Predictive Performance Comparison of Decision Policies Under Confounding

Luke Guerdan [1] **Amanda Coston** [2] **Kenneth Holstein** [1] **Zhiwei Steven Wu** [1]

Abstract

Predictive models are often introduced to decision-making tasks under the rationale that they improve performance over an existing decision-making policy. However, it is challenging to compare predictive performance against an existing decision-making policy that is generally under-specified and dependent on unobservable factors. These sources of uncertainty are often addressed in practice by making strong assumptions about the data-generating mechanism. In this work, we propose a method to compare the predictive performance of decision policies under a variety of modern identification approaches from the causal inference and off-policy evaluation literatures (e.g., instrumental variable, marginal sensitivity model, proximal variable). Key to our method is the insight that there are regions of uncertainty that we can safely ignore in the policy comparison. We develop a practical approach for finite-sample estimation of regret intervals under no assumptions on the parametric form of the status quo policy. We verify our framework theoretically and via synthetic data experiments. We conclude with a real-world application using our framework to support a pre-deployment evaluation of a proposed modification to a healthcare enrollment policy.

1. Introduction

Predictive models are often introduced under the rationale that they improve performance over an existing decision-making policy (Grove et al., 2000). For example, models have been developed with the goal of improving the accuracy of human decisions in domains such as healthcare, criminal justice, and education (Baker & Hawn, 2021; Rambachan et al., 2021). Given the high-stakes nature of these domains, regulatory frameworks have called for organizations to provide explicit comparisons of predictive models against the status quo they are intended to replace (Wyden; Johnson & Zhang, 2022). For example, consider two common policy comparison settings of real-world significance.

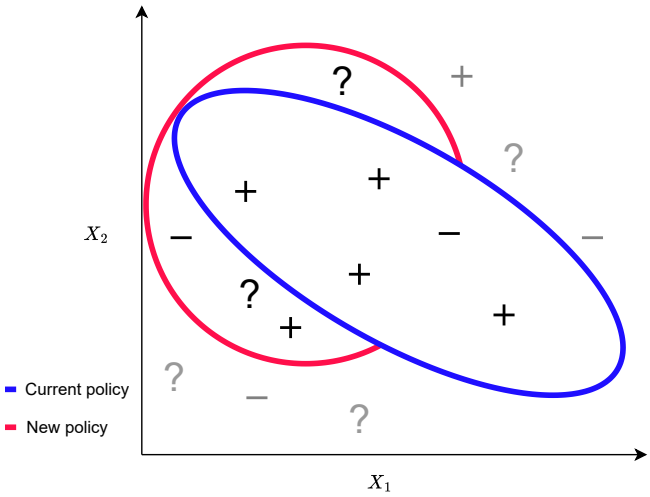


Figure 1. Illustration of uncertainty in comparing two policies in a toy setting with $X \in \mathcal{R}^2$. Points are labelled by their outcome: positive (+), negative (-) or unknown (?). Ovals denote the selection region of a policy. Points that neither policy selects (denoted by grey) are irrelevant to the comparison. Our method leverages this to reduce policy comparison uncertainty.

Example 1.1 (Human vs algorithm decisions). Let the status quo policy consist of human-only decisions (e.g., pre-trial release decisions (Kleinberg et al., 2018), medical testing decisions (Mullainathan & Obermeyer, 2019)) made using covariates and unobserved contextual information. We would like to evaluate whether a model recommending actions by thresholding risk predictions would constitute an improvement over the current human-only policy (e.g., Rambachan et al. (2021); Chouldechova et al. (2018)).

Example 1.2 (Human+algorithm vs algorithm decisions). Let the status quo policy consist of humans making decisions with the support of a predictive model, covariates, and unobserved contextual information. We would like to assess whether decisions made using the predictive model alone would yield a performance improvement over the human-algorithm combination (e.g., Cheng et al. (2022)).

[1]Carnegie Mellon University. [2]Microsoft Research. Correspondence to: Luke Guerdan <lguerdan@cs.cmu.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

Both of these examples can be formalized as confounded off-policy evaluation in the contextual bandits setting (Jung et al., 2020; Swaminathan & Joachims, 2015; Tennenholtz et al., 2021; Rambachan et al., 2022). Given observational data collected under a confounded status quo policy, our goal is to compare predictive performance against an alternative policy which would assign different actions. A common strategy for off-policy evaluation under confounding involves *partial identification* of policy performance within an uncertainty interval (Kallus & Zhou, 2021; Rambachan et al., 2021; 2022; Pu & Zhang, 2021; Zhang et al., 2021). Yet constructing informative performance intervals is challenging because it often requires imposing untestable assumptions on the structure and magnitude of unmeasured confounders impacting the status quo policy.

