper bounds. Third, in terms of the MAE and RMSE, (Direct) uniformly outperforms the competing methods. We note that the EER, MAE, and RMSE cannot be calculated in real-world applications, as these measures require knowledge of the optimal PDI, which is not available. Lastly, in terms of classification performance measures, (Direct) uniformly exhibits higher accuracy, F1 score, MCC, and Cohen's kappa compared to the other methods. We further note that (True) generally yields the highest classification performance measures. Consequently, these measures, which can be calculated in real-world applications without knowledge of the optimal PDI, can serve as practical proxies for the EER, MAE, and RMSE in such scenarios.

Two additional observations are worth emphasizing. First, (Ind-Para) remains reasonably competitive despite misspecification because the posited logistic model still captures the dominant quadratic hormetic shape of the true dose-probability curve. However, extracting an interval by thresholding the estimated curve remains more sensitive to local bias than directly optimizing the proposed loss. Second, the substantially worse performance of (Ind-Ens) suggests that flexible nuisance estimation alone does not guarantee good interval

recovery: local irregularities in the estimated level set, together with the grid-search step used to extract the interval, can amplify non-monotonicity and increase the frequency of invalid PDIs. These comparisons clarify that the main practical benefit of the direct formulation is not merely improved curve fitting, but rather stable estimation of both bounds under a joint optimization criterion.

Combining all numerical results, we conclude that the proposed direct method (Direct) outperforms the other methods by almost always producing valid intervals, resulting in a smaller bias and fewer type I and II errors.

Table 1: Summary of the Simulation Results. Each number shows an average of over 1000 simulation repetitions. A lower value indicates better performance for the first four evaluation metrics, while a higher value indicates better performance for the latter five metrics. The best performance among the three competing methods is highlighted in bold text. The evaluation metrics for (True) is aggregated across all N.

α	Estimator	N	EER	Invalid PDI (%)	MAE (×1000)	RMSE (×1000)	Accuracy	F1	MCC	Recall	Cohen's kappa
0.6	(Ind-Para)	500	3.278	0.031	16.798	15.425	0.897	0.855	0.775	0.852	0.774
		1000	0.366	0.003	11.914	10.810	0.899	0.858	0.780	0.856	0.780
		1500	0.083	0.000	9.461	8.562	0.901	0.861	0.784	0.858	0.784
		2000	0.037	0.000	8.407	7.595	0.902	0.862	0.786	0.859	0.786
	(Ind-Ens)	500	75.191	0.749	20.894	22.913	0.891	0.844	0.762	0.827	0.760
		1000	30.929	0.307	18.150	18.963	0.893	0.846	0.765	0.829	0.764
		1500	19.378	0.192	17.177	17.706	0.894	0.848	0.767	0.828	0.767
		2000	13.396	0.132	16.287	16.533	0.895	0.849	0.769	0.829	0.768
	(Direct)	500	**0.184**	**0.001**	**13.895**	**12.977**	**0.899**	**0.858**	**0.779**	**0.856**	**0.779**
		1000	**0.070**	**0.000**	**10.045**	**9.314**	**0.900**	**0.860**	**0.783**	**0.859**	**0.783**
		1500	**0.041**	**0.000**	**8.088**	**7.432**	**0.902**	**0.862**	**0.786**	**0.860**	**0.786**
		2000	**0.016**	**0.000**	**7.267**	**6.645**	**0.902**	**0.863**	**0.787**	**0.861**	**0.787**
	(True)		0.000	0.000	0.000	0.000	0.903	0.864	0.788	0.861	0.788
0.65	(Ind-Para)	500	7.222	0.071	17.543	16.179	0.893	0.847	0.766	0.827	0.765
		1000	0.658	0.006	12.394	11.259	0.896	0.850	0.771	0.831	0.770
		1500	0.100	0.000	9.833	8.902	0.897	0.853	0.775	0.832	0.774
		2000	0.047	0.000	8.724	7.880	0.898	0.853	0.776	0.833	0.775
	(Ind-Ens)	500	179.484	1.792	24.247	27.649	0.883	0.826	0.741	0.786	0.738
		1000	90.584	0.904	20.814	23.002	0.885	0.831	0.747	0.790	0.744
		1500	62.496	0.623	19.398	21.273	0.886	0.832	0.750	0.790	0.747
		2000	46.383	0.462	18.242	19.790	0.888	0.834	0.752	0.793	0.749
	(Direct)	500	**0.275**	**0.002**	**14.348**	**13.476**	**0.895**	**0.850**	**0.770**	**0.831**	**0.769**
		1000	**0.090**	**0.000**	**10.184**	**9.474**	**0.897**	**0.852**	**0.774**	**0.835**	**0.773**
		1500	**0.028**	**0.000**	**8.108**	**7.502**	**0.898**	**0.854**	**0.777**	**0.835**	**0.776**
		2000	**0.026**	**0.000**	**7.220**	**6.662**	**0.899**	**0.855**	**0.778**	**0.836**	**0.777**
	(True)		0.000	0.000	0.000	0.000	0.899	0.855	0.779	0.835	0.778
0.7	(Ind-Para)	500	15.677	0.155	18.581	17.232	0.888	0.835	0.753	0.796	0.750
		1000	1.437	0.014	13.094	11.929	0.890	0.838	0.757	0.800	0.755
		1500	0.171	0.001	10.409	9.442	0.891	0.840	0.761	0.800	0.758
		2000	0.096	0.001	9.235	8.356	0.892	0.841	0.762	0.801	0.760
	(Ind-Ens)	500	409.985	4.097	29.186	33.773	0.869	0.798	0.711	0.731	0.703
		1000	228.890	2.286	24.775	28.383	0.874	0.806	0.721	0.740	0.714
		1500	173.203	1.730	22.839	26.199	0.875	0.809	0.724	0.742	0.717
		2000	133.861	1.337	21.328	24.364	0.877	0.812	0.728	0.746	0.721
	(Direct)	500	**0.499**	**0.004**	**14.799**	**14.005**	**0.890**	**0.839**	**0.759**	**0.802**	**0.756**
		1000	**0.067**	**0.000**	**10.587**	**9.892**	**0.891**	**0.841**	**0.761**	**0.804**	**0.759**
		1500	**0.041**	**0.000**	**8.328**	**7.760**	**0.893**	**0.842**	**0.763**	**0.804**	**0.761**
		2000	**0.020**	**0.000**	**7.392**	**6.872**	**0.893**	**0.843**	**0.764**	**0.805**	**0.762**
	(True)		0.000	0.000	0.000	0.000	0.893	0.843	0.765	0.803	0.762

6 Application: Warfarin Dosing

We applied our method to explore the two-sided PDI in the context of warfarin dosing. Warfarin is a medicine for preventing harmful blood clots from forming and growing bigger. It is particularly useful for

patients suffering from conditions that can cause fatal blood clots such as stroke and heart attack. However, excessive doses of warfarin can lead to severe complications, including life-threatening bleeding and tissue death. Therefore, it is essential to prescribe warfarin in a suitable therapeutic dose to ensure its benefits to patients. As recommended by the American Heart Association, warfarin should be prescribed to keep a patient's international normalized ratio (INR) within a desired range, typically between 2 and 3 (January et al., 2014). The INR level increases with the warfarin dose, leading to a dose-probability curve with an inverted U-shaped pattern, reflecting the hormetic response (Blann et al., 2003).

We used the dataset presented in The International Warfarin Pharmacogenetics Consortium (2009) to estimate the PDI of warfarin. Specifically, each patient's target INR range, the reported INR value, and the warfarin dose measured in mg/week were considered as the desired range $\mathcal{T}$, the observed outcome Y, and the observed dose level A, respectively. Accordingly, $R = \mathbb{1}(Y \in \mathcal{T})$ indicates whether the INR of a patient was within the desired range. Furthermore, we included the following pharmacogenetic and clinical variables as pre-treatment covariates $\boldsymbol{X}$: gender (male/female), weight (in kg), height (in cm), age (nine discretized levels), and Amiodarone receipt status (took Amiodarone/did not take Amiodarone). Although our framework allows the desired range $\mathcal{T}$ to depend on covariates, we restricted our analysis to Asians because the target INR ranges and empirical dose-response patterns differed substantially across racial groups, making a common specification of the nuisance functions and a common hormetic shape assumption less credible in the pooled sample. The subgroup restriction therefore served to create a more homogeneous population in which Assumption 4 was empirically more plausible and overlap was better behaved. Lastly, we discarded two observations with outlying dose levels greater than 80 mg/week to guarantee the positivity assumption 3. Consequently, we used 1407 patients in the analysis. In this application, the PDI is interpreted as a patient-specific interval of weekly warfarin doses such that, if an Asian patient were assigned any dose within that interval, the probability that the patient's INR falls within the target range would be at least α. Thus, the lower and upper bounds of the PDI represent personalized dose thresholds within which warfarin treatment is expected to be therapeutically acceptable with sufficiently high probability.

In our analysis, we implemented the proposed direct method, which corresponds to (Direct) in the simulation study, as follows. We split the sample into two sets, with 1000 patients in the training set $\mathcal{D}_{\text{train}}$ and 407 patients in the test set $\mathcal{D}_{\text{test}}$. We repeated the split 100 times, and for each iteration, we used the training set to estimate the PDI, and the test set to evaluate the estimated PDI. Using the training set, we estimated the propensity score based on the linear regression model of $\log(A)$ on $\boldsymbol{X}$ with normal error and the dose-probability curve based on the logistic regression model of R on A, A^2, and $\boldsymbol{X}$, which are denoted by $\widehat{e}$ and $\widehat{\mu}$, respectively. We note that $\log(A)$ was used in the propensity score model because it suggested better model diagnostics; see Section A.4 for details. We then estimated the PDI from the ERM (4) where the hyperparameters were chosen based on the same procedure as in the simulation study. The probability level α was chosen from $\{0.6, 0.65, 0.7\}$.

For comparison, we also implemented the two indirect methods (Ind-Para) and (Ind-Ens) considered in the simulation study.

In Table 2, we report the proportion of producing an invalid PDI and the classification performance measures that were used in the simulation. First, we find that the proportion of invalid PDI estimates follows a similar pattern to the simulation study: invalid PDI estimates are more frequent with the indirect methods than with the direct method. Second, the classification performance measures suggest that the direct method yields more accurate estimates. Specifically, the direct method uniformly shows higher values. Combining all results from both simulation and data analysis, the direct method can be a valuable tool for practitioners seeking to determine the therapeutic dose of warfarin.

We conclude the data analysis by providing the summary of the PDI estimates obtained from the proposed direct method at $\alpha = 0.7$. Table 3 summarizes the relationship between the PDI estimates and two covariates, age and gender. In terms of age, we find that the PDI interval is wider for younger patients and narrower for older patients. This result is consistent with previous medical studies which have established a negative correlation between age and the therapeutic dose range of warfarin (Khoury & Sheikh-Taha, 2014; Shendre et al., 2018). In terms of gender, we find that the PDI interval is wider for female patients, but the relationship between gender and the therapeutic warfarin dose range remains uncertain based on previous medical studies.

Table 2: Summary of the Warfarin Dosing Analysis. Each number shows an average of over 100 data splits. A Lower value indicates better performance for the proportion of invalid PDI, while a higher value indicates better performance for the latter five metrics. The best performance among the three competing methods is highlighted in bold text.

α	Estimator	Invalid PDI (%)	Accuracy	F1	MCC	Recall	Cohen's kappa
0.6	(Ind-Para)	0.020	0.730	0.834	0.147	0.913	0.130
	(Ind-Ens)	1.401	0.655	0.768	0.089	0.774	0.085
	(Direct)	**0.000**	**0.743**	**0.845**	**0.168**	**0.939**	**0.136**
0.65	(Ind-Para)	0.052	0.706	0.811	0.161	0.846	0.156
	(Ind-Ens)	1.845	0.605	0.715	0.071	0.676	0.068
	(Direct)	**0.000**	**0.724**	**0.826**	**0.178**	**0.879**	**0.169**
0.7	(Ind-Para)	0.216	0.667	0.769	0.176	0.744	0.174
	(Ind-Ens)	2.961	0.564	0.660	0.099	0.574	0.090
	(Direct)	**0.012**	**0.684**	**0.786**	**0.177**	**0.782**	**0.176**

Some studies did not find significant differences in the therapeutic warfarin dose range between male and female patients (Poli et al., 2009; Miao et al., 2007), while others suggested that gender is an important factor, especially for Asian patients (Choi et al., 2011; Liew et al., 2013). Our analysis, which only included Asian patients, seems to support the latter view. However, more research is needed to establish the role of gender in determining the therapeutic dose of warfarin.

Table 3: Average of the PDI Estimates at $\alpha = 0.7$ Obtained from 100 Data Split (unit: mg/week).

Gender	Age						
	Under 29	30-39	40-49	50-59	60-69	70-79	Over 80
Female	(10.2, 73.0)	(11.3, 71.7)	(11.9, 71.1)	(12.7, 70.2)	(13.6, 69.2)	(14.6, 68.2)	(15.4, 67.5)
Male	(12.3, 70.3)	(13.3, 69.4)	(14.2, 68.6)	(15.1, 67.7)	(16.1, 66.6)	(17.3, 65.3)	(19.0, 63.7)

7 Concluding Remarks

In this paper, we present a method for estimating the personalized two-sided PDI. To directly estimate the PDI from an ERM, we develop a loss function that is robust to misspecification of the nuisance components (the propensity score and the dose-probability curve), and design a computationally tractable surrogate loss function for obtaining PDI estimators over the product RKHS. We show that the excess risk bound of our estimated PDI converges to the true optimal PDI at a rate that depends on the estimation error of the nuisance parameters and the approximation error of the PDI using functions within the product RKHS. The simulation results show that our direct method can yield PDIs that are more likely to be valid and more accurate in terms of bias and classification performance measures when compared to indirect methods. Although optimal single-dose rules and personalized dose intervals are both useful summaries of individualized treatment effects, they answer different scientific questions. A single-dose rule identifies the dose maximizing a target criterion, whereas a PDI identifies the entire set of doses that achieve a clinically acceptable outcome level. In that sense, a single recommended dose or the midpoint of an estimated PDI may serve as a convenient summary, but it cannot replace the interval itself because it does not reveal the range of feasible doses available to a practitioner.

The proposed methodology offers several opportunities for extension. First, our framework is well-suited to settings with a moderate number of covariates, as illustrated by the two real-world applications in Section 6, involving 5 and 11 variables, respectively. To increase its versatility, the method could be extended to accommodate high-dimensional covariates, expanding its applicability to a broader range of scenarios. Second, many real-world datasets are collected in online settings, where the optimal decision rule must be updated sequentially as new data arrive. While our framework is developed for an offline setting, which aligns with the structure of the two real-world datasets analyzed, it would be valuable to extend the framework to an online setting and investigate the corresponding regret bounds. Third, while the ERM in (4) is solved within the RKHS framework, alternative function spaces—such as those generated by (deep) neural networks (LeCun et al., 2015; Goodfellow et al., 2016), random forests (Breiman, 2001), or generalized additive models

(Hastie & Tibshirani, 1986)—could be used in principle. However, establishing the convergence analysis of the PDI estimates in these alternative spaces requires substantial effort, as it involves distinct approximation theories that differ fundamentally from the RKHS approximation theory. Furthermore, the parameterization of these spaces differs from that of the RKHS, which may render the proposed DC algorithm incapable of preserving convexity under their parameterization. Therefore, we defer the exploration of these directions to future work. Lastly, we also note that our method relies on the hormesis, i.e., Assumption 4. This assumption restricts the shape of the dose-probability curve, but can be empirically verified in certain cases, such as the data examples in Section 6. However, in cases where this assumption is difficult to verify due to insufficient observations or heterogeneous data sources, an investigator should be aware of this assumption and, if applicable, consider focusing on uncovering the mechanism between treatment and outcome through randomized experiments.