In this work, we develop a novel partial identification method which reduces uncertainty in policy performance comparisons by isolating confounding-related uncertainty in the disagreement region of the action space (Figure 1). Our uncertainty cancellation approach yields tighter regret intervals than existing frameworks designed for non-comparative performance evaluations (Rambachan et al., 2022; Namkoong et al., 2020), which account for redundant uncertainty in the agreement region of the action space. Our simple approach operates by (1) decomposing policy performance measures into sets of identified and partially identified comparison statistics, then (2) cancelling common partially identified terms in the regret estimand.

We show that a diverse set of causal assumptions studied in prior off-policy evaluation literature imply uncertainty sets over partially identified comparison statistics. As a result, while existing off-policy evaluation frameworks are often tied to a fixed causal assumption (e.g., the marginal sensitivity model (Tan, 2006)) that may be unjustified in some contexts, ours is interoperable with a broad family of modern identification strategies (e.g., Rosenbaum's Γ model (Namkoong et al., 2020; Zhang et al., 2021), instrumental variable (Lakkaraju et al., 2017; Kleinberg et al., 2018), proximal variable (Ghassami et al., 2023), or the marginal sensitivity model (MSM) (Kallus & Zhou, 2021)).

We propose a flexible plug-in algorithm for finite sample estimation of regret bounds under no parametric assumptions on the data-generating policy. While this procedure inherits slow non-parametric convergence rates of regression functions used to estimate bounds on comparison statistics, we also show how to construct assumption-tailored doubly robust estimators which attain fast root-n convergence rates under no parametric assumptions. We conduct synthetic experiments verifying the coverage of our regret intervals. We conclude by illustrating how our framework can be used to support a pre-deployment evaluation of a proposed modification to a healthcare enrollment policy (Obermeyer et al., 2019). Our results demonstrate that, in some cases, our improved regret interval supports more conclusive pre-deployment assessments of decision policies than would be possible via existing off-policy evaluation approaches.

We make the following main contributions:

- We formulate the problem of comparative predictive performance evaluations for decision making policies under unmeasured confounding (§ 3).
- We propose a novel partial identification technique that yields informative bounds on predictive performance differences by isolating comparison-relevant uncertainty (§ 4). Our technique is interoperable with a range of modern identification approaches from causal inference and off-policy evaluation literature (§ 5).
- We develop flexible methods for finite sample estimation of regret bounds under no parametric assumptions on the confounded status quo policy (§ 6).
- We validate our framework theoretically and via synthetic data experiments (§ 7). We demonstrate how our framework can support a pre-deployment evaluation of a proposed healthcare enrollment policy (§ 8). We present all proofs in the Appendix.[1]

2. Related Work

Off-policy evaluation (OPE) is a widely studied problem in reinforcement learning (Uehara et al., 2022; Jiang & Li, 2016; Precup, 2000; Namkoong et al., 2020; Kallus & Zhou, 2020), contextual bandits (Si et al., 2020; Tennenholtz et al., 2021; Swaminathan & Joachims, 2015; Xu et al., 2021; Dudık et al., 2014), and econometrics (Kitagawa & Tetenov, 2018; Huber, 2021) literature. We study the confounded offline contextual bandits setting (Jung et al., 2020; Swaminathan & Joachims, 2015; Tennenholtz et al., 2021), in which observational data is collected under a behavior policy $\pi_0(X_i, U_i)$ which acts on covariates X_i and unmeasured confounders U_i.

Existing OPE frameworks typically estimate the *value function* $V(\pi) = \mathbb{E}[Y(\pi(X))]$ of a new policy $\pi(X_i)$ via observational data (Kallus & Zhou, 2020; Namkoong et al., 2020; Zhang et al., 2021; Uehara et al., 2022; Si et al., 2020; Kallus & Zhou, 2020; Hatt et al., 2022; Ek et al., 2023; Tennenholtz et al., 2021; Ishikawa & He, 2023).[2] In this work, we instead target predictive performance measures (e.g., Accuracy, TNR, PPV), which are often of interest during policy evaluation of algorithmic risk assessments. While OPE of predictive performance measures introduces estimand-specific challenges (i.e., conditioning

[1] All code for experiments is publicly available here.