We revisit the distinction between two approaches to personalized dose recommendations in continuous treatment. The first focuses on optimal single-dose rules that maximize outcomes (e.g., Chen et al. (2016)), while the second targets therapeutic dose intervals that achieve outcomes within a desired range. The former is often preferred due to its interpretability and ease of estimation. Nevertheless, interval-based methods are equally important in many applications, as illustrated in our motivating examples. From a clinical or policy perspective, recommending a single fixed dose may be impractical when considerations such as feasibility, adherence, toxicity, or implementation constraints require flexibility. In such settings, a PDI identifies the set of doses that yield clinically acceptable outcomes rather than replacing clinical judgment with a fully automated decision. Accordingly, our loss function evaluates whether observed outcomes support a candidate interval based on whether the administered dose lies within the interval and whether the outcome falls in the desired set, allowing the method to learn which intervals are compatible with favorable outcomes.

Building on this perspective, we conclude by outlining how PDIs can be used in practice and how they complement point-based approaches. To initiate treatment, one may estimate both the optimal single-dose rule and the PDI. By increasing α to impose a more stringent criterion, one can assess whether the point recommendation remains within the PDI. If it does, the interval-based analysis reinforces the recommendation by indicating robustness to stricter thresholds; if not, it suggests weaker support, and the clinician may adjust toward doses that remain within the admissible interval. While a single dose or the midpoint of a PDI provides a convenient summary, it cannot replace the interval itself, as it does not convey the range of feasible options or the robustness of the recommendation. Thus, interval-based approaches complement single-dose rules by offering flexibility, transparency, and practical value in real-world decision-making.

References

Walter Ageno, Alexander S Gallus, Ann Wittkowsky, Mark Crowther, Elaine M Hylek, and Gualtiero Palareti. Oral anticoagulant therapy: antithrombotic therapy and prevention of thrombosis: American college of chest physicians evidence-based clinical practice guidelines. *Chest*, 141(2):e44S–e88S, 2012.

Daniel Albalate, Germà Bel, and Jordi Rosell. An evaluation of optimal scale and jurisdiction size to improve efficiency in metropolitan bus systems. *Research in Transportation Business & Management*, 46:100822, 2023.

American Association of Clinical Endocrinologists and Others. Management of common comorbidities of diabetes, 2019.

American Diabetes Association. 13. Older Adults: Standards of Care in Diabetes—2024. *Diabetes Care*, 47 (Supplement_1):S244–S257, 2023.

Jeffrey L. Anderson, Benjamin D. Horne, Scott M. Stevens, Amanda S. Grove, Stephanie Barton, Zachery P. Nicholas, Samera F.S. Kahn, Heidi T. May, Kent M. Samuelson, Joseph B. Muhlestein, and John F. Carlquist. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. *Circulation*, 116(22):2563–2570, 2007.

Philipp Bach, Malte S. Kurz, Victor Chernozhukov, Martin Spindler, and Sven Klaassen. Doubleml: An object-oriented implementation of double machine learning in r. *Journal of Statistical Software*, 108(3): 1–56, 2024.

Heejung Bang and James M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Alexandre Belloni and Victor Chernozhukov. ℓ_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82 – 130, 2011.

Alexandre Belloni and Victor Chernozhukov. Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521 – 547, 2013.

Andrew D Blann, David A Fitzmaurice, and Gregory Y H Lip. Anticoagulation in hospitals and general practice. *BMJ*, 326(7381):153–156, 2003.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Hengrui Cai, Chengchun Shi, Rui Song, and Wenbin Lu. Jump interval-learning for individualized decision making with continuous treatments. *Journal of Machine Learning Research*, 24(140):1–92, 2023. URL http://jmlr.org/papers/v24/21-0843.html.

Edward J. Calabrese. Hormesis and medicine. *British Journal of Clinical Pharmacology*, 66(5):594–617, 2008.

Edward J Calabrese and Mark P Mattson. How does hormesis impact biology, toxicology, and medicine? *NPJ aging and mechanisms of disease*, 3(1):1–8, 2017.

Bibhas Chakraborty and Erica E. M. Moodie. *Statistical Methods for Dynamic Treatment Regimes*, volume 2. Springer, New York, 2013.

Guanhua Chen, Donglin Zeng, and Michael R. Kosorok. Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association*, 111(516):1509–1521, 2016.

Guanhua Chen, Xiaomao Li, and Menggang Yu. Policy learning for optimal individualized dose intervals. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 1671–1693. PMLR, 28–30 Mar 2022.

Xiaohong Chen and Halbert White. Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Transactions on Information Theory*, 45(2):682–691, 1999.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Hugh A. Chipman, Edward I. George, Robert E. McCulloch, and Thomas S. Shively. mBART: Multidimensional Monotone BART. *Bayesian Analysis*, 17(2):515 – 544, 2022.

Jung Ran Choi, Jeong-Oh Kim, Dae Ryong Kang, Seong Yoon, Jung-Young Shin, XiangHua Zhang, Mee Ork Roh, Hyung Joo Hong, Young-Pil Wang, Keon-Hyon Jo, et al. Proposal of pharmacogenetics-based warfarin dosing algorithm in korean patients. *Journal of Human Genetics*, 56(4):290–295, 2011.

Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1):37–46, 1960.

Tamlin S Conner, Aimee C Richardson, and Jody C Miller. Optimal serum selenium concentrations are associated with lower depressive symptoms and negative mood among young adults1, 2, 3. *The Journal of Nutrition*, 145(1):59–65, 2015.

Ralph Cook and Edward J. Calabrese. The importance of hormesis to public health. *Environmental Health Perspectives*, 114(11):1631–1635, 2006. doi: 10.1289/ehp.8606. URL https://ehp.niehs.nih.gov/doi/abs/10.1289/ehp.8606.

Mona Eberts and Ingo Steinwart. Optimal regression rates for svms using gaussian kernels. *Electronic Journal of Statistics*, 7:1–42, 2013.

Max H. Farrell, Tengyuan Liang, and Sanjog Misra. Deep neural networks for estimation and inference. *Econometrica*, 89(1):181–213, 2021.

John M. Flack and Bemi Adekola. Blood pressure and the new acc/aha hypertension guidelines. *Trends in Cardiovascular Medicine*, 30(3):160–164, 2020.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232, 2001.

Damien Garreau, Wittawat Jitkrittum, and Motonobu Kanagawa. Large sample analysis of the median heuristic. *Preprint arXiv:1707.07269*, 2017.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

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

Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical Science*, 1(3):297 – 310, 1986. URL https://doi.org/10.1214/ss/1177013604.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, 2 edition, 2009.

David Haussler. Decision theoretic generalizations of the pac model for neural net and other learning applications. *Information and Computation*, 100(1):78–150, 1992.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton, 2020.

Keisuke Hirano and Guido W. Imbens. The propensity score with continuous treatments. In Andrew Gelman and Xiao-Li Meng (eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, chapter 7, pp. 73–84. John Wiley & Sons, Ltd, New York, 2004.

Joey Hong, Aviral Kumar, and Sergey Levine. Confidence-conditioned value functions for offline reinforcement learning. *Preprint arXiv:2212.04607*, 2022.

William Hua, Hongyuan Mei, Sarah Zohar, Magali Giral, and Yanxun Xu. Personalized dynamic treatment regimes in continuous time: a bayesian approach for optimizing clinical decisions with timing. *Bayesian Analysis*, 17(3):849–878, 2022.

Guido W. Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87 (3):706–710, 2000.

Craig T January, L Samuel Wann, Joseph S Alpert, Hugh Calkins, Joaquin E Cigarroa, Joseph C Cleveland, Jamie B Conti, Patrick T Ellinor, Michael D Ezekowitz, Michael E Field, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: "A" report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *Journal of the American College of Cardiology*, 64(21):e1–e76, 2014.

Leslie Pack Kaelbling. *Learning in Embedded Systems*. The MIT Press, Cambridge, 1993.

Nathan Kallus and Angela Zhou. Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics*, pp. 1243–1251, 2018.

Nikos Karampatziakis, John Langford, and Paul Mineiro. Empirical likelihood for contextual bandits. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20. Curran Associates Inc., 2020. ISBN 9781713829546.

Ghada Khoury and Marwan Sheikh-Taha. Effect of age and sex on warfarin dosing. *Clinical Pharmacology: Advances and Applications*, 6:103, 2014.

George S. Kimeldorf and Grace Wahba. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *Annals of Mathematical Statistics*, 41(2):495–502, 1970.

Eric B Laber and Ying-Qi Zhao. Tree-based methods for individualized treatment regimes. *Biometrika*, 102 (3):501–514, 2015.

Tze Leung Lai. Adaptive treatment allocation and the multi-armed bandit problem. *The Annals of Statistics*, 15(3):1091 – 1114, 1987.

Hoai An Le Thi and Tao Pham Dinh. Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. *Journal of Global Optimization*, 11(3):253–285, 1997.

Hoai An Le Thi and Tao Pham Dinh. Dc programming and dca: thirty years of developments. *Mathematical Programming*, 169(1):5–68, 2018.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

Gaofeng Li, Yexiong Li, Junjie Wang, Xianshu Gao, Qiuzi Zhong, Liru He, Chunmei Li, Ming Liu, Yueping Liu, Mingwei Ma, Hao Wang, Xuan Wang, and Hui Zhu. Guidelines for radiotherapy of prostate cancer (2020 edition). *Precision Radiation Oncology*, 5(3):160–182, 2021.

Qi Li and Jeffrey S Racine. Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, 26(4):423–434, 2008.

Xiaomao Li. *Optimal recommendation of individual dose intervals*. PhD thesis, University of Wisconsin–Madison, 2018.

Chooi-Lan Liew, Jui-Hung Yen, An-Bang Liu, and Ingrid Y Liu. Sex differences in the effective warfarin dosage in han and aboriginal taiwanese patients with the "vkorc1-1639aa" genotype. *Tzu Chi Medical Journal*, 25(4):213–217, 2013.

Ang Liu, Yuguang Wei, Qi Xiu, Hao Yao, and Jia Liu. How learning time allocation make sense on secondary school students' academic performance: A chinese evidence based on pisa 2018. *Behavioral Sciences*, 13 (3), 2023.

Alexander R. Luedtke and Mark J. van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, 44(2):713 – 742, 2016.

Jared K. Lunceford and Marie Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23(19):2937–2960, 2004.

Ye Luo and Martin Spindler. High-dimensional $l_2$boosting: Rate of convergence. *arXiv preprint*, 2016.

Markus Luster, SE Clarke, M Dietlein, M Lassmann, P Lind, WJG Oyen, Jan Tennvall, and E Bombardieri. Guidelines for radioiodine therapy of differentiated thyroid cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 35:1941–1959, 2008.

B.W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2):442–451, 1975.

Mark P. Mattson. Hormesis defined. *Ageing Research Reviews*, 7(1):1–7, 2008.

Liyan Miao, Jian Yang, Chenrong Huang, and Zhenya Shen. Contribution of age, body weight, and cyp2c9 and vkorc1 genotype to the anticoagulant response to warfarin: proposal for a new dosing regimen in chinese patients. *European Journal of Clinical Pharmacology*, 63(12):1135–1141, 2007.

Erica EM Moodie, Bibhas Chakraborty, and Michael S Kramer. Q-learning for estimating optimal dynamic treatment rules from observational data. *Canadian Journal of Statistics*, 40(4):629–645, 2012.

Georg Northoff and Shankar Tumati. "average is good, extremes are bad" – non-linear inverted u-shaped relationship between neural mechanisms and functionality of mental features. *Neuroscience & Biobehavioral Reviews*, 104:11–25, 2019.

Chan Park, Guanhua Chen, Menggang Yu, and Hyunseung Kang. Minimum resource threshold policy under partial interference. *Journal of the American Statistical Association*, 119(548):2881–2894, 2024.

Daniela Poli, Emilia Antonucci, Elisa Grifoni, Rosanna Abbate, Gian Franco Gensini, and Domenico Prisco. Gender differences in stroke risk of atrial fibrillation patients on oral anticoagulant treatment. *Thromb Haemost*, 101(05):938–942, 2009.

Amir Qaseem, Timothy J Wilt, Robert Rich, Linda L Humphrey, Jennifer Frost, Mary Ann Forciea, Clinical Guidelines Committee of the American College of Physicians, the Commission on Health of the Public, and Science of the American Academy of Family Physicians. Pharmacologic treatment of hypertension in adults aged 60 years or older to higher versus lower blood pressure targets: A clinical practice guideline from the american college of physicians and the american academy of family physicians. *Annals of Internal Medicine*, 166(6):430–437, 2017.

Min Qian and Susan A. Murphy. Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180 – 1210, 2011.

B. D. Ripley. Neural networks and related methods for classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3):409–456, 1994.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688, 1974.

Donald B. Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58, 1978.

Areo G Saffarzadeh, Maureen Canavan, Benjamin J Resio, Samantha L Walters, Kaitlin M Flores, Roy H Decker, and Daniel J Boffa. Optimal radiation dose for stage iii lung cancer—should "definitive" radiation doses be used in the preoperative setting? *JTO Clinical and Research Reports*, 2(8):100201, 2021.

Daniel O. Scharfstein, Andrea Rotnitzky, and James M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.

Anton Schick. On asymptotically efficient estimation in semiparametric models. *The Annals of Statistics*, 14(3):1139–1151, 1986.

Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. A generalized representer theorem. In David Helmbold and Bob Williamson (eds.), *Computational Learning Theory*, pp. 416–426. Springer, Berlin, Heidelberg, 2001.

Juliana Schulz and Erica E. M. Moodie. Doubly robust estimation of optimal dosing strategies. *Journal of the American Statistical Association*, 116(533):256–268, 2021.

Ingmar Schuster, Mattes Mollenhauer, Stefan Klus, and Krikamol Muandet. Kernel conditional density operators. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 993–1004. PMLR, 2020.

Aditi Shendre, Gaurav M. Parmar, Chrisly Dillon, Timothy Mark Beasley, and Nita A. Limdi. Influence of age on warfarin dose, anticoagulation control, and risk of hemorrhage. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 38(6):588–596, 2018.

Chengchun Shi, Wenbin Lu, and Rui Song. Breaking the curse of nonregularity with subagging — inference of the mean outcome under optimal treatment regimes. *Journal of Machine Learning Research*, 21(176): 1–67, 2020.

Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer-Verlag, New York, 2008.

Kathleen Stergiopoulos and David L. Brown. Genotype-Guided vs Clinical Dosing of Warfarin and Its Analogues: Meta-analysis of Randomized Clinical Trials. *JAMA Internal Medicine*, 174(8):1330–1338, 2014.

Charles J. Stone. Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, 8 (6):1348–1360, 1980.

Alexander L. Strehl and Michael L. Littman. A theoretical analysis of model-based interval estimation. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, pp. 856–863, New York, NY, USA, 2005. Association for Computing Machinery.

Alexander L. Strehl and Michael L. Littman. An analysis of model-based interval estimation for markov decision processes. *Journal of Computer and System Sciences*, 74(8):1309–1331, 2008.

Vasilis Syrgkanis and Manolis Zampetakis. Estimation and inference with trees and forests in high dimensions. *Preprint arXiv:2007.03210*, 2020.

The International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 360(8):753–764, 2009. doi: 10.1056/NEJMoa0809329. URL https://doi.org/10.1056/NEJMoa0809329. PMID: 19228618.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer New York, New York, NY, 2009.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, 1998.

Aad W van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, 1996.

J Wadsley, N Armstrong, V Bassett-Smith, M Beasley, R Chandler, L Cluny, AJ Craig, K Farnell, Karcez Garcez, N Garnham, et al. Patient preparation and radiation protection guidance for adult patients undergoing radioiodine treatment for thyroid cancer in the uk. *Clinical Oncology*, 35(1):42–56, 2023.

Stefan Wager and Guenther Walther. Adaptive concentration of regression trees, with application to random forests. *arXiv preprint*, 2016.

Chang Wang and Lu Wang. Non-greedy tree-based learning for estimating global optimal dynamic treatment decision rules with continuous treatment dosage. *arXiv preprint arXiv:2302.02015*, 2023.

Daniel M Witt, Nathan P Clark, Scott Kaatz, Terri Schnurr, and Jack E Ansell. Guidance for the practical management of warfarin therapy in the treatment of venous thromboembolism. *Journal of thrombosis and thrombolysis*, 41:187–205, 2016.

Stephen J Wright. Coordinate descent algorithms. *Mathematical Programming*, 151(1):3–34, 2015.

Yichao Wu and Yufeng Liu. Robust truncated hinge loss support vector machines. *Journal of the American Statistical Association*, 102(479):974–983, 2007.

Wenzhuo Zhou, Ruoqing Zhu, and Donglin Zeng. A parsimonious personalized dose-finding model via dimension reduction. *Biometrika*, 108(3):643–659, 2021.

Liangyu Zhu, Wenbin Lu, Michael R Kosorok, and Rui Song. Kernel assisted learning for personalized dose finding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 56–65, 2020.

Yeying Zhu, Donna L. Coffman, and Debashis Ghosh. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *Journal of Causal Inference*, 3(1):25–40, 2015.

Joshua Graff Zivin and Matthew Neidell. Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26, 2014.

Appendix

A Details of the Main Paper

A.1 Existence of the Optimal PDI

We discuss the existence of the optimal PDI, (f_L^*, f_U^*), under Assumptions 4 and 5. It is trivial that $a_{\boldsymbol{X}}$ in 4 belongs to the level set $\{a \,|\, \mu^*(a, \boldsymbol{X}) \geq \alpha\}$, which is uniquely determined based on Assumption 5. We now show that $\{a \,|\, \mu^*(a, \boldsymbol{X}) \geq \alpha\}$ is not a point, but an interval. Let L_μ be the Lipschitz constant of $\mu^*(\cdot, \boldsymbol{X})$, i.e., $\big|\mu^*(a_1, \boldsymbol{X}) - \mu^*(a_2, \boldsymbol{X})\big| \leq L_\mu \big|a_1 - a_2\big|$. Since $\mu^*(a_{\boldsymbol{X}}, \boldsymbol{X}) \geq \alpha + c_\mu$, we find $\mu^*(a_{\boldsymbol{X}} - c_\mu/L_\mu, \boldsymbol{X}) > \alpha$ and $\mu^*(a_{\boldsymbol{X}} + c_\mu/L_\mu, \boldsymbol{X}) > \alpha$, indicating that $[a_{\boldsymbol{X}} - c_\mu/L_\mu, a_{\boldsymbol{X}} + c_\mu/L_\mu] \subseteq \{a \,|\, \mu^*(a, \boldsymbol{X}) \geq \alpha\}$. As a result, we find

$$f_L^*(\boldsymbol{X}) = \inf_a \{a \,|\, \mu^*(a, \boldsymbol{X}) \geq \alpha\} \leq a_{\boldsymbol{X}} - \frac{c_\mu}{L_\mu}\,, \qquad f_U^*(\boldsymbol{X}) = \sup_a \{a \,|\, \mu^*(a, \boldsymbol{X}) \geq \alpha\} \geq a_{\boldsymbol{X}} + \frac{c_\mu}{L_\mu}\,,$$

indicating $f_L^*(\boldsymbol{X}) < f_U^*(\boldsymbol{X})$.

A.2 Details of the AIPW Loss Function

Recall that

$$\begin{aligned}
&\mathcal{L}_{\mathrm{AIPW}}(\boldsymbol{O}, f_L, f_U; \mu, e) \\
&= \alpha \bigg[\frac{\{\mu(A, \boldsymbol{X}) - R\} \mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}}{e(A \,|\, \boldsymbol{X})} + \int \{1 - \mu(a, \boldsymbol{X})\} \mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \bigg] \\
&+ (1 - \alpha) \bigg[\frac{\{R - \mu(A, \boldsymbol{X})\} \mathbb{1}\{A \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}}{e(A \,|\, \boldsymbol{X})} + \int \mu(a, \boldsymbol{X}) \mathbb{1}\{a \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \bigg]\,,
\end{aligned}$$

We first introduce the surrogate loss function of $\mathcal{L}_{\mathrm{AIPW}}$.

$$\begin{aligned}
&\mathcal{L}_{\mathrm{AIPW}}(\boldsymbol{O}, f_L, f_U; \mu, e) \\
&= \alpha \bigg[\frac{\{\mu(A, \boldsymbol{X}) - R\} \Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)}{e(A \,|\, \boldsymbol{X})} + \int \{1 - \mu(a, \boldsymbol{X})\} \mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \bigg] \\
&+ (1 - \alpha) \bigg[\frac{\{R - \mu(A, \boldsymbol{X})\} \{1 - \Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)\}}{e(A \,|\, \boldsymbol{X})} + \int \mu(a, \boldsymbol{X}) \mathbb{1}\{a \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \bigg]\,,
\end{aligned}$$

Here $\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}$ is replaced with $\Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big)$.

We find the two IPW terms in $\mathcal{L}_{\mathrm{AIPW}}$ are

$$\begin{aligned}
&\frac{\alpha\{\mu(A, \boldsymbol{X}) - R\} \mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\} + (1 - \alpha)\{R - \mu(A, \boldsymbol{X})\} \mathbb{1}\{A \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}}{e(A \,|\, \boldsymbol{X})} \\
&= \frac{\alpha\{\mu(A, \boldsymbol{X}) - R\} \mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\} - (1 - \alpha)\{\mu(A, \boldsymbol{X}) - R\} \big[1 - \mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\big]}{e(A \,|\, \boldsymbol{X})} \\
&= \frac{\{\mu(A, \boldsymbol{X}) - R\} \big[-1 + \alpha + \mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\big]}{e(A \,|\, \boldsymbol{X})} \\
&= \frac{\{\mu(A, \boldsymbol{X}) - R\} \big[\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\big]}{e(A \,|\, \boldsymbol{X})} + B_1(\boldsymbol{O})
\end{aligned}$$

where $B_1(\boldsymbol{O}) = \{\mu(A, \boldsymbol{X}) - R\}(\alpha - 1)/e(A \,|\, \boldsymbol{X})$. The two outcome-integrated terms are

$$
\begin{aligned}
&\alpha \int \big\{1 - \mu(a, \boldsymbol{X})\big\} \mathbb{1}\big\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\, da + (1-\alpha) \int \mu(a, \boldsymbol{X}) \mathbb{1}\big\{a \notin [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\, da \\
&= \int \alpha \mathbb{1}\big\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\} + (1-\alpha)\mu(a, \boldsymbol{X}) - \mu(a, \boldsymbol{X}) \mathbb{1}\big\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\, da \\
&= \int \big\{\alpha - \mu(a, \boldsymbol{X})\big\} \mathbb{1}\big\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\, da + \int (1-\alpha)\mu(a, \boldsymbol{X})\, da \\
&= \int \big\{\alpha - \mu(a, \boldsymbol{X})\big\} \mathbb{1}\big\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\, da + B_2(\boldsymbol{O}) \ .
\end{aligned}
$$

where $B_2(\boldsymbol{O}) = \int (1-\alpha)\mu(a, \boldsymbol{X})\, da$. Therefore, $\mathcal{L}_{\mathrm{AIPW}} - \mathcal{L}^{(1)}$ does not depend on (f_L, f_U).

A.3 Details of the Median Heuristic and Cross-validation

First, we provide details on the median heuristic (Garreau et al., 2017). The bandwidth parameter γ is set as the median of the pairwise distances between the observed covariates, i.e., $\gamma = \mathrm{median}_{i,j} \left\| \boldsymbol{X}_i - \boldsymbol{X}_j \right\|$. This choice ensures that the entries of the Gram matrix, given by $\exp\big(-\left\| \boldsymbol{X}_i - \boldsymbol{X}_j \right\|_2^2 / \gamma^2\big)$, remain within a reasonable range, avoiding extreme values. This approach is referred to as the median heuristic.

Next, we summarize the cross-validation procedure for selecting λ in Algorithm 3.

Algorithm 3. Cross-validation

Require: Candidates of hyperparameters Θ; Number of folds M; Data $\mathcal{D}$

```
Split the data into equally sized L folds, say D_1, ..., D_M
for t = 1, ..., T do
    Let θ_t ∈ Θ (t = 1, ..., T) be the tth hyperparameter candidate
    for m = 1, ..., M do
        Let (ξ̂_L, ξ̂_U) be the solution to (5) obtained by using D_m^c and θ_t
        Let L̂_{t,m} be the empirical loss evaluated over D_m
    end for
    Let L̂_t be the average of L̂_{t,1}, ..., L̂_{t,M}
end for
    return Optimal hyperparameter θ* = arg min_{θ_t} L̂_t
```

A.4 Details on the Data Analysis

We first present a graphical summary of the estimated dose-probability curve. Figure 2 shows the estimated dose-probability curve, which has an inverted U-shape for almost all observations. Additionally, visual inspection guarantees the existence of the PDI for $\alpha \in [0.6, 0.7]$. Specifically, the PDIs are well-defined for 1397 observations (99.29%) at $\alpha = 0.6$ and 1403 observations (99.72%) at $\alpha = 0.7$. This suggests that Assumptions 4 and 5 are likely satisfied.

Next, we validate the propensity score model. Figure 3 shows the empirical distribution of the observed dose level A_i and its log-transformed value $\log(A_i)$, and the QQ plots of the residuals obtained from the parametric propensity score model where A_i or $\log(A_i)$ is regressed on the covariates $\boldsymbol{X}_i$, respectively. The visual diagnosis suggests that the propensity score model using the log-transformed dose is better than that using the raw dose.

Lastly, we compare the desired ranges $\mathcal{T}$ for Asians and those for other races. Table 4 shows the frequencies of $\mathcal{T}$ for Asians and non-Asians who have a complete set of covariates. We find that the desired ranges for Asians are very different from those for non-Asians. Therefore, we only focus on the Asian subgroup to make $\mathcal{T}$ homogeneous.

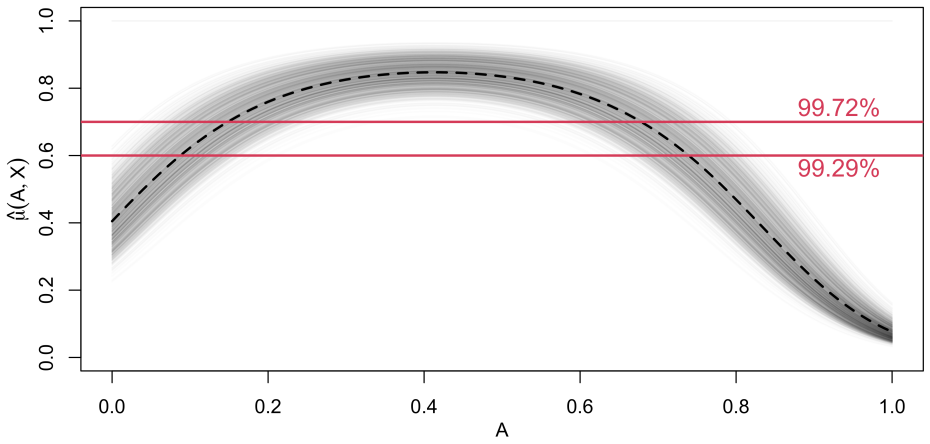

Figure 2: A graphical summary of the estimated dose-probability curve. The gray curves show $\widehat{\mu}(a, \boldsymbol{X}_i)$ for $a \in [0, 1]$ and patient i's covariates $\boldsymbol{X}_i$. The black dashed line shows the mean of the dose-probability curve across all $N = 1407$ patients. The red solid lines show the lower and upper ends of $\alpha = 0.6$ and 0.7, respectively. The red percentages represent the proportion of observations with a valid PDI estimate at $\alpha = 0.6, 0.7$.

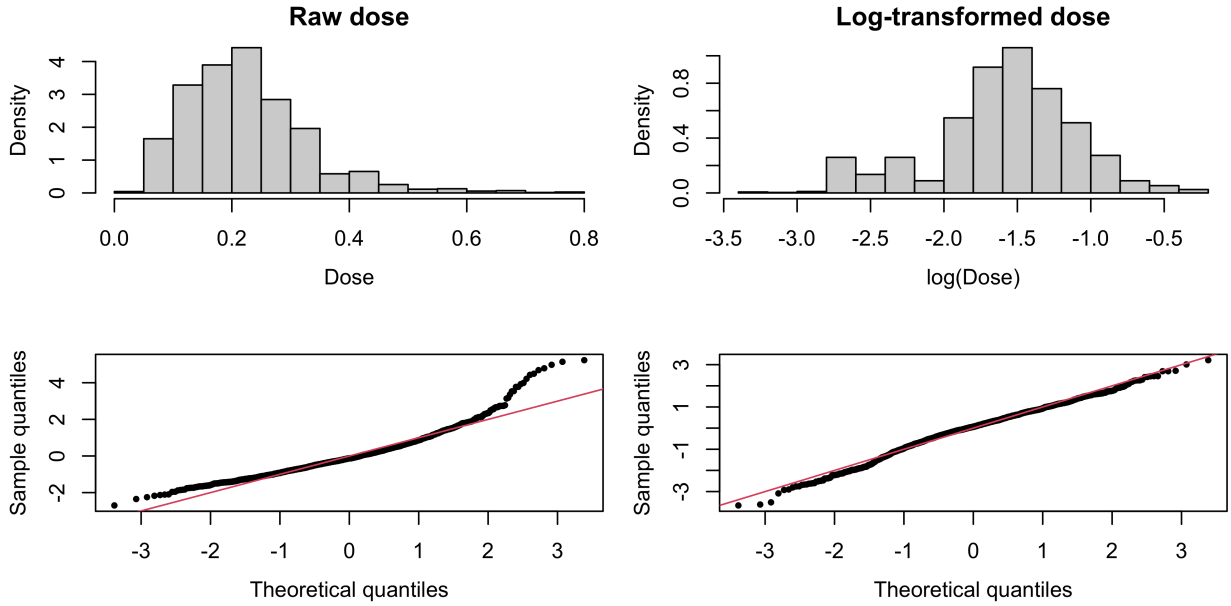

Figure 3: Graphical summaries of propensity score model diagnosis. The left and right panels show the raw dose A_i and the log-transformed dose $\log(A_i)$, respectively. The top panel shows histograms of the (log) dose levels of $N = 1407$ patients, and the bottom panel shows the QQ plot obtained from the parametric propensity score model, respectively. The red solid lines in the bottom panel visually guide the QQ line for the standard normal distribution.

Table 4: Frequency Table of $\mathcal{T}$ among Asians and Non-Asians.