[2] Work in the confounded offline setting typically studies updated policies that only observe covariates. When $\pi(X_i)$ also observes U_i regret intervals tend to be vacuous.

on a partially-observed potential outcome), approaches developed for OPE of value functions are also applicable in the algorithmic decision-making context (e.g., doubly robust estimation, reweighting).[3]

In particular, Rambachan et al. (2022) develop a partial identification framework for robust learning and evaluation of algorithmic risk assessments under confounding. This framework leverages doubly-robust approaches for OPE of predictive performance measures. One natural strategy for extending this framework to the policy comparison setting involves computing bounds on $V(\pi)$ and $V(\pi_0)$ *independently* before computing the regret interval by taking a difference. However, we show that this approach yields an overly conservative regret interval by introducing redundant uncertainty in the agreement region of the policy action space. We introduce an alternative δ-regret interval, which we show yields more informative bounds than this baseline approach. Moreover, we extend the Mean Outcome Sensitivity Model (MOSM) introduced by Rambachan et al. (2022) to develop uncertainty sets around partially identified policy comparison statistics which are consistent with a range of modern causal assumptions (e.g., instrumental variable, MSM, Rosenbaum's Γ sensitivity model).

Our work also builds upon prior literature studying policy learning and selection under confounding (Zhang et al., 2021; Konyushova et al., 2021; Yang et al., 2022; Kuzborskij et al., 2021; Zhang et al., 2021; Kallus & Zhou, 2021; 2018; Hatt et al., 2022; Ek et al., 2023; Gao & Yin, 2023; Balachandar et al., 2023). Zhang et al. (2021) develop a framework for ranking individualized treatment policies under Rosenbaum's Γ sensitivity model. Namkoong et al. (2020) also use Rosenbaum's Γ sensitivity model to partially identify $V(\pi)$ in the sequential setting. Kallus & Zhou (2021; 2018) develop an approach for learning minimax optimal decision-policies under the MSM. In principle, confounding-robust policy learning frameworks such as these can also be used to recover δ-regret intervals against the confounded data-generating policy (Hatt et al., 2022; Ek et al., 2023; Gao & Yin, 2023). However, these techniques (1) target the mean potential outcome rather than predictive performance measures, and (2) are tied to a fixed causal assumption which may not be applicable in all contexts. In contrast, our framework is interoperable with a range of assumptions (e.g., instrumental variable, MSM, proximal variable) studied in prior OPE literature.

Finally, our work relates to literature studying comparison of human versus algorithmic decision-making policies. Kleinberg et al. (2018) develop a framework for comparing human decisions and algorithmic predictions under selective labels and unobservables. The contraction technique underpinning this approach compares the failure rate of judicial decisions against a predictive model under an assumption that judges have heterogeneous selection rates (Lakkaraju et al., 2017). Later developments formalize this approach into an instrumental variable framework (Chen et al., 2023; Rambachan et al., 2021; 2022). Rambachan et al. (2021) develop a framework for identifying systematic prediction mistakes in historical human decisions. This framework enables utility-based comparisons of human versus algorithmic decision-making policies under varying sets of econometric assumptions. Most recently, Ben-Michael et al. (2024) devise an experimental framework for evaluating human+algorithm hybrid workflows against the quality of decisions that a human or algorithm would make alone. This framework supports *post-deployment* policy comparisons following randomized assignment to a treatment (algorithmic recommendation) or control (no algorithmic recommendation) condition. In contrast, our framework is designed to support *pre-deployment* policy comparisons given observational data collected under a status quo policy.

3. Preliminaries

Let $\pi_0 : \mathcal{X} \times \mathcal{U} \to \mathcal{A}$ be a status quo decision-making policy assigning binary actions given measured covariates $X \in \mathcal{X}$ and unmeasured confounders $U \in \mathcal{U}$. Let $\pi : \mathcal{X} \to \mathcal{A}$ be a proposed policy that assigns binary actions only via covariates.[4] For example, in a medical testing context, π_0 is an existing (e.g., physician based) testing policy, while π is a proposed algorithmic policy. Let D^{π_0}, T^{π} be random variables indicating actions selected under the status quo and proposed policies, respectively.[5] Let $Y(1) \in \mathcal{Y}$ be the outcome of interest to the policy comparison. This is the potential outcome which *would be observed* given a positive decision ($D^{\pi_0} = 1$) selected under the status quo policy (Rubin, 2005). For example, $Y(1)$ denotes the disease status of a patient, or repayment if a lendee is granted a loan. Let $Y \in \mathcal{Y}$ be the binary outcome observed in observational data. In our selective labels setting (Lakkaraju et al., 2017), Y is only observed when an instance received a positive decision ($D^{\pi_0} = 1$) under π_0. We make the following standard consistency assumption on Y (Rubin, 2005).