Lower end of $\mathcal{T}$	0.8	1.2	1.25	1.3	1.5	1.7			1.75	1.8		
Upper end of $\mathcal{T}$	1.8	2.2	2.25	2.3	2.5	2.7	2.8	3.3	2.75	2.2	2.5	2.8
Non-Asian	1	1	6	1	4	184	0	0	4	1	3	5
Asian	0	0	0	0	631	1	236	249	0	0	0	0

Lower end of $\mathcal{T}$	2		2.1	2.2	2.25	2.3	2.5		2.75	3	3
Upper end of $\mathcal{T}$	3	3.5	3.1	3.2	3.25	3.3	3	3.5	3.75	3.5	4
Non-Asian	2805	435	1	11	1	3	2	335	1	2	10
Asian	286	1	0	0	0	0	0	5	0	0	0

B Proof

B.1 Proof of Lemma 3.1

Let $(f_L^{\#}, f_U^{\#})$ be the minimzer of the risk function $\mathcal{R}$. Based on the form of the loss function $\mathcal{L}$, the minimizer should satisfy $f_L^{\#}(\boldsymbol{X}) \le f_U^{\#}(\boldsymbol{X})$ for all $\boldsymbol{X}$. For a PDI candidate satisfying $f_L(\boldsymbol{X}) \le f_U(\boldsymbol{X})$, we find

$$
\begin{aligned}
&\mathrm{E}\Big\{\mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu, e) \,\Big|\, \boldsymbol{X}\Big\} \\
&= \mathrm{E}\left[\frac{\big\{\mu(A,\boldsymbol{X}) - R\big\}\big[\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\big]}{e(A \,|\, \boldsymbol{X})} \,\middle|\, \boldsymbol{X}\right] + \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\alpha - \mu(a,\boldsymbol{X})\big\}\, da \\
&= \mathrm{E}\left[\frac{\big\{\mu(A,\boldsymbol{X}) - \mu^*(A,\boldsymbol{X})\big\}\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}}{e(A \,|\, \boldsymbol{X})} \,\middle|\, \boldsymbol{X}\right] + \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\alpha - \mu(a,\boldsymbol{X})\big\}\, da \\
&= \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\mu(a,\boldsymbol{X}) - \mu^*(a,\boldsymbol{X})\big\}\frac{e^*(a \,|\, \boldsymbol{X})}{e(a \,|\, \boldsymbol{X})}\, da + \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\alpha - \mu(a,\boldsymbol{X})\big\}\, da\,. \qquad (9)
\end{aligned}
$$

If $\mu = \mu^*$, (9) reduces to

$$
\begin{aligned}
&\mathrm{E}\Big\{\mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu^*, e) \,\Big|\, \boldsymbol{X}\Big\} \\
&= \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \underbrace{\big\{\mu^*(a,\boldsymbol{X}) - \mu^*(a,\boldsymbol{X})\big\}}_{=0}\frac{e^*(a \,|\, \boldsymbol{X})}{e(a \,|\, \boldsymbol{X})}\, da + \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\alpha - \mu^*(a,\boldsymbol{X})\big\}\, da \\
&= \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\alpha - \mu^*(a,\boldsymbol{X})\big\}\, da\,.
\end{aligned}
$$

Likewise, if $e = e^*$, (9) becomes

$$
\begin{aligned}
&\mathrm{E}\Big\{\mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu, e^*) \,\Big|\, \boldsymbol{X}\Big\} \\
&= \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\mu(a,\boldsymbol{X}) - \mu^*(a,\boldsymbol{X})\big\}\underbrace{\frac{e^*(a \,|\, \boldsymbol{X})}{e^*(a \,|\, \boldsymbol{X})}}_{=1}\, da + \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\alpha - \mu(a,\boldsymbol{X})\big\}\, da \\
&= \int_{f_L(\boldsymbol{X})}^{f_U(\boldsymbol{X})} \big\{\alpha - \mu^*(a,\boldsymbol{X})\big\}\, da\,.
\end{aligned}
$$

Therefore, the minimizer $(f_L^{\#}, f_U^{\#})$ must have a form

$$
\begin{aligned}
f_L^{\#}(\boldsymbol{X}) &= \inf_{\ell}\big\{(\ell, u) \,|\, {}^{\forall}a \in (\ell, u) \;\Rightarrow\; \mu^*(a,\boldsymbol{X}) \ge \alpha\big\}\,, \\
f_U^{\#}(\boldsymbol{X}) &= \sup_{u}\big\{(\ell, u) \,|\, {}^{\forall}a \in (\ell, u) \;\Rightarrow\; \mu^*(a,\boldsymbol{X}) \ge \alpha\big\}\,,
\end{aligned}
$$

which is equivalent to the definition of the PDI, i.e., $(f_L^*, f_U^*) = (f_L^{\#}, f_U^{\#})$. This completes the proof.

B.2 Proof of Lemma 4.1

The proof is part of the proof of Theorem 4.2; see the derivation of (11) for details.

B.3 Proof of Theorem 4.2

Let $(f_L^\dagger, f_U^\dagger)$ and $(f_L^\#, f_U^\#)$ be the intermediate quantities that satisfy

$$(f_L^\dagger, f_U^\dagger) = \operatorname*{arg\,min}_{(f_L, f_U)} \mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) \ ,$$
$$(f_L^\#, f_U^\#) = \operatorname*{arg\,min}_{(f_L, f_U)} \mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L, f_U; \mu^*, e^*) \ .$$

We decompose the difference between the two risk functions as follows.

$$\begin{aligned}
&\mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) \\
&= \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) \\
&\quad + \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\#, f_U^\#; \mu^*, e^*\big) \\
&\quad + \underbrace{\mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\#, f_U^\#; \mu^*, e^*\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big)}_{\leq 0} \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) \\
&\leq \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) && \Leftarrow \quad (A) \\
&\quad + \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) && \Leftarrow \quad (B) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) && \Leftarrow \quad (C) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\#, f_U^\#; \mu^*, e^*\big) && \Leftarrow \quad (D) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) \ . && \Leftarrow \quad (E)
\end{aligned}$$

Likewise, we have

$$\begin{aligned}
&\mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) \\
&= \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\#, f_U^\#; \mu^*, e^*\big) \\
&\quad + \underbrace{\mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\#, f_U^\#; \mu^*, e^*\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big)}_{\leq 0} \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) \\
&\leq \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) && \Leftarrow \quad (B) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) && \Leftarrow \quad (C) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^\#, f_U^\#; \mu^*, e^*\big) && \Leftarrow \quad (D) \\
&\quad + \mathcal{R}_{\text{sur}}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(f_L^*, f_U^*; \mu^*, e^*\big) \ . && \Leftarrow \quad (E)
\end{aligned}$$

In (10), (11), (12), (28), and (29), we establish:

(10) : With probability not less than $1-\Delta_N$, $(A) \lesssim N^{-r_\mu - r_e}$;

(11) : $(B) \lesssim N^b \epsilon$;

(12) : With probability not less than $1-3e^{-\tau}$,

$$(C) \lesssim c_1 \lambda \gamma^{-d} + c_2 \gamma^\beta + c_3 \Big\{ \gamma^{(1-p)(1+\epsilon)d} \lambda^p N \Big\}^{-\frac{1}{2-p}} + c_4 N^{-1/2} \tau^{1/2} + c_5 N^{-1} \tau \ ;$$

(28) : With probability not less than $1-\Delta_N$, $(D) \lesssim C N^{-r_\mu - r_e} + C' \epsilon \leq C N^{-r_\mu - r_e} + C' N^b \epsilon$;

(29) : $(E) \lesssim \epsilon \leq N^b \epsilon$.

In the rest of the proof, we obtain the rate of each term, which completes the proof.

B.3.1 Upper bound of (A)

For any measurable functions $(f_L, f_U) : \mathcal{X}^{\otimes 2} \to \mathbb{R}^{\otimes 2}$, the difference between the loss function at (μ^*, e^*) and (μ', e') is

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{O}, f_L, f_U; \mu^*, e^*) - \mathcal{L}(\boldsymbol{O}, f_L, f_U; \mu', e') \\
&= \frac{\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}}{e^*(A \,|\, \boldsymbol{X}) e'(A \,|\, \boldsymbol{X})} \begin{bmatrix} R\{e^*(A \,|\, \boldsymbol{X}) - e'(A \,|\, \boldsymbol{X})\} \\ +\mu^*(A, \boldsymbol{X}) e'(A \,|\, \boldsymbol{X}) - \mu'(A, \boldsymbol{X}) e^*(A \,|\, \boldsymbol{X}) \end{bmatrix} \\
&\quad + \int \Big\{ \mu'(a, \boldsymbol{X}) - \mu^*(a, \boldsymbol{X}) \Big\} \Big\{ \mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\} \Big\} da \qquad \text{if } f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X}), \\
&= 0 \qquad \text{if } f_L(\boldsymbol{X}) > f_U(\boldsymbol{X}).
\end{aligned}$$

Note that for all (f_L, f_U), we have

$$\begin{aligned}
&\mathrm{E}\left[\frac{\mathbb{1}\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\} [\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}]}{e^*(A \,|\, \boldsymbol{X}) e'(A \,|\, \boldsymbol{X})} \begin{bmatrix} R\{e^*(A \,|\, \boldsymbol{X}) - e'(A \,|\, \boldsymbol{X})\} \\ +\mu^*(A, \boldsymbol{X}) e'(A \,|\, \boldsymbol{X}) \\ -\mu'(A, \boldsymbol{X}) e^*(A \,|\, \boldsymbol{X}) \end{bmatrix} \right] \\
&= \mathrm{E}\left[\frac{\mathbb{1}\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\} [\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}] \{\mu^*(A, \boldsymbol{X}) - \mu'(A, \boldsymbol{X})\}}{e'(A \,|\, \boldsymbol{X})} \right],
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{E}\Big[\mathbb{1}\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\} \int \Big\{ \mu'(a, \boldsymbol{X}) - \mu^*(a, \boldsymbol{X}) \Big\} \Big\{ \mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\} \Big\} da \Big] \\
&= \mathrm{E}\left[\frac{\mathbb{1}\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\} [\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}] \{\mu'(A, \boldsymbol{X}) - \mu^*(A, \boldsymbol{X})\}}{e^*(A \,|\, \boldsymbol{X})} \right].
\end{aligned}$$

As a consequence, (A) is upper bounded as follows.

$$\begin{aligned}
&\Big| \mathcal{R}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*\big) - \mathcal{R}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) \Big| \\
&= \Big| \mathrm{E}^{\mathcal{D}} \Big\{ \mathcal{L}(\boldsymbol{O}, \widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{L}(\boldsymbol{O}, \widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) \Big\} \Big| \\
&= \left| \mathrm{E}^{\mathcal{D}} \left[\begin{array}{r} \frac{\mathbb{1}\{\widehat{f}_L^{(s)}(\boldsymbol{X}) \leq \widehat{f}_U^{(s)}(\boldsymbol{X})\} \mathbb{1}\{A \in [\widehat{f}_L^{(s)}(\boldsymbol{X}), \widehat{f}_U^{(s)}(\boldsymbol{X})]\}}{\widehat{e}^{(-s)}(A \,|\, \boldsymbol{X}) e^*(A \,|\, \boldsymbol{X})} \\ \times \{e^*(A \,|\, \boldsymbol{X}) - \widehat{e}^{(-s)}(A \,|\, \boldsymbol{X})\} \{\mu^*(A, \boldsymbol{X}) - \widehat{\mu}^{(-s)}(A, \boldsymbol{X})\} \end{array} \right] \right| \\
&\leq \left\| \frac{\mathbb{1}\{\widehat{f}_L^{(s)}(\boldsymbol{X}) \leq \widehat{f}_U^{(s)}(\boldsymbol{X})\} \mathbb{1}\{A \in [\widehat{f}_L^{(s)}(\boldsymbol{X}), \widehat{f}_U^{(s)}(\boldsymbol{X})]\}}{\widehat{e}^{(-s)}(A \,|\, \boldsymbol{X}) e^*(A \,|\, \boldsymbol{X})} \right\|_{P,\infty} \\
&\qquad \times \Big| \mathrm{E}\Big[\{e^*(A \,|\, \boldsymbol{X}) - \widehat{e}^{(-s)}(A \,|\, \boldsymbol{X})\} \{\mu^*(A, \boldsymbol{X}) - \widehat{\mu}^{(-s)}(A, \boldsymbol{X})\} \Big] \Big| \\
&\precsim \big\| \widehat{e}^{(-s)} - e^* \big\|_{P,2} \big\| \widehat{\mu}^{(-s)} - \mu^* \big\|_{P,2} .
\end{aligned}$$

The two inequalities are obtained by applying the Hölder's inequality. In particular, we used the assumption that the (estimated) propensity scores are away from zero. Consequently, we have a constant C_1 satisfying

$$(A) \leq C_1 \big\| \widehat{e}^{(-s)} - e^* \big\|_{P,2} \big\| \widehat{\mu}^{(-s)} - \mu^* \big\|_{P,2} \lesssim N^{-r_\mu - r_e} \ , \tag{10}$$

where the second inequality holds from (RC5).

B.3.2 Upper bound of (B)

For any measurable functions $(f_L, f_U) : \mathcal{X}^{\otimes 2} \to \mathbb{R}^{\otimes 2}$, we consider the following three cases, namely: (i) $f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})$, (ii) $f_U(\boldsymbol{X}) + \epsilon \leq f_L(\boldsymbol{X})$, and (iii) $f_L(\boldsymbol{X}) \in (f_U(\boldsymbol{X}), f_U(\boldsymbol{X}) + \epsilon)$.

— Case 1: $f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})$

We find $\Phi_\epsilon\big(f_L(\boldsymbol{X}), f_U(\boldsymbol{X})\big) = \mathbb{1}\big\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\big\}$. Thus,

$$\begin{aligned}
&\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(s)}, \widehat{e}^{(s)}) - \mathcal{L}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(s)}, \widehat{e}^{(s)}) \\
&= \mathbb{1}\big\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\big\} \frac{\big\{\widehat{\mu}^{(-s)}(A, \boldsymbol{X}) - R\big\}\big[\Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big) - \mathbb{1}\big\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\big]}{\widehat{e}^{(-s)}(A \,|\, \boldsymbol{X})} \ .
\end{aligned}$$

Since $\{\widehat{\mu}^{(-s)}(A, \boldsymbol{X}) - R\}/\widehat{e}^{(-s)}(A \,|\, \boldsymbol{X})$ is uniformly bounded under Assumption (R1), we have

$$\begin{aligned}
&\Big| \mathrm{E}^{\mathcal{D}}\Big[\mathbb{1}\big\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\big\}\Big\{\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{L}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\Big\}\Big] \Big| \\
&\precsim \Big| \mathrm{E}\Big[\mathbb{1}\big\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\big\}\big[\Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big) - \mathbb{1}\big\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\big]\Big] \Big| \ .
\end{aligned}$$

Given $f_L \leq f_U$, we find

$$\begin{aligned}
&\Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big) - \mathbb{1}\big\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\} \\
&= \frac{A - f_L(\boldsymbol{X}) + \epsilon}{\epsilon} \mathbb{1}\big\{A \in [f_L(\boldsymbol{X}) - \epsilon, f_L(\boldsymbol{X})]\big\} + \frac{f_U(\boldsymbol{X}) - A + \epsilon}{\epsilon} \mathbb{1}\big\{A \in [f_U(\boldsymbol{X}), f_U(\boldsymbol{X}) + \epsilon]\big\} \ ,
\end{aligned}$$

which implies

$$\begin{aligned}
&\mathrm{E}\big\{\Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big) - \mathbb{1}\big\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\} \,|\, \boldsymbol{X}\big\} \\
&= \int_{f_L(\boldsymbol{X}) - \epsilon}^{f_L(\boldsymbol{X})} \frac{a - f_L(\boldsymbol{X}) + \epsilon}{\epsilon} P(A = a \,|\, \boldsymbol{X})\, da + \int_{f_U(\boldsymbol{X})}^{f_U(\boldsymbol{X}) + \epsilon} \frac{f_U(\boldsymbol{X}) - a + \epsilon}{\epsilon} P(A = a \,|\, \boldsymbol{X})\, da \\
&\leq c_e \int_{f_L(\boldsymbol{X}) - \epsilon}^{f_L(\boldsymbol{X})} \frac{a - f_L(\boldsymbol{X}) + \epsilon}{\epsilon}\, da + c_e \int_{f_U(\boldsymbol{X})}^{f_U(\boldsymbol{X}) + \epsilon} \frac{f_U(\boldsymbol{X}) - a + \epsilon}{\epsilon}\, da \\
&= c_e \cdot \epsilon \ .
\end{aligned}$$

The inequality is from the positivity assumption 3, and the last equality is trivial. Combining the result, we get

$$\begin{aligned}
&\Big| \mathrm{E}^{\mathcal{D}}\Big[\mathbb{1}\big\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\big\}\Big\{\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{L}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\Big\}\Big] \Big| \\
&\precsim \Big| \mathrm{E}\Big[\mathbb{1}\big\{f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})\big\}\big[\Psi_\epsilon\big(f_L(\boldsymbol{X}), A, f_U(\boldsymbol{X})\big) - \mathbb{1}\big\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\big\}\big]\Big] \Big| \\
&\leq c_e \cdot \epsilon \ .
\end{aligned}$$

— Case 2: $f_U(\boldsymbol{X}) + \epsilon \leq f_L(\boldsymbol{X})$

We have $\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{L}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) = 0$ since $\mathcal{L} = \mathcal{L}_{\text{sur}} \equiv C_{\mathcal{L}}$.

— Case 3: $f_L(\boldsymbol{X}) \in (f_U(\boldsymbol{X}), f_U(\boldsymbol{X}) + \epsilon)$

From the range of the loss functions, we have

$$\left|\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{L}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \leq 2C_{\mathcal{L}} \ .$$

Consequently,

$$\begin{aligned}
&\Big|\mathrm{E}^{\mathcal{D}}\Big[\mathbb{1}\big\{f_L(\boldsymbol{X}) \in (f_U(\boldsymbol{X}), f_U(\boldsymbol{X}) + \epsilon)\big\} \\
&\qquad\qquad \times \Big\{\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{L}(\boldsymbol{O}, f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\Big\}\Big]\Big| \\
&\leq 2C_{\mathcal{L}} \cdot \mathrm{E}^{\mathcal{D}}\big[\mathbb{1}\big\{f_L(\boldsymbol{X}) - f_U(\boldsymbol{X}) \in (0, \epsilon)\big\}\big] \ .
\end{aligned}$$

Combining three cases, we find

$$\begin{aligned}
&\left|\mathcal{R}^{\mathcal{D}}_{\text{sur}}(f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \\
&\leq c_e \cdot \epsilon + 0 + 2C_{\mathcal{L}} \cdot \mathrm{E}^{\mathcal{D}}\big[\mathbb{1}\big\{f_L(\boldsymbol{X}) - f_U(\boldsymbol{X}) \in (0, \epsilon)\big\}\big] \ .
\end{aligned}$$

At the optimal PDI, we have $\mathrm{E}^{\mathcal{D}}\big[\mathbb{1}\big\{f^*_L(\boldsymbol{X}) - f^*_U(\boldsymbol{X}) \in (0, \epsilon)\big\}\big] = 0$. At the estimated PDI, this term is assumed to be upper bounded by a term having a rate $N^b\epsilon$. As a consequence, we get an upper bound of (B) for some constant C_2:

$$\begin{aligned}
&\left|\mathcal{R}^{\mathcal{D}}_{\text{sur}}(f^*_L, f^*_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f^*_L, f^*_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \precsim \epsilon \\
&\left|\mathcal{R}^{\mathcal{D}}_{\text{sur}}(\widehat{f}^{(s)}_L, \widehat{f}^{(s)}_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(\widehat{f}^{(s)}_L, \widehat{f}^{(s)}_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \precsim N^b\epsilon \ .
\end{aligned}$$

The second result implies

$$(B) \leq C_2 \cdot N^b\epsilon \ . \tag{11}$$

Note that this provides the proof of Lemma 4.1. In addition, if $\epsilon \asymp N^{-\beta/(2\beta+d)-b}$, we get

$$(B) \leq C_2 N^{-\beta/(2\beta+d)} \ .$$

B.3.3 Upper bound of (C)

The outline of the proof is similar to that of Theorem 2 of Chen et al. (2016). We remark that lemmas and theorems presented in Steinwart & Christmann (2008) and Eberts & Steinwart (2013) are essential for the proof. For notational brevity, let $L(\boldsymbol{O}, \ell, u) = \mathcal{L}_{\text{sur}}(\boldsymbol{O}, \ell, u\cdot; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})$, $R(\ell, u) = \mathrm{E}^{\mathcal{D}}\big\{\mathcal{L}_{\text{sur}}(\boldsymbol{O}, \ell, u; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\big\}$, and the expectation involving L is calculated conditioning on the estimation split data, i.e., $\mathrm{E}(L) := \mathrm{E}^{\mathcal{D}}\big\{\mathcal{L}_{\text{sur}}(\cdot; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\big\}$. Let n be the number of observations in the ERM dataset $\mathcal{D}_s$, which is proportional to N.

We require L to satisfy the following conditions:

(C1) For all $(\boldsymbol{o}, \ell, u)$, there exists a constant $B > 0$ satisfying $0 \leq L(\boldsymbol{o}, \ell, u) \leq B$.

(C2) $L(\boldsymbol{o}, \ell, u)$ is locally Lipschitz continuous with respect to (ℓ, u) for $\ell \leq u$.

(C3) For all $(t, \boldsymbol{o})$, we have $L(\boldsymbol{o}, \ell^W, u^W) \leq L(\boldsymbol{o}, \ell, u)$ where $\ell^W = \ell \cdot \mathbb{1}\big\{|\ell| \leq c_0\big\} + \mathrm{sign}(\ell)c_0 \cdot \mathbb{1}\big\{c_0 < |\ell|\big\}$.

(C4) $\mathrm{E}\big[\{L(\boldsymbol{O}, f_L^W, f_U^W) - L(\boldsymbol{O}, f_L^\dagger, f_U^\dagger)\}^2\big] \leq V \cdot \big[\mathrm{E}\{L(\boldsymbol{O}, f_L^W, f_U^W) - L(\boldsymbol{O}, f_L^\dagger, f_U^\dagger)\}\big]^v$ is satisfied for constant $v \in [0,1]$, $V \geq B^{2-v}$, and for all $f_L, f_U \in \mathcal{H}$.

With some algebra, we find that the surrogate loss function L satisfies these conditions by (i) adding a sufficiently large constant to make L non-negative and (ii) adding a very large constant for PDI candidates outside of the proper dose range $\mathbb{R}^2 \setminus [0,1]^2$. Then, we find each condition is satisfied as follows:

(C1) Trivial after adding a baseline constant.

(C2) The surrogate loss function is Lipschitz continuous for given ϵ due to its construction.

(C3) Trivial after adding a penalizing constant over $\mathbb{R}^2 \setminus [0,1]^2$.

(C4) This holds with $v = 0$ and $V = 4B^2$.

Our ultimate goal is to show the following inequality holds

$$\begin{aligned}
&\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}\big(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) \\
&= \mathrm{E}\big\{L(\boldsymbol{O}, \widehat{f}_L^W, \widehat{f}_U^W) - L(\boldsymbol{O}, f_L^\dagger, f_U^\dagger)\big\} \\
&\leq c_1 \lambda \gamma^{-d} + c_2 \gamma^\beta + c_3 \Big\{\gamma^{(1-p)(1+\epsilon)d} \lambda^p N\Big\}^{-\frac{1}{2-p}} + c_4 N^{-1/2} \tau^{1/2} + c_5 N^{-1} \tau \ . \qquad (12)
\end{aligned}$$

with probability $P_{\boldsymbol{O}}^n$ not less than $1 - 3e^{-\tau}$ where $c_1, \ldots, c_5$ does not depend on N. Then, taking $\gamma \asymp N^{-1/(2\beta+d)}$ and $\lambda \asymp N^{-(\beta+d)/(2\beta+d)}$, we get the following asymptotic rate with probability not less than $1 - 3e^{-\tau}$:

$$\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}\big(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) - \mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}\big(f_L^*, f_U^*; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}\big) = O_P\big(N^{-\beta/(2\beta+d)}\big) \ . \qquad (13)$$

Since the estimated nuisance functions are random (depending on $\mathcal{D}_s^c$), we use O_P-notation instead of O-notation.

To show the result, we first present a formal result by extending Theorem 7.23 of Steinwart & Christmann (2008) to our setting.

Theorem B.1 (Extension of Theorem 7.23. of Steinwart & Christmann (2008)) *Let L be a loss function having a non-negative value. Also, let $\mathcal{H}$ be a Gaussian RKHS with bandwidth parameter γ over $\mathcal{X} \subseteq \mathbb{R}^d$. Consider the ERM of the form*

$$(\widehat{f}_L, \widehat{f}_U) = \operatorname*{arg\,min}_{(f_L, f_U) \in \mathcal{H}^{\otimes 2}} \left[\frac{1}{N} \sum_{i=1}^{N} L(\boldsymbol{O}_i, f_L, f_U) + \lambda \left\{ \|f_L\|_{\mathcal{H}}^2 + \lambda \|f_U\|_{\mathcal{H}}^2 \right\} \right] . \tag{14}$$

Let $P_{\boldsymbol{O}}$ and $P_{\boldsymbol{X}}$ be distributions of $\boldsymbol{O}$ and $\boldsymbol{X}$, respectively. Furthermore, suppose Regularity Conditions (R2), (R3), and , (R5) hold. Assume that for fixed $n \geq 1$, there exist constant $p \in (0,1)$, and $\mathfrak{a} \geq B$ such that

$$E_{\mathcal{O} \sim P_{\boldsymbol{O}}^{|\mathcal{D}_s|}} \left[e_n \big(\textit{identity map} : \mathcal{H}^{\otimes 2} \to L_2^{\otimes 2}(\mu) \big) \right] \leq \mathfrak{a} \cdot i^{-\frac{1}{2p}} , \quad i \geq 1 . \tag{15}$$

Finally, fix $(f_{L,0}, f_{U,0}) \in \mathcal{H}^{\otimes 2}$ and a constant $B_0 \geq B$ such that $\big\| L(O, f_{L,0}, f_{U,0}) \big\|_{\infty} \leq B_0$. Then, for all fixed $\tau > 0$, and $\lambda > 0$, the ERM (14) using $\mathcal{H}$ and L satisfies

$$\begin{aligned}
&\lambda \|\widehat{f}_L\|_{\mathcal{H}}^2 + \lambda \|\widehat{f}_U\|_{\mathcal{H}}^2 + R(\widehat{f}_L^W, \widehat{f}_U^W) - R(f_L^\dagger, f_U^\dagger) \\
&\leq 9 \{ \lambda \|f_{L,0}\|_{\mathcal{H}}^2 + \lambda \|f_{U,0}\|_{\mathcal{H}}^2 + R(f_{L,0}, f_{U,0}) - R(f_L^\dagger, f_U^\dagger) \} \\
&\quad + K_0 \left\{ \frac{\mathfrak{a}^{2p}}{\lambda^p n} \right\}^{\frac{1}{2-p-v+vp}} + 3 \left(\frac{72 V \tau}{n} \right)^{\frac{1}{2-v}} + \frac{15 B_0 \tau}{n} .
\end{aligned}$$

with probability $P_{\boldsymbol{O}}^n$ not less than $1 - 3e^{-\tau}$, where where $K_0 \geq 1$ is a constant only depending on p, c_0, B, v, and V.

We establish Theorem B.1 following the proof of Theorem 7.23 of Steinwart & Christmann (2008). When $a^{2p} > \lambda^p n$, we find

$$\begin{aligned}
&\lambda \|\widehat{f}_L\|_{\mathcal{H}}^2 + \lambda \|\widehat{f}_U\|_{\mathcal{H}}^2 + R(\widehat{f}_L^W, \widehat{f}_U^W) - R(f_L^\dagger, f_U^\dagger) \\
&\leq \lambda \|\widehat{f}_L\|_{\mathcal{H}}^2 + \lambda \|\widehat{f}_U\|_{\mathcal{H}}^2 + \widehat{R}(\widehat{f}_L^W, \widehat{f}_U^W) + B \\
&\leq \widehat{R}(0,0) + B \leq 2B \leq 2B \left(\frac{a^{2p}}{\lambda^p n} \right)^{\frac{1}{1-(1-p)(1-v)}} ,
\end{aligned}$$

where $\widehat{R}(\ell, u)$ is the empirical risk $\sum_{i \in \mathcal{D}_s} L(\boldsymbol{O}_i, \ell, u)/n$. Therefore, the claim holds whenever $K_0 \leq 2B$.

Next, we consider $a^{2p} \leq \lambda^p n$. Let r^* be

$$r^* := \inf_{f_L, f_U \in \mathcal{H}} \Big\{ \underbrace{\lambda \|f_L\|_{\mathcal{H}}^2 + \lambda \|f_U\|_{\mathcal{H}}^2}_{\Gamma(f_L, f_U)} + R(f_L^W, f_U^W) - R(f_L^\dagger, f_U^\dagger) \Big\} ,$$

and we further define the following sets for $r > r^*$:

$$\begin{aligned}
(\mathcal{H}_L \otimes \mathcal{H}_U)_r &:= \big\{ f_L, f_U \in \mathcal{H} \,|\, \Gamma(f_L, f_U) + R(f_L^W, f_U^W) - R(f_L^\dagger, f_U^\dagger) \leq r \big\} \\
\mathcal{L}_r &:= \big\{ L(\boldsymbol{O}, f_L^W, f_U^W) - L(\boldsymbol{O}, f_L^\dagger, f_U^\dagger) \,|\, (f_L, f_U) \in (\mathcal{H}_L \otimes \mathcal{H}_U)_r \big\} .
\end{aligned}$$

Let $B_{\mathcal{H}}$ be the closed unit ball of $\mathcal{H}$. For $\boldsymbol{f} \in (\mathcal{H}_L \otimes \mathcal{H}_U)_r$, we have

$$\Gamma(f_L, f_U) = \lambda \|f_L\|_{\mathcal{H}}^2 + \lambda \|f_U\|_{\mathcal{H}}^2 \leq \lambda \|f_L\|_{\mathcal{H}}^2 + \lambda \|f_U\|_{\mathcal{H}}^2 + R(f_L^W, f_U^W) - R(f_L^\dagger, f_U^\dagger) \leq r .$$

This implies $\boldsymbol{f} \in (\mathcal{H}_L \otimes \mathcal{H}_U)_r$ also belong to $((r/\lambda)^{1/2} B_{\mathcal{H}})^{\otimes 2}$.