Assumption 3.1 (Consistency)**.** $D^{\pi_0} = 1 \implies Y = Y(1)$.

For example, consistency would be violated if the testing decision for one patient impacts the disease status of another (Hudgens & Halloran, 2008). Additionally, we assume that

[3] In the algorithmic decision-making context, the term *selective labels* is often used to describe a setting in which a potential outcome is only observed under one of the possible actions (Lakkaraju et al., 2017; Wei, 2021; Coston et al., 2021; De-Arteaga et al., 2018).

[4] Our framework is compatible with new policies which are stochastic and deterministic. However we assume that π is known.

[5] We sometimes omit policy superscripts over random variables to ease notation.

each instance has some probability of both decisions under the status quo policy.

Assumption 3.2 (Positivity)**.** $p(D^{\pi_0} = a \mid X = x, U = u) > 0,\ \forall a \in \mathcal{A}, x \in \mathcal{X}, u \in \mathcal{U}$.

3.1. Problem Formulation

Given the target distribution $(X, U, D^{\pi_0}, T^{\pi}, Y(1), Y) \sim p^*(\cdot)$, our goal is to evaluate the regret $R^*(\pi, \pi_0; m) = m^*(\pi) - m^*(\pi_0)$, where $m^* : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}_+$ is a policy performance measure computed with respect to $Y(1)$. We study the following performance measures, which are analogously defined for the baseline policy.

Predictive performance: Let $m^*_y(\pi) = p(T^{\pi} = 1 \mid Y(1) = y)$ be the positive ($y = 1$) and negative ($y = 0$) class predictive performance.[6] For example, a positive false positive rate regret $m^*_{y=0}(\pi) - m^*_{y=0}(\pi_0) > 0$ indicates that the proposed policy recommends tests for healthy patients more frequently than the status quo. Similarly, let $m^*_a(\pi) = \mathbb{E}[Y(1) = a \mid T^{\pi} = a]$ be the positive ($a = 1$) and negative ($a = 0$) predictive value of π.

Utility: Let $u_{a,y} \in \mathbb{R}_+$ be the utility of outcome y under action a. We let

$$m^*_u(\pi) = \mathbb{E}\left[\sum_{a,y} u_{ay} \cdot \mathbb{I}\{A^{\pi} = a, Y(1) = y\}\right]$$

be the expected utility of π.[7] A positive utility regret $m^*_u(\pi) - m^*_u(\pi_0) > 0$ indicates that the updated policy has an overall welfare benefit in comparison to the status quo (Rambachan et al., 2021). For example, this measure reflects settings in which turning away sick patients incurs lower utility than testing healthy ones ($u_{01} << u_{10}$) as well as accuracy ($u_{11} = u_{00} = 1$, $u_{10} = u_{01} = 0$).

Because of the selective labels problem, the policy regret is *partially identified* within the interval $R^*(\pi, \pi_0; m) \in [\underline{R}(\pi, \pi_0; m), \overline{R}(\pi, \pi_0; m)]$. Our goal in this work is to recover the most informative regret interval possible given observational data $O = \{(X_i, T^{\pi}_i, D^{\pi_0}_i, Y_i) : i = 1, ..., n\} \sim p(X, D^{\pi_0}, T^{\pi}, Y)$ generated under the status-quo policy.

3.2. Partial Identification of Policy Performance

To tightly characterize policy performance differences, we introduce notation that enables us to isolate confounding-related uncertainty in performance measures. We decompose performance measures into a set of sufficient v-statistics $\mathbf{v} = \{v_y(t, d) : \forall y, t, d\}$, where each term $v_y(t, d) = p(T^{\pi} = t, D^{\pi_0} = d, Y(1) = y)$ is the joint probability of policy actions and the potential outcome.

The subset $\mathbf{v}_1 = \{v_y(t, 1) : \forall y, t\}$ contains all v-statistics known from observational data. By consistency (3.1), each term in $\mathbf{v}_1$ is given by

$$p(T = t, D = 1, Y(1) = 1) = p(T = t, D = 1, Y = 1).$$

Similarly, $\mathbf{v}_0 = \{v_y(t, 0) : \forall y, t\}$ contains all partially identified v-statistics. Each term in $\mathbf{v}_0$ is bounded in the interval $v_y(t, 0) \in [0, p(T = t, D = 0)]$ because the potential outcome is unobserved when $D = 0$. More generally, we let $v_y(t, 0) \in [\underline{v}_y(t, 0), \overline{v}_y(t, 0)]$ be bounds on partially identified v-statistics. We let $\mathbf{v}^*_0$ be the true value of partially identified terms under the target distribution and let

$$\mathcal{V}(p) = \left\{\mathbf{v}_0 : \sum_y v_y(t, 0) = p(T = t, D = 0),\ \forall t\right\}$$

be an uncertainty set of feasible values consistent with the observed data distribution. We can bound policy performance measures over this uncertainty set via

$$\underline{m}(\pi_0; \mathcal{V}) = \min_{\mathbf{v}_0 \in \mathcal{V}(p)} m(\mathbf{v}_0, \mathbf{v}_1; \pi_0),$$
$$\overline{m}(\pi_0; \mathcal{V}) = \max_{\mathbf{v}_0 \in \mathcal{V}(p)} m(\mathbf{v}_0, \mathbf{v}_1; \pi_0),$$

where $m(\mathbf{v}^*_0, \mathbf{v}_1; \pi_0)$ is a v-statistic decomposition of $m^*(\pi_0)$. In the following section, we show that this reduction to a set of common sufficient statistics enables uncertainty cancellation when comparing across policies.

4. Regret Bound Identification

We now introduce our novel approach for comparing policies under confounding. Our approach yields informative regret intervals by eliminating redundant uncertainty that does not contribute to differential policy performance. We first focus on asymptotic performance bounds, and provide finite-sample analyses in Section 6.

A baseline approach for bounding regret involves partially identifying the performance of each policy individually, and then taking a difference across policy-specific bounds. We refer to this as the baseline regret interval.

Definition 4.1 (Baseline regret interval)**.** The baseline regret interval over $\mathcal{V}(p)$ has lower and upper endpoints:

$$\underline{R}(\pi, \pi_0; m, \mathcal{V}) = \underline{m}(\pi; \mathcal{V}) - \overline{m}(\pi_0; \mathcal{V}),$$
$$\overline{R}(\pi, \pi_0; m, \mathcal{V}) = \overline{m}(\pi; \mathcal{V}) - \underline{m}(\pi_0; \mathcal{V}).$$

In contrast, our proposed approach directly bounds the difference in the oracle regret $\delta_m(\mathbf{v}^*_0, \mathbf{v}_1) = m(\mathbf{v}^*_0, \mathbf{v}_1; \pi) - m(\mathbf{v}^*_0, \mathbf{v}_1; \pi_0)$.

[6] While our definition of $m^*_y(\pi)$ assumes that $T^{\pi} = 1$ (i.e, the TPR and FPR) the TNR ($y = 0$) and FNR ($y = 1$) can be recovered by taking $1 - R^*(\pi, \pi_0; m_y)$.

[7] We model utilities of binary classification outcomes rather than the vector-valued potential outcome $Y(a) \in \{0, 1\}^{|A|}$ studied in the standard off policy evaluation setup.

Definition 4.2 (δ-regret interval)**.** The δ-regret interval over $\mathcal{V}(p)$ has lower and upper endpoints:

$$\underline{R}_\delta(\pi, \pi_0; m, \mathcal{V}) = \min_{\mathbf{v}_0 \in \mathcal{V}(p)} \delta_m(\mathbf{v}_0, \mathbf{v}_1),$$
$$\overline{R}_\delta(\pi, \pi_0; m, \mathcal{V}) = \max_{\mathbf{v}_0 \in \mathcal{V}(p)} \delta_m(\mathbf{v}_0, \mathbf{v}_1).$$