To proceed, we adopt the approach used in the proof of Lemma 7.17 of Steinwart & Christmann (2008) to our setting. Given that $L \circ (\mathcal{H}_L \otimes \mathcal{H}_U)_r^W \subseteq L_2(P_{\boldsymbol{O}})$, we have the relationship for (dyadic) entropy number

e_n:

$$e_n\big(\mathcal{L}_r, \|\cdot\|_{L_2(P_{\boldsymbol{O}})}\big) \le e_n\big(L\circ(\mathcal{H}_L\otimes\mathcal{H}_U)_r, \|\cdot\|_{L_2(P_{\boldsymbol{O}})}\big)\ .$$

Now let $\{\boldsymbol{f}_1,\ldots,\boldsymbol{f}_{2^{n-1}}\}$ ($\boldsymbol{f}=(f_{L,i},f_{U,i})$) be an ϵ-net of $(\mathcal{H}_L\otimes\mathcal{H}_U)_r$ with respect to $\|\cdot\|_{L_2(P_{\boldsymbol{X}})}$ where

$$\|\boldsymbol{f}\|^2_{L_2(P_{\boldsymbol{X}})} = \int \{f_L(\boldsymbol{X})\}^2 + \{f_U(\boldsymbol{X})\}^2\, dP_{\boldsymbol{X}}\ .$$

See Section A.5.6 of Steinwart & Christmann (2008) for details on the dyadic entropy number. This means that, for $\boldsymbol{f}\in(\mathcal{H}_L\otimes\mathcal{H}_U)_r$, there exists an $i\in\{1,\ldots,2^{M-1}\}$ such that $\|\boldsymbol{f}-\boldsymbol{f}\|_{L_2(P_{\boldsymbol{X}})}\le\epsilon$. Thus,

$$\begin{aligned}
\Big\|L\circ\boldsymbol{f}^W - L\circ\boldsymbol{f}^W\Big\|^2_{L_2(P_{\boldsymbol{O}})} &= \int \Big|L\big(\boldsymbol{O}, f_L^W(\boldsymbol{X}), f_U^W(\boldsymbol{X})\big) - L\big(\boldsymbol{O}, f_{L,i}^W(\boldsymbol{X}), f_{U,i}^W(\boldsymbol{X})\big)\Big|^2\, dP_{\boldsymbol{O}} \\
&\le L_c^2\Big[\int \Big|f_L^W(\boldsymbol{X}) - f_{L,i}^W(\boldsymbol{X})\Big|^2\, dP_{\boldsymbol{X}} + \int \Big|f_U^W(\boldsymbol{X}) - f_{U,i}^W(\boldsymbol{X})\Big|^2\, dP_{\boldsymbol{X}}\Big] \\
&\le 4L_c^2\epsilon^2\ .
\end{aligned}$$

Therefore, $\{L\circ\boldsymbol{f}_1^W,\ldots,L\circ\boldsymbol{f}_{2^{n-1}}^W\}$ is an $(2L_c\epsilon)$-net of $L\circ(\mathcal{H}_L\otimes\mathcal{H}_U)_r$. As a consequence,

$$e_n\big(\mathcal{L}_r, \|\cdot\|_{L_2(P_{\boldsymbol{O}})}\big) \le 2L_c e_n\big((\mathcal{H}_L\otimes\mathcal{H}_U)_r, \|\cdot\|_{L_2(P_{\boldsymbol{X}})}\big)\ .$$

Taking expectation (only over the data used in ERM which is $\mathcal{D}_s$), we have

$$\begin{aligned}
&\mathrm{E}_{\mathcal{O}\sim P_{\boldsymbol{O}}^{|\mathcal{D}_s|}}\Big[e_n\big(\text{identity map}: \mathcal{L}_r\to L_2(\mathcal{O})\big)\Big] \\
&\le 2L_c\cdot \mathrm{E}_{\mathcal{X}\sim P_{\boldsymbol{X}}^{|\mathcal{D}_s|}}\Big[\underbrace{e_n\big(\text{identity map}: (\mathcal{H}_L\otimes\mathcal{H}_U)_r\to L_2^{\otimes 2}(\mathcal{X})\big)}_{=\text{(Term 1)}}\Big]\ .
\end{aligned}$$

Here, $L_2^{\otimes 2}(\mathcal{X}) = \{(f,g)\,|\,\int f^2(\boldsymbol{X})\,dP_{\boldsymbol{X}} + \int g^2(\boldsymbol{X})\,dP_{\boldsymbol{X}} < \infty\}$.

The quantity (Term 1) can be further upper bounded using Theorem 7.34 of Steinwart & Christmann (2008), which we state below for completeness:

Theorem 7.34 (*Entropy Numbers for Gaussian Kernels*; Steinwart & Christmann (2008)) Let μ be a distribution on $\mathbb{R}^d$ having tail exponent $\tau\in(0,\infty]$, and $\mathcal{H}_\gamma(\mathbb{R}^d)$ be the RKHS endowed with the Gaussian kernel with bandwidth γ. Then, for all $\epsilon>0$ and $d/(d+\tau)<p<1$, there exists a constant $c\ge 1$ such that

$$e_n(\text{identity map}: \mathcal{H}_\gamma(\mathbb{R}^d)\to L_2(\mu)) \le c_{\epsilon,p}\gamma^{-\frac{(1-p)(1+\epsilon)d}{2p}} i^{-\frac{1}{2p}}$$

for all $i\ge 1$ and $\gamma\in(0,1]$.

Since $\mathcal{H}$ is induced by the Gaussian kernel, we have

$$\begin{aligned}
\text{(Term 1)} &= e_n\big(\text{identity map}: (\mathcal{H}_L\otimes\mathcal{H}_U)_r\to L_2^{\otimes 2}(\mathcal{X})\big) \\
&\le e_n\big(\text{identity map}: ((r/\lambda)^{1/2}B_{\mathcal{H}})^{\otimes 2}\to \ell_\infty^{\otimes 2}((r/\lambda)^{1/2}B_{\mathcal{H}})\big) \\
&\le \left(\frac{r}{\lambda}\right)^{1/2} e_n\big(\text{identity map}: (B_{\mathcal{H}})^{\otimes 2}\to \ell_\infty^{\otimes 2}(B_{\mathcal{H}})\big)\ . \qquad (16)
\end{aligned}$$

From Lemma 6.21 and Excercise 6.8 of Steinwart & Christmann (2008), we have

$$\begin{aligned}
&e_n\big(\text{identity map}: (B_{\mathcal{H}})^{\otimes 2}\to \ell_\infty^{\otimes 2}(B_{\mathcal{H}})\big) \lesssim n^{-1/q} \text{ for some } q>0 \\
\Leftrightarrow\quad &\log\mathcal{N}(\epsilon, (B_{\mathcal{H}})^{\otimes 2}, \|\cdot\|_\infty) \lesssim (1/\epsilon)^q \text{ for some } q>0 \qquad (17)
\end{aligned}$$

where $\mathcal{N}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the ϵ-covering number of a space $\mathcal{F}$ with respect to a metric $\|\cdot\|$; see Definition 6.19 of Steinwart & Christmann (2008) for details. Therefore, to derive the upper bound of the right-hand side (16), it suffices to characterize an upper bound of $\log \mathcal{N}(\epsilon, (B_{\mathcal{H}})^{\otimes 2}, \|\cdot\|_\infty)$. In order to do so, let $C_b^m(S)$ be all continuous functions from $S \subseteq \mathbb{R}^d$ with bounded m-th derivative. We further introduce Theorem 2.7.1 of van der Vaart & Wellner (1996) and an extension of Lemma 7 of Haussler (1992) which are provided below for completeness:

Theorem 2.7.1 (van der Vaart & Wellner (1996)) Let $\mathcal{X}$ be a bounded, convex subset of $\mathbb{R}^d$ with nonempty interior. There exists a constant K depending only on m and d such that

$$\log \mathcal{N}(\epsilon, C_b^m(S), \|\cdot\|_\infty) \lesssim (1/\epsilon)^{d/m} \ . \tag{18}$$

Extension of Lemma 7 (Haussler (1992))

$$\log \mathcal{N}(\epsilon, C_b^m(S)^{\otimes 2}, \|\cdot\|_\infty) \leq 2 \log \mathcal{N}(\epsilon, C_b^m(S), \|\cdot\|_\infty) \ . \tag{19}$$

We briefly show (19). Let U be an ϵ-net of $C_b^m(S)$ having n_ϵ elements. Let $\mathcal{U} = U^{\otimes 2} = \{(u_1, u_2) \,|\, u_1, u_2 \in U\}$ having n_ϵ^2 by construction. It suffices to show that $\mathcal{U}$ is an ϵ-net of $C_b^m(S)^{\otimes 2}$. We arbitrarily take any function $(g_1, g_2) \in C_b^m(S)^{\otimes 2}$. Then, we can find $u_j \in U$ such that $\|g_j - u_j\|_\infty \leq \epsilon$, which implies

$$\big\|(g_1, g_2) - (u_1, u_2)\big\|_\infty = \max_{i=1,2} \big\|g_i - u_i\big\|_\infty \leq \epsilon \ .$$

Therefore, $\mathcal{N}(\epsilon, C_b^m(S)^{\otimes 2}, \|\cdot\|_\infty) \leq |\mathcal{U}| = n_\epsilon^2 = \mathcal{N}(\epsilon, C_b^m(S), \|\cdot\|_\infty)$.

Combining equations (16)-(19), we establish that (Term 1) $\lesssim (r/\gamma)^{1/2} n^{-m/d}$, which is essentially the same as the result of Theorems 6.26, 6.27, 7.33 of Steinwart & Christmann (2008). Therefore, the remainder of the proof of Theorem 7.34 of Steinwart & Christmann (2008) applies to our setting. Therefore, from Theorem 7.34 of Steinwart & Christmann (2008), we get

$$\begin{aligned}
\text{(Term 1)} &= e_n\big(\text{identity map} : (B_{\mathcal{H}})^{\otimes 2} \to \ell_\infty^{\otimes 2}(B_{\mathcal{H}})\big) \\
&\leq e_n\big(\text{identity map} : \mathcal{H}^{\otimes 2} \to L_2^{\otimes 2}(\mu)\big) \\
&\leq c_{\epsilon,p} \gamma^{-\frac{(1-p)(1+\epsilon)d}{2p}} i^{-\frac{1}{2p}} \ ,
\end{aligned}$$

for all $i \geq 1$ and $\gamma \in (0, 1]$. which satisfies condition (15) with $\mathfrak{a} = c_{\epsilon,p} \gamma^{-\frac{(1-p)(1+\epsilon)d}{2p}}$. This further implies the following result:

$$\begin{aligned}
&\mathrm{E}_{\mathcal{O} \sim P_{\boldsymbol{O}}^{|\mathcal{D}_s|}} \Big[e_n\big(\text{identity map} : \mathcal{L}_r \to L_2(\mathcal{O})\big) \Big] \\
&\leq 2L_c \cdot \mathrm{E}_{\mathcal{X} \sim P_{\boldsymbol{X}}^{|\mathcal{D}_s|}} \Big[\underbrace{e_n\big(\text{identity map} : (\mathcal{H}_L \otimes \mathcal{H}_U)_r \to L_2^{\otimes 2}(\mathcal{X})\big)}_{\text{(Term 1)}} \Big] \\
&\leq 2L_c \left(\frac{r}{\lambda}\right)^{1/2} \cdot \mathfrak{a} \cdot i^{-\frac{1}{2p}} \quad \text{for any } i \geq 1 \ .
\end{aligned}$$

The rest of the proof of Theorem B.1 is the same as that in Steinwart & Christmann (2008), which yields

$$\begin{aligned}
&\lambda\|\widehat{f}_L\|_{\mathcal{H}}^2 + \lambda\|\widehat{f}_U\|_{\mathcal{H}}^2 + R(\widehat{f}_L^W, \widehat{f}_U^W) - R(f_L^\dagger, f_U^\dagger) \\
&\leq 9\{\underbrace{\lambda\|f_{L,0}\|_{\mathcal{H}}^2 + \lambda\|f_{U,0}\|_{\mathcal{H}}^2 + R(f_{L,0}, f_{U,0}) - R(f_L^\dagger, f_U^\dagger)}_{=B(f_{L,0}, f_{U,0})}\} \\
&\quad + K_0\left\{\frac{\gamma^{-(1-p)(1+\epsilon)d}}{\lambda^p n}\right\}^{\frac{1}{2-p}} + 3\left(\frac{72V\tau}{n}\right)^{\frac{1}{2-v}} + \frac{15B_0\tau}{n} \\
&\overset{\text{take } v=0}{=} B(f_{L,0}, f_{U,0}) + K_0\left\{\frac{\gamma^{-(1-p)(1+\epsilon)d}}{\lambda^p n}\right\}^{\frac{1}{2-p}} + 36\sqrt{2}B\sqrt{\frac{\tau}{n}} + 15B\frac{\tau}{n}\ . \qquad (20)
\end{aligned}$$

Since the above result holds for any $(f_{L,0}, f_{U,0}) \in \mathcal{H}^{\otimes 2}$, we can further bound $B(f_{L,0}, f_{U,0})$ by carefully choosing $(f_{L,0}, f_{U,0})$ in a way presented in Eberts & Steinwart (2013). We first define a function $Q_{r,\gamma} : \mathbb{R}^d \to \mathbb{R}$ as

$$Q_{r,\gamma}(\boldsymbol{z}) = \sum_{j=1}^{r}\binom{r}{j}(-1)^{1-j}\frac{1}{j^d}\left(\frac{2}{\gamma^2}\right)^{\frac{d}{2}}\mathcal{K}_{j\gamma/\sqrt{2}}(\boldsymbol{z})\ ,\ \mathcal{K}_\gamma(\boldsymbol{z}) = \exp\left\{-\gamma^{-2}\|\boldsymbol{z}\|_2^2\right\}\ . \qquad (21)$$

for $r \in \{1, 2, \ldots\}$ and $\gamma > 0$. If $(f_L^\dagger, f_U^\dagger) \in L_2(\mathbb{R}^d) \cap L_\infty(\mathbb{R}^d)$, we can define $(f_{L,0}, f_{U,0})$ by convolving $Q_{r,\gamma}$ with $(f_L^\dagger, f_U^\dagger)$ as follows (Eberts & Steinwart, 2013).

$$f_{L,0}(\boldsymbol{X}) = \big(Q_{r,\gamma} * f_L^*\big)(\boldsymbol{X}) = \int_{\mathbb{R}^d} Q_{r,\gamma}(\boldsymbol{X} - \boldsymbol{z}) f_L^*(\boldsymbol{z})\, d\boldsymbol{z}\ .$$

Next, we introduce Theorems 2.2 and 2.3 of Eberts & Steinwart (2013).

Theorem 2.2 (Eberts & Steinwart (2013)) Let us fix some $q \in [1, \infty)$. Furthermore, assume that $P_{\boldsymbol{X}}$ is a distribution on $\mathbb{R}^d$ that has a Lebesgue density $p_{\boldsymbol{X}} \in L_p(\mathbb{R}^d)$ for some $p \in [1, \infty]$. Let $f : \mathbb{R}^d \to \mathbb{R}$ be such that $f \in L_q(\mathbb{R}^d) \cap L_\infty(\mathbb{R}^d)$. Then, for $r \in \{1, 2, \ldots\}$, $\gamma > 0$, and $s \geq 1$ with $1 = s^{-1} + p^{-1}$, we have

$$\|Q_{r,\gamma} * f - f\|_{L_q(P_{\boldsymbol{X}})}^q \leq C_{r,q} \cdot \|p_{\boldsymbol{X}}\|_{L_p(\mathbb{R}^d)} \cdot \omega_{r, L_{qs}(\mathbb{R}^d)}^q(f, \gamma/2)$$

where $C_{r,q}$ is a constant only depending on r and q.

Theorem 2.3 (Eberts & Steinwart (2013)) Let $f \in L_2(\mathbb{R}^d)$, $\mathcal{H}$ be the RKHS of the Gaussian kernel with parameter $\gamma > 0$, and $Q_{r,\gamma}$ be defined by (21) for a fixed $r \in \{1, 2, \ldots\}$. Then we have $Q_{r,\gamma} * f \in \mathcal{H}$ with

$$\|Q_{r,\gamma} * f\|_{\mathcal{H}} \leq \big(\gamma\sqrt{\pi}\big)^{-\frac{d}{2}}\big(2^r - 1\big)\|f\|_{L_2(\mathbb{R}^d)}\ .$$

Moreover, if $f \in L_\infty(\mathbb{R}^d)$, we have $|Q_{r,\gamma} * f| \leq (2^r - 1)\|f\|_{L_\infty(\mathbb{R}^d)}$.

As a result, we obtain

$$
\begin{aligned}
&\lambda\|f_{L,0}\|_{\mathcal{H}}^2+\lambda\|f_{U,0}\|_{\mathcal{H}}^2+R(f_{L,0},f_{U,0})-R(f_L^\dagger,f_U^\dagger) \\
&=\lambda\|Q_{r,\gamma}*f_L^\dagger\|_{\mathcal{H}}^2+\lambda\|Q_{r,\gamma}*f_U^\dagger\|_{\mathcal{H}}^2+R(Q_{r,\gamma}*f_L^\dagger,Q_{r,\gamma}*f_U^\dagger)-R(f_L^\dagger,f_U^\dagger) \\
&\le\lambda\big(\gamma\sqrt{\pi}\big)^{-d}\big(2^r-1\big)^2\big\{\|f_L^\dagger\|_{L_2(\mathbb{R}^d)}^2+\|f_U^\dagger\|_{L_2(\mathbb{R}^d)}^2\big\}+R(Q_{r,\gamma}*f_L^\dagger,Q_{r,\gamma}*f_U^\dagger)-R(f_L^\dagger,f_U^\dagger) \\
&\le\lambda\big(\gamma\sqrt{\pi}\big)^{-d}\big(2^r-1\big)^2\big\{\|f_L^\dagger\|_{L_2(\mathbb{R}^d)}^2+\|f_U^\dagger\|_{L_2(\mathbb{R}^d)}^2\big\} \\
&\qquad+B'\cdot\|Q_{r,\gamma}*f_L^\dagger-f_L^\dagger\|_{L_1(P_{\boldsymbol{X}})}+B'\cdot\|Q_{r,\gamma}*f_U^\dagger-f_U^\dagger\|_{L_1(P_{\boldsymbol{X}})} \\
&\le\lambda\big(\gamma\sqrt{\pi}\big)^{-d}\big(2^r-1\big)^2\big\{\|f_L^\dagger\|_{L_2(\mathbb{R}^d)}^2+\|f_U^\dagger\|_{L_2(\mathbb{R}^d)}^2\big\} \\
&\qquad+B'\cdot C_{r,1}\cdot\|p_{\boldsymbol{X}}\|_{L_\infty(\mathbb{R}^d)}\cdot\big\{\omega_{r,L_1(\mathbb{R}^d)}(f_L^\dagger,\gamma/2)+\omega_{r,L_1(\mathbb{R}^d)}(f_U^\dagger,\gamma/2)\big\} \\
&\le\lambda\big(\gamma\sqrt{\pi}\big)^{-d}\big(2^r-1\big)^2\big\{\|f_L^\dagger\|_{L_2(\mathbb{R}^d)}^2+\|f_U^\dagger\|_{L_2(\mathbb{R}^d)}^2\big\}+B'c'\cdot C_{r,1}\cdot\|p_{\boldsymbol{X}}\|_{L_\infty(\mathbb{R}^d)}\gamma^\beta\ .
\end{aligned}
\tag{22}
$$

The first equality is from the construction of $(f_{L,0},f_{U,0})$. The first inequality is from Theorem 2.3 of Eberts & Steinwart (2013). The second inequality is from Lipschitz continuity of L. The third inequality is from Theorem 2.2 of Eberts & Steinwart (2013) with $q=s=1$ and $p=\infty$. The last inequality holds for some constant c' since $f_L^\dagger,f_U^\dagger\in\mathcal{B}_{1,\infty}^\beta(\mathbb{R}^d)$ implies $\omega_{r,L_1(\mathbb{R}^d)}(f^\dagger,\gamma/2)\le c'\gamma^\beta$ from the definition of a Besov space.

Combining the results in (20) and (22), we have the following result.

$$
\begin{aligned}
&R\big(\widehat{f}_L^W,\widehat{f}_U^W\big)-R\big(f_L^\dagger,f_U^\dagger\big) \\
&\le 9\big[\lambda\big(\gamma\sqrt{\pi}\big)^{-d}\big(2^r-1\big)^2\big\{\|f_L^\dagger\|_{L_2(\mathbb{R}^d)}^2+\|f_U^\dagger\|_{L_2(\mathbb{R}^d)}^2\big\}+B'c'\cdot C_{r,1}\cdot\|p_{\boldsymbol{X}}\|_{L_\infty(\mathbb{R}^d)}\gamma^\beta\big] \\
&\qquad\qquad+K_0\left\{\frac{\gamma^{-(1-p)(1+\epsilon)d}}{\lambda^p n}\right\}^{\frac{1}{2-p}}+36\sqrt{2}B\sqrt{\frac{\tau}{n}}+15B\frac{\tau}{n} \\
&\le c_1\lambda\gamma^{-d}+c_2\gamma^\beta+c_3\Big\{\gamma^{(1-p)(1+\epsilon)d}\lambda^p N\Big\}^{-\frac{1}{2-p}}+c_4N^{-1/2}\tau^{1/2}+c_5N^{-1}\tau\ .
\end{aligned}
$$

The latter inequality is from the fact that n is proportional to N. This establishes (12).

B.3.4 Upper bound of (D)

Since $(f_L^\#,f_U^\#)$ is the minimizer of $\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L,f_U;\mu^*,e^*)$, we find

$$
\begin{aligned}
&\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)}) \\
&=\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})+\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\mu^*,e^*)-\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\mu^*,e^*) \\
&\ge\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})+\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\mu^*,e^*)-\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\mu^*,e^*)\ .
\end{aligned}
\tag{23}
$$

Likewise, $(f_L^\dagger,f_U^\dagger)$ is the minimizer of $\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L,f_U;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})$, we find

$$
\begin{aligned}
&\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\mu^*,e^*) \\
&=\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\mu^*,e^*)+\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})-\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)}) \\
&\ge\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\mu^*,e^*)+\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})-\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})\ .
\end{aligned}
\tag{24}
$$

Results (23) and (24) imply that

$$
\begin{aligned}
&\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})-\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\mu^*,e^*) \\
&\le\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\dagger,f_U^\dagger;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})-\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\mu^*,e^*) \\
&\le\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\widehat{\mu}^{(-s)},\widehat{e}^{(-s)})-\mathcal{R}_{\mathrm{sur}}^{\mathcal{D}}(f_L^\#,f_U^\#;\mu^*,e^*)\ .
\end{aligned}
$$

Therefore, we find

$$\begin{aligned}&\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\#, f_U^\#; \mu^*, e^*)\right| \\ &\leq \max\left\{ \begin{array}{l} \left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\#, f_U^\#; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\#, f_U^\#; \mu^*, e^*)\right|, \\ \left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \mu^*, e^*)\right| \end{array} \right\}. \end{aligned} \tag{25}$$

In addition, we remark that $f_L^\#(\boldsymbol{X}) \leq f_U^\#(\boldsymbol{X})$ and $f_L^\dagger(\boldsymbol{X}) \leq f_U^\dagger(\boldsymbol{X})$ for all $\boldsymbol{X}$, because $\mathcal{L}_{\text{sur}}$ is defined in a way that $\mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L, f_U; \mu, e) \leq \mathcal{L}_{\text{sur}}(\boldsymbol{O}, f_L', f_U'; \mu, e)$ for any pair (f_L, f_U) and (f_L', f_U') satisfying $f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X})$ and $f_L(\boldsymbol{X}) > f_U(\boldsymbol{X})$. Consequently, following the steps in Section B.3.2, we establish

$$\begin{aligned}&\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\#, f_U^\#; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f_L^\#, f_U^\#; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \leq c_e \cdot \epsilon\ , \\ &\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\#, f_U^\#; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^\#, f_U^\#; \mu^*, e^*)\right| \leq c_e \cdot \epsilon\ , \\ &\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)})\right| \leq c_e \cdot \epsilon\ , \\ &\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \mu^*, e^*)\right| \leq c_e \cdot \epsilon\ . \end{aligned} \tag{26}$$

Results (25) and (26) imply that

$$\begin{aligned}&\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\#, f_U^\#; \mu^*, e^*)\right| \\ &\leq \max\left\{ \begin{array}{l} \left|\mathcal{R}^{\mathcal{D}}(f_L^\#, f_U^\#; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f_L^\#, f_U^\#; \mu^*, e^*)\right|, \\ \left|\mathcal{R}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \mu^*, e^*)\right| \end{array} \right\} + 2c_e \cdot \epsilon\ . \end{aligned} \tag{27}$$

Following the derivation for (10), we find the following result holds with probability at least $1 - \Delta_N$:

$$\begin{aligned}&\left|\mathcal{R}^{\mathcal{D}}(f_L, f_U; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}^{\mathcal{D}}(f_L, f_U; \mu^*, e^*)\right| \\ &\lesssim \left\|\widehat{e}^{(-s)} - e^*\right\|_{P,2} \left\|\widehat{\mu}^{(-s)} - \mu^*\right\|_{P,2} \\ &\lesssim N^{-r_\mu - r_e}\ . \end{aligned}$$

Therefore, (27) reduces to

$$\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\dagger, f_U^\dagger; \widehat{\mu}^{(-s)}, \widehat{e}^{(-s)}) - \mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^\#, f_U^\#; \mu^*, e^*)\right| \leq C N^{-r_\mu - r_e} + C'\epsilon \tag{28}$$

for some constants C and C' with probability at least $1 - \Delta_N$.

B.3.5 Upper bound of (E)

Following the steps in Section B.3.2, we establish

$$\left|\mathcal{R}_{\text{sur}}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*)\right| \leq c_e \cdot \epsilon\ . \tag{29}$$

B.4 Proof of Theorem 4.3

Suppose that the result in Theorem 4.2 holds, which occurs with a probability not less than $1 - 3e^{-\tau} - \Delta_N$. We denote this event by $E_{\text{Thm 4.2}}$.

For readability, we restate the loss function in (2):

$$\begin{aligned}\mathcal{L}(\boldsymbol{O}, f_L, f_U; \mu^*, e^*) &= \left\{ \begin{array}{ll} \mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu^*, e^*) & \text{if } f_L(\boldsymbol{X}) \leq f_U(\boldsymbol{X}) \\ C_{\mathcal{L}} & \text{if } f_L(\boldsymbol{X}) > f_U(\boldsymbol{X}) \end{array} \right., \\ \mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu^*, e^*) &= \left[\begin{array}{l} \left\{\mu^*(A, \boldsymbol{X}) - R\right\} \mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\} / e^*(A \,|\, \boldsymbol{X}) \\ + \int \left\{\alpha - \mu^*(a, \boldsymbol{X})\right\} \mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \end{array} \right], \end{aligned}$$

where $C_{\mathcal{L}} = 2 \sup_{\boldsymbol{O}, f_L, f_U} \left|\mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu^*, e^*)\right|$.

Suppose that $\Pr^{\mathcal{D}}\{\widehat{f}_L^{(s)}(\boldsymbol{X}) > \widehat{f}_U^{(s)}(\boldsymbol{X})\} = \int \mathbb{1}\{\widehat{f}_L^{(s)}(\boldsymbol{X}) > \widehat{f}_U^{(s)}(\boldsymbol{X})\}\, dP_{\boldsymbol{X}} \to \underline{c}$ for a positive constant $\underline{c}$ as $N \to \infty$. Then, we find

$$
\begin{aligned}
&\mathcal{R}^{\mathcal{D}}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*) \\
&= \int \big\{\mathcal{L}(\boldsymbol{O}, \widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{L}(\boldsymbol{O}, f_L^*, f_U^*; \mu^*, e^*)\big\}\, dP_{\boldsymbol{O}} \\
&= \int \mathbb{1}\big\{\widehat{f}_L^{(s)}(\boldsymbol{X}) > \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}\big\{\mathcal{L}(\boldsymbol{O}, \widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{L}(\boldsymbol{O}, f_L^*, f_U^*; \mu^*, e^*)\big\}\, dP_{\boldsymbol{O}} \\
&\qquad + \int \mathbb{1}\big\{\widehat{f}_L^{(s)}(\boldsymbol{X}) \le \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}\big\{\mathcal{L}(\boldsymbol{O}, \widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{L}(\boldsymbol{O}, f_L^*, f_U^*; \mu^*, e^*)\big\}\, dP_{\boldsymbol{O}} \\
&\overset{(a)}{\ge} \int \mathbb{1}\big\{\widehat{f}_L^{(s)}(\boldsymbol{X}) > \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}\big\{\mathcal{L}(\boldsymbol{O}, \widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{L}(\boldsymbol{O}, f_L^*, f_U^*; \mu^*, e^*)\big\}\, dP_{\boldsymbol{O}} \\
&= \int \mathbb{1}\big\{\widehat{f}_L^{(s)}(\boldsymbol{X}) > \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}\big\{C_{\mathcal{L}} - \mathcal{L}(\boldsymbol{O}, f_L^*, f_U^*; \mu^*, e^*)\big\}\, dP_{\boldsymbol{O}} \\
&\overset{(b)}{=} 0.5 C_{\mathcal{L}} \int \mathbb{1}\big\{\widehat{f}_L^{(s)}(\boldsymbol{X}) > \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}\, dP_{\boldsymbol{X}} \\
&\to 0.5 \underline{c} C_{\mathcal{L}} > 0\ .
\end{aligned}
$$

Inequality $\overset{(a)}{\ge}$ holds from the fact that (f_L^*, f_U^*) is the minimizer of $\int \mathcal{L}(f_L, f_U; \mu^*, e^*)$ given that $f_L \le f_U$. Equality $\overset{(b)}{=}$ holds from the fact that $C_{\mathcal{L}} = 2\sup_{\boldsymbol{O}, f_L, f_U} \big|\mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu^*, e^*)\big|$. Therefore, $\mathcal{R}^{\mathcal{D}}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*)$ converges to a positive value as N grows to infinity. This contradicts the result of Theorem 4.2. Therefore, we conclude that

$$
\int \mathbb{1}\{\widehat{f}_L^{(s)}(\boldsymbol{X}) > \widehat{f}_U^{(s)}(\boldsymbol{X})\}\, dP_{\boldsymbol{X}} \to 0 \quad \text{as} \quad N \to \infty\ . \tag{30}
$$

We henceforth assume that $\widehat{f}_L^{(s)} \le \widehat{f}_U^{(s)}$, which happens with probability tending to 1.

Provided that $f_L \le f_U$, we find $\mathcal{R}^{\mathcal{D}}(f_L, f_U; \mu^*, e^*)$ is represented as follows:

$$
\begin{aligned}
&\mathcal{R}^{\mathcal{D}}(f_L, f_U; \mu^*, e^*) \\
&= \mathrm{E}^{\mathcal{D}}\big\{\mathcal{L}^{(1)}(\boldsymbol{O}, f_L, f_U; \mu^*, e^*)\big\} \\
&= \mathrm{E}^{\mathcal{D}}\left[\begin{array}{l} \{\mu^*(A, \boldsymbol{X}) - R\}\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}/e^*(A \,|\, \boldsymbol{X}) \\ + \int \{\alpha - \mu^*(a, \boldsymbol{X})\}\mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \end{array}\right] \\
&= \mathrm{E}^{\mathcal{D}}\left[\mathrm{E}^{\mathcal{D}}\left[\begin{array}{l} \{\mu^*(A, \boldsymbol{X}) - R\}\mathbb{1}\{A \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}/e^*(A \,|\, \boldsymbol{X}) \\ + \int \{\alpha - \mu^*(a, \boldsymbol{X})\}\mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da \end{array}\,\middle|\, A, \boldsymbol{X}\right]\right] \\
&= \mathrm{E}^{\mathcal{D}}\left[\int \{\alpha - \mu^*(a, \boldsymbol{X})\}\mathbb{1}\{a \in [f_L(\boldsymbol{X}), f_U(\boldsymbol{X})]\}\, da\right]\ .
\end{aligned} \tag{31}
$$

Based on equation (31), we find the excess risk is simplified as follows, given that $\widehat{f}_L^{(s)} \le \widehat{f}_U^{(s)}$:

$$
\begin{aligned}
&\mathcal{R}^{\mathcal{D}}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*) \\
&= \mathrm{E}^{\mathcal{D}}\left[\int \{\alpha - \mu^*(a, \boldsymbol{X})\}\left[\begin{array}{l} \mathbb{1}\{a \in [\widehat{f}_L^{(s)}(\boldsymbol{X}), \widehat{f}_U^{(s)}(\boldsymbol{X})]\} \\ -\mathbb{1}\{a \in [f_L^*(\boldsymbol{X}), f_U^*(\boldsymbol{X})]\} \end{array}\right] da\right] \\
&= \mathrm{E}^{\mathcal{D}}\left[\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da\right]\ .
\end{aligned} \tag{32}
$$

From the proof of Lemma 3.1 in Section B.1, we find

$$
f_L^*(\boldsymbol{X}) = \inf\big\{a \,|\, \mu^*(a, \boldsymbol{X}) \ge \alpha\big\}\ , \quad f_U^*(\boldsymbol{X}) = \sup\big\{a \,|\, \mu^*(a, \boldsymbol{X}) \ge \alpha\big\}\ .
$$