The δ-regret interval yields tighter regret bounds by eliminating redundant uncertainty irrelevant to policy comparison. For example, consider the accuracy regret decomposition

$$\delta_u(\mathbf{v}_0^*, \mathbf{v}_1) = \sum_y v_y(y, 1-y) - v_y(1-y, y).$$

Because $v_0(0,0) \in \mathbf{v}_0$ cancels when taking a difference across polices, this term does not contribute uncertainty to the δ-regret interval, as it would with the baseline interval.[8] We now provide a theoretical result characterizing the improvement offered by the δ-interval more generally. We let $I_\delta(m, \mathcal{V}) = \overline{R}_\delta(\pi, \pi_0; m, \mathcal{V}) - \underline{R}_\delta(\pi, \pi_0; m, \mathcal{V})$ and $I(m, \mathcal{V}) = \overline{R}(\pi, \pi_0; m, \mathcal{V}) - \underline{R}(\pi, \pi_0; m, \mathcal{V})$ be the length of the δ-regret and baseline regret intervals, respectively.

Theorem 4.3 (Regret separation)**.** *Let $\Delta(m, \mathcal{V}) = I(m, \mathcal{V}) - I_\delta(m, \mathcal{V})$. Then the δ-regret interval offers the following improvement over the baseline regret interval*

$$\Delta(m_y, \mathcal{V}) \geq \frac{2\alpha \cdot v_y(1,1)}{(\overline{\gamma}_y)^2}, \quad \Delta(m_{a=1}, \mathcal{V}) = 0,$$
$$\Delta(m_u, \mathcal{V}) \geq 2\alpha \cdot (u_{00} + u_{01}),$$
$$\Delta(m_{a=0}, \mathcal{V}) \geq \frac{2\alpha}{\max\{\psi_0(\pi), \psi_0(\pi_0)\}},$$

where $\alpha = \overline{v}_0(0,0) - \underline{v}_0(0,0)$, $\psi_0(\pi) = p(A^\pi = 0)$, and $\overline{\gamma}_y = \sum_a \sum_{a'} \overline{v}_y(a, a')$.

Observe that the improvement offered by the δ-regret interval is proportional to the magnitude of uncertainty in the cancellation term $\alpha = \overline{v}_0(0,0) - \underline{v}_0(0,0)$. Therefore, our approach offers the largest benefit under a high degree of uncertainty around partially identified v-statistics.

Because regret measures inherit a monotonic dependence on v-statistics, we can compute closed form analytic bounds by maximizing and minimizing regret measures over uncertainty intervals around partially identified terms $[\underline{v}_y(t,0), \overline{v}_y(t,0)]$. We provide δ-regret and baseline regret intervals for policy performance measures in Appendix A.

5. Mapping Causal Assumptions to Informative Regret Bounds

The regret interval we recover without imposing additional causal assumptions on the data generating policy tends to be uninformative. Therefore, we now introduce a *pointwise bounding functions* assumption which we use to tighten regret intervals by shrinking the uncertainty set around partially-identified v-statistics. We show that this assumption is implied by a range of causal assumptions studied in prior OPE literature (e.g., MSM, Rosenbaum's Γ, IV) in Appendix B. Thus, our δ-regret interval yields tighter uncertainty quantification than the baseline interval for any causal assumption implying pointwise bounding functions (Figure 2).

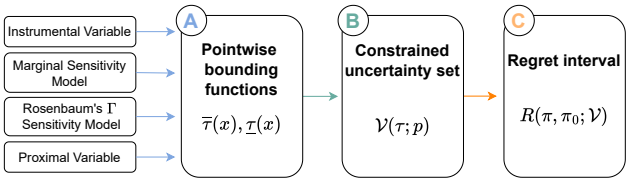


Figure 2. Flow of assumptions in our framework. (A) Traditional causal assumptions imply *pointwise bounding functions* on the unobserved outcome (Appendix B); (B) Pointwise bounding functions imply constrained uncertainty sets (Lemma 5.2); (C) Constrained uncertainty sets imply policy regret bounds (Section 4).

Assumption 5.1 (Pointwise bounding functions)**.** Let $\underline{\tau}, \overline{\tau} : \mathcal{X} \to [0,1]$ be a pair of bounding functions satisfying

$$\underline{\tau}(x) \leq \mathbb{E}[Y(1) \mid D = 0, X = x] \leq \overline{\tau}(x), \quad \forall x \in X.$$

Intuitively, this assumption requires that each individual screened out under the status quo policy has risk which is bounded by $[\underline{\tau}(x), \overline{\tau}(x)]$. Rambachan et al. (2022) show that a variant of this assumption is implied by the IV framework, the MSM, and Rosenbaum's Γ sensitivity model.[9] Next, we show that bounding functions can be used to construct intervals on partially identified v-statistics.