Under Assumption 5, we have that:

$$\begin{aligned}
&\mu^*(f_L^*(\boldsymbol{X}), \boldsymbol{X}) = \mu^*(f_U^*(\boldsymbol{X}), \boldsymbol{X}) = \alpha \ , \\
&\mu^*(a, \boldsymbol{X}) \geq \alpha \ , \quad \forall a \in [f_L^*(\boldsymbol{X}), f_U^*(\boldsymbol{X})] \ , \\
&\mu^*(a, \boldsymbol{X}) < \alpha \ , \quad \forall a \notin [f_L^*(\boldsymbol{X}), f_U^*(\boldsymbol{X})] \ .
\end{aligned} \tag{33}$$

From (33), we find that the following conditions hold for any $\widehat{f}_L^{(s)}(\boldsymbol{X})$ and $\widehat{f}_U^{(s)}(\boldsymbol{X})$:

$$\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \geq 0 \ , \qquad \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \geq 0 \ . \tag{34}$$

In addition, Regularity Condition (RC6) implies that

$$\begin{aligned}
&\alpha - \mu^*(a, \boldsymbol{X}) = \mu^*(f_L^*(\boldsymbol{X}), \boldsymbol{X}) - \mu^*(a, \boldsymbol{X}) \geq \underline{L}\big\{f_L^*(\boldsymbol{X}) - a\big\} \geq 0 \ , && \forall a \in [f_L^*(\boldsymbol{X}) - \mathfrak{b}, f_L^*(\boldsymbol{X})] \ , \\
&\alpha - \mu^*(a, \boldsymbol{X}) = \mu^*(f_L^*(\boldsymbol{X}), \boldsymbol{X}) - \mu^*(a, \boldsymbol{X}) \leq \underline{L}\big\{f_L^*(\boldsymbol{X}) - a\big\} \leq 0 \ , && \forall a \in [f_L^*(\boldsymbol{X}), f_L^*(\boldsymbol{X}) + \mathfrak{b}] \ , \\
&\alpha - \mu^*(a, \boldsymbol{X}) = \mu^*(f_U^*(\boldsymbol{X}), \boldsymbol{X}) - \mu^*(a, \boldsymbol{X}) \geq \underline{L}\big\{a - f_U^*(\boldsymbol{X})\big\} \geq 0 \ , && \forall a \in [f_U^*(\boldsymbol{X}), f_U^*(\boldsymbol{X}) + \mathfrak{b}] \ , \\
&\alpha - \mu^*(a, \boldsymbol{X}) = \mu^*(f_U^*(\boldsymbol{X}), \boldsymbol{X}) - \mu^*(a, \boldsymbol{X}) \leq \underline{L}\big\{a - f_U^*(\boldsymbol{X})\big\} \leq 0 \ , && \forall a \in [f_U^*(\boldsymbol{X}) - \mathfrak{b}, f_U^*(\boldsymbol{X})] \ .
\end{aligned}$$

This implies that

$$\begin{aligned}
&\int_{f_L^*(\boldsymbol{X})-\mathfrak{b}}^{f_L^*(\boldsymbol{X})} \big\{\alpha - \mu^*(a, \boldsymbol{X})\big\}\, da \geq \underline{L}\mathfrak{b}^2/2 \ , && \int_{f_L^*(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})+\mathfrak{b}} \big\{\mu^*(a, \boldsymbol{X}) - \alpha\big\}\, da \geq \underline{L}\mathfrak{b}^2/2 \ , \\
&\int_{f_U^*(\boldsymbol{X})-\mathfrak{b}}^{f_U^*(\boldsymbol{X})} \big\{\mu^*(a, \boldsymbol{X}) - \alpha\big\}\, da \geq \underline{L}\mathfrak{b}^2/2 \ , && \int_{f_U^*(\boldsymbol{X})}^{f_U^*(\boldsymbol{X})+\mathfrak{b}} \big\{\alpha - \mu^*(a, \boldsymbol{X})\big\}\, da \geq \underline{L}\mathfrak{b}^2/2 \ .
\end{aligned} \tag{35}$$

Let $E_L := \big\{\boldsymbol{X} \,|\, \widehat{f}_L^{(s)}(\boldsymbol{X}) \notin [f_L^*(\boldsymbol{X}) - \mathfrak{b}, f_L^*(\boldsymbol{X}) + \mathfrak{b}]\big\}$ or $E_U := \big\{\boldsymbol{X} \,|\, \widehat{f}_U^{(s)}(\boldsymbol{X}) \notin [f_U^*(\boldsymbol{X}) - \mathfrak{b}, f_U^*(\boldsymbol{X}) + \mathfrak{b}]\big\}$ where a constant $\mathfrak{b}$ satisfies (35). We consider the following subevents of E_L:

- (Case 1) $\widehat{f}_L^{(s)}(\boldsymbol{X}) < f_L^*(\boldsymbol{X}) - \mathfrak{b}$:

  We find

$$\begin{aligned}
&\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&\geq \int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&\geq \int_{f_L^*(\boldsymbol{X})-\mathfrak{b}}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&\geq \underline{L}\mathfrak{b}^2/2 \ .
\end{aligned}$$

  The first and second inequalities are from (33) and (34). The third inequality is from (35).

- (Case 2) $f_L^*(\boldsymbol{X}) + \mathfrak{b} < \widehat{f}_L^{(s)}(\boldsymbol{X}) \leq f_U^*(\boldsymbol{X})$:

We find

$$\begin{aligned}
&\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&= \int_{f_L^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&\geq \int_{f_L^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da \\
&\geq \int_{f_L^*(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})+\mathfrak{b}} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da \\
&\geq \underline{L}\mathfrak{b}^2/2 \ .
\end{aligned}$$

The first four lines are from (33) and (34). The last line holds from (35).

- (Case 3) $f_U^*(\boldsymbol{X}) < \widehat{f}_L^{(s)}(\boldsymbol{X})$:
  We find

$$\begin{aligned}
&\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&= \int_{f_L^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&= \int_{f_L^*(\boldsymbol{X})}^{f_U^*(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&= \int_{f_L^*(\boldsymbol{X})}^{f_U^*(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da - \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&= \int_{f_L^*(\boldsymbol{X})}^{f_U^*(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da + \int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\
&\geq \int_{f_L^*(\boldsymbol{X})}^{f_U^*(\boldsymbol{X})} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da \\
&\geq \int_{f_L^*(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})+\mathfrak{b}} \{\mu^*(a, \boldsymbol{X}) - \alpha\}\, da \\
&\geq \underline{L}\mathfrak{b}^2/2 \ .
\end{aligned}$$

  The equalities are obtained from straightforward algebra. The first and second inequalities are from (33). The last line holds from (35).

Therefore, $\boldsymbol{X} \in E_L$ implies that

$$\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \geq \underline{L}\mathfrak{b}^2/2 \ . \tag{36}$$

Likewise, one can also show that $\boldsymbol{X} \in E_U$ implies that

$$\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \geq \underline{L}\mathfrak{b}^2/2 \ . \tag{37}$$

Therefore, we find

$$
\begin{aligned}
&\mathcal{R}^{\mathcal{D}}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*) \\
&= \mathrm{E}^{\mathcal{D}}\left[\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da\right] \\
&= \Pr^{\mathcal{D}}(\boldsymbol{X} \in E_L \cup E_U) \times \mathrm{E}^{\mathcal{D}}\left[\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \,\middle|\, \boldsymbol{X} \in E_L \cup E_U\right] \\
&\quad + \Pr^{\mathcal{D}}(\boldsymbol{X} \in E_L^c \cap E_U^c) \times \mathrm{E}^{\mathcal{D}}\left[\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \,\middle|\, \boldsymbol{X} \in E_L^c \cap E_U^c\right] \\
&\overset{(a)}{\geq} 0.5 \Pr^{\mathcal{D}}(\boldsymbol{X} \in E_L \cup E_U) \times \underline{L}\mathfrak{b}^2 \\
&\quad + \Pr^{\mathcal{D}}(\boldsymbol{X} \in E_L^c \cap E_U^c) \times \mathrm{E}^{\mathcal{D}}\left[\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \,\middle|\, \boldsymbol{X} \in E_L^c \cap E_U^c\right] \\
&\overset{(b)}{\geq} 0.5 \Pr^{\mathcal{D}}(\boldsymbol{X} \in E_L \cup E_U) \times \underline{L}\mathfrak{b}^2 \ .
\end{aligned}
$$

Inequality $\overset{(a)}{\geq}$ holds from (36) and (37). Inequality $\overset{(b)}{\geq}$ holds from (34).

Suppose that $\Pr^{\mathcal{D}}(\boldsymbol{X} \in E_L \cup E_U) \to \underline{c}$ for a positive constant $\underline{c}$ as $N \to \infty$. Then, we find

$$
\mathcal{R}^{\mathcal{D}}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*) \to 0.5\underline{c}\underline{L}\mathfrak{b}^2 > 0 \ .
$$

Again, this contradicts the result of Theorem 4.2. Therefore, we conclude that

$$
\Pr^{\mathcal{D}}(\boldsymbol{X} \in E_L \cup E_U) \to 0 \quad \text{as} \quad N \to \infty \ . \tag{38}
$$

We now consider the following event E_f:

$$
E_f := \left\{ \begin{array}{l} \{\widehat{f}_L^{(s)}(\boldsymbol{X}) \leq \widehat{f}_U^{(s)}(\boldsymbol{X})\} \\ \cap\, \{\widehat{f}_L^{(s)}(\boldsymbol{X}) \in [f_L^*(\boldsymbol{X}) - \mathfrak{b}, f_L^*(\boldsymbol{X}) + \mathfrak{b}]\} \\ \cap\, \{\widehat{f}_U^{(s)}(\boldsymbol{X}) \in [f_U^*(\boldsymbol{X}) - \mathfrak{b}, f_U^*(\boldsymbol{X}) + \mathfrak{b}]\} \end{array} \right\} \ ,
$$

which occurs with probability not less than $1 - \delta_N$ where $\delta_N \to 0$ as $N \to \infty$ from (30) and (38).

Given $E_{\text{Thm 4.2}} \cap E_f$, which happens with probability not less than $1 - 3e^{-\tau} - \Delta_N - \delta_N$, we find the following results:

- (Case 1 for f_L) $\quad \widehat{f}_L^{(s)} \in [f_L^* - \mathfrak{b}, f_L^*]$

  We find

$$
\begin{aligned}
\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \left\{\alpha - \mu^*(a, \boldsymbol{X})\right\} da &= \int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \left\{\mu^*(f_L^*(\boldsymbol{X}), \boldsymbol{X}) - \mu^*(a, \boldsymbol{X})\right\} da \\
&\geq \underline{L} \int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \left\{f_L^*(\boldsymbol{X}) - a\right\} da \\
&= 0.5\underline{L}\left\{f_L^*(\boldsymbol{X}) - \widehat{f}_L^{(s)}(\boldsymbol{X})\right\}^2 \ .
\end{aligned}
$$

  The first line holds from (33). The second line holds from (35). The last line holds from straightforward algebra.

- (Case 2 for f_L) $\quad \widehat{f}_L^{(s)} \in [f_L^*, f_L^* + \mathfrak{b}]$
  We find

$$\begin{aligned}\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \big\{\alpha - \mu^*(a, \boldsymbol{X})\big\}\, da &= \int_{f_L^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \big\{\mu^*(a, \boldsymbol{X}) - \alpha\big\}\, da \\ &= \int_{f_L^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \big\{\mu^*(a, \boldsymbol{X}) - \mu^*(f_L^*(\boldsymbol{X}), \boldsymbol{X})\big\}\, da \\ &\geq \underline{L} \int_{f_L^*(\boldsymbol{X})}^{\widehat{f}_L^{(s)}(\boldsymbol{X})} \big\{a - f_L^*(\boldsymbol{X})\big\}\, da \\ &= 0.5\underline{L}\big\{f_L^*(\boldsymbol{X}) - \widehat{f}_L^{(s)}(\boldsymbol{X})\big\}^2 .\end{aligned}$$

  The first line holds from straightforward algebra. The second line holds from (33). The third line holds from (35). The last line again holds from straightforward algebra.

- (Case 1 for f_U) $\quad \widehat{f}_U^{(s)} \in [f_U^*, f_U^* + \mathfrak{b}]$
  By following the analogous steps as in (Case 1 for f_L), one can establish

$$\int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \big\{\alpha - \mu^*(a, \boldsymbol{X})\big\}\, da \geq 0.5\underline{L}\big\{f_U^*(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}^2 .$$

- (Case 2 for f_U) $\quad \widehat{f}_U^{(s)} \in [f_U^* - \mathfrak{b}, f_U^*]$
  By following the analogous steps as in (Case 2 for f_L), one can establish

$$\int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \big\{\alpha - \mu^*(a, \boldsymbol{X})\big\}\, da \geq 0.5\underline{L}\big\{f_U^*(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}^2 .$$

Therefore, given $E_{\text{Thm 4.2}} \cap E_f$, we have

$$\begin{aligned}&\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da \\ &\geq 0.5\underline{L}\big[\big\{f_L^*(\boldsymbol{X}) - \widehat{f}_L^{(s)}(\boldsymbol{X})\big\}^2 + \big\{f_U^*(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}^2\big] .\end{aligned} \tag{39}$$

Accordingly, we find the following representation of the excess risk in (32):

$$\begin{aligned}&\mathcal{R}^{\mathcal{D}}(\widehat{f}_L^{(s)}, \widehat{f}_U^{(s)}; \mu^*, e^*) - \mathcal{R}^{\mathcal{D}}(f_L^*, f_U^*; \mu^*, e^*) \\ &= \mathrm{E}^{\mathcal{D}}\bigg[\int_{\widehat{f}_L^{(s)}(\boldsymbol{X})}^{f_L^*(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da + \int_{f_U^*(\boldsymbol{X})}^{\widehat{f}_U^{(s)}(\boldsymbol{X})} \{\alpha - \mu^*(a, \boldsymbol{X})\}\, da\bigg] \\ &\geq 0.5\underline{L} \times \mathrm{E}^{\mathcal{D}}\Big[\big\{f_L^*(\boldsymbol{X}) - \widehat{f}_L^{(s)}(\boldsymbol{X})\big\}^2 + \big\{f_U^*(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}^2\Big] .\end{aligned}$$

The last line holds from (39). Consequently, this completes the proof that the following result holds with probability not less than $1 - 3e^{-\tau} - \Delta_N - \delta_N$:

$$\begin{aligned}&\mathrm{E}^{\mathcal{D}}\Big[\big\{f_L^*(\boldsymbol{X}) - \widehat{f}_L^{(s)}(\boldsymbol{X})\big\}^2 + \big\{f_U^*(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}^2\Big] \\ &\leq c_1'\lambda\gamma^{-d} + c_2'\gamma^{\beta} + c_3'\big\{\gamma^{(1-p)(1+\epsilon)d}\lambda^p N\big\}^{-\frac{1}{2-p}} + c_4' N^{-1/2}\tau^{1/2} + c_5' N^{-1}\tau + c_6'\epsilon N^b + c_7' N^{-r_e - r_\mu}\end{aligned}$$

where constants $c_1', \ldots, c_7'$ do not depend on N. Taking $\gamma \asymp N^{-1/(2\beta+d)}$ and $\lambda \asymp N^{-(\beta+d)/(2\beta+d)}$, we get the following asymptotic rate with probability not less than $1 - 3e^{-\tau} - \Delta_N - \delta_N$:

$$\mathrm{E}^{\mathcal{D}}\Big[\big\{f_L^*(\boldsymbol{X}) - \widehat{f}_L^{(s)}(\boldsymbol{X})\big\}^2 + \big\{f_U^*(\boldsymbol{X}) - \widehat{f}_U^{(s)}(\boldsymbol{X})\big\}^2\Big] = O_P\big(N^{-\beta/(2\beta+d)} + N^{-r_e - r_\mu}\big) . \tag{40}$$

This completes the proof.