Lemma 5.2 (Assumption mapping)**.** *Let $\underline{\tau}, \overline{\tau} : \mathcal{X} \to [0,1]$ be a pair of bounding functions satisfying Assumption 5.1. Then partially identified v-statistics are bounded by*

$$\mathbb{E}[\underline{\tau}(x) \cdot e_0(x) \cdot \pi_t(x)] \leq v_1(t,0) \leq \mathbb{E}[\overline{\tau}(x) \cdot e_0(x) \cdot \pi_t(x)],$$

where $e_d(x) = p(D = d \mid X = x)$ and $\pi_t(x) = p(T = t \mid X = x)$.

Lemma 5.2 provides a convenient approach for bounding $\mathbf{v}_0^*$ because, given a pair of bounding functions, it only requires knowledge of *nominal* (i.e., confounded) status quo policy probabilities $e_d(x)$ and the new policy $\pi_t(x)$. Nominal probabilities and bounding functions can be learned from observational data, while $\pi_t(x)$ is known in advance by model developers. Our goal is to tighten regret intervals

[8] We provide additional intuition for the benefits of our uncertainty cancellation approach in Appendix E.

[9] Our pointwise bounding function assumption is also related to the notion of "bounded outcomes" studied by Manski (1990) and "natural outcomes" studied by Pearl (2009).

Algorithm 1 Plug-in regret bound estimator

Input: Data $\mathcal{O} = \{(X_i, D_i^{\pi_0}, T_i^{\pi}, Y_i)\}_{i=1}^n \sim p$, Folds K
for $k = 1$ **to** K **do**
  Learn $\hat{\eta}_k = (\hat{e}_{1,k}, \hat{\tau}_k)$ using $\mathcal{O}_{-k}$
  Estimate $\mathcal{H}(\hat{v}_{1,k}(t,0); \hat{\tau})$ via e.q. (1) using $\mathcal{O}_k$
  Construct $\hat{\mathcal{V}}_k(p; \hat{\tau})$
  Compute $\underline{\hat{R}}_{\delta,k}(\pi, \pi_0; m, \hat{\mathcal{V}}_k), \hat{\overline{R}}_{\delta,k}(\pi, \pi_0; m, \hat{\mathcal{V}}_k)$
end for
Set $\underline{\hat{R}}_{\delta}(\pi, \pi_0; m, \hat{\mathcal{V}}) = \frac{1}{K}\sum_k \underline{\hat{R}}_{\delta,k}(\pi, \pi_0; m, \hat{\mathcal{V}}_k)$,
Set $\hat{\overline{R}}_{\delta}(\pi, \pi_0; m, \hat{\mathcal{V}}) = \frac{1}{K}\sum_k \hat{\overline{R}}_{\delta,k}(\pi, \pi_0; m, \hat{\mathcal{V}}_k)$

by using bounding functions to shrink the uncertainty set around $\mathbf{v}_0^*$. Let $\mathcal{H}(v_1(t,0); \tau) = [\underline{v}_y(t,0), \overline{v}_1(t,0)]$ be the interval implied by Lemma 5.2 and let $\rho_{td} = p(T = t, D = d)$. We define the constrained uncertainty set over $\mathbf{v}_0^*$ as

$$\mathcal{V}(p;\tau) = \left\{ v_y(t,0) \colon \begin{array}{c} v_1(0,0) \in \mathcal{H}(v_1(0,0);\tau) \\ v_0(0,0) = \rho_{00} - v_1(0,0) \\ v_1(1,0) \in \mathcal{H}(v_1(1,0);\tau) \\ v_0(1,0) = \rho_{10} - v_1(1,0) \end{array} \right\}.$$

We only use intervals $\mathcal{H}(v_1(t,0);\tau)$ for this set definition because an analogous bound is implied over $v_0(t,0)$ by the constraint requiring that $\sum_y v_y(t,0) = \rho_{t0}$. We quantify the size of $\mathcal{V}(p;\tau)$ via the Lebesgue measure, which is defined as the cartesion product over partially identified intervals $\lambda(\mathcal{V}(p;\tau)) = |\overline{v}_1(1,0) - \underline{v}_1(1,0)||\overline{v}_1(0,0) - \underline{v}_1(0,0)|$. We now show that $\mathcal{V}(p;\tau)$ is the smallest uncertainty set one could construct based on the observational distribution and a pair of bounding functions.

Theorem 5.3 (Minimality)**.** *$\mathcal{V}(p;\tau)$ is the minimal uncertainty set—that is, no strict subset of $\mathcal{V}(p;\tau)$ is consistent with $p(X, D, T, Y)$ and user-specified bounding functions.*

6. Regret Bound Estimation

In this section, we construct finite sample estimates of regret intervals using observational data $\mathcal{O} = \{(X_i, D_i^{\pi_0}, T_i^{\pi}, Y_i)\}_{i=1}^n \sim p$ collected under the status quo policy. We estimate bounds by applying closed-form expressions for the δ-regret interval over an estimate of the uncertainty set $\hat{\mathcal{V}}(p;\hat{\tau})$. We can directly compute finite-sample estimates $\hat{\mathbf{v}}_1$ for identified v-statistics by taking a sample average. As a result, the technical challenge underpinning regret bound estimation boils down to estimating the intervals $\mathcal{H}(\hat{v}_y(t,0);\hat{\tau}) = [\underline{\hat{v}}_y(t,0), \hat{\overline{v}}_y(t,0)]$ around $\mathbf{v}_0^*$.

6.1. Plug-in Estimator

A direct approach for bound estimation involves learning $\hat{\eta} = (\hat{e}, \hat{\tau})$, then estimating $\mathcal{H}(\hat{v}_1(t,0);\hat{\tau})$ over a held-out sample by invoking Lemma 5.2.[10][11] We assume that $\pi(x)$ is known by model developers and thus does not need to be learned. We split the data into K disjoint folds, where we denote $\mathcal{O}_k, \mathcal{O}_{-k}$, as the sample inside and outside of fold k, respectively. We then define the plug-in estimator over $\mathcal{O}_k$ as

$$\hat{\overline{v}} = \frac{1}{|\mathcal{O}_k|} \sum_{x_i \in \mathcal{O}_k} \pi(x_i) \cdot \hat{e}(x_i) \cdot \hat{\overline{\tau}}(x_i) \tag{1}$$

where $\hat{\eta}$ is learned over fold $\mathcal{O}_{-k}$.[12] We recover regret estimates at full data efficiency via a cross-fitting approach outlined in Algorithm 1. The following result shows consistency of $\hat{\overline{v}}$ and characterizes its convergence rate to the true upper bound $\overline{v}$, where we define $\|f\|^2 := \int (f(x))^2 dP(x)$ to be the squared $L_2(O)$ norm of a function f.

Theorem 6.1. *Let $f(x) = e(x) \cdot \overline{\tau}(x)$ and assume $||\hat{f} - f|| = o_p(1)$. Then the plug-in estimator satisfies*

$$\hat{\overline{v}} - \overline{v} = O_P\left(\|e(x) - \hat{e}(x)\| + \|\overline{\tau}(x) - \hat{\overline{\tau}}(x)\|\right) + O_P\left(\frac{1}{\sqrt{n}}\right).$$

Because policy performance measures are linear combinations (m_u^*) or ratios (m_a^*, m_y^*) of v-statistics, this result also characterizes the convergence rate of the estimated upper and lower regret bounds. Observe that the convergence rate of the plug-in estimator inherits the rate of the nuisance functions. As a result, $\hat{v}$ will tend to converge slowly when using flexible machine learning methods to fit nuisance functions under no parametric assumptions.

6.2. Doubly Robust Estimator

Therefore, we propose a doubly robust approach for estimating $\hat{v}$ that corrects the bias in the plug-in estimator. Because doubly robust estimators have an error term which is a product of nuisance function errors, they attain fast convergence rates, even when estimating nuisance functions at slow non-parametric rates (Kennedy et al., 2023). For example, to attain $n^{-1/2}$ rates for the estimator, it is sufficient to estimate nuisance functions at $n^{-1/4}$, a rate achieved by many non-parametric machine learning techniques. The form of the doubly robust estimator depends on the pointwise bounding functions implied by a causal assumption.

We provide the doubly robust estimator for the commonly studied MSM. The MSM assumes there is some $\Lambda \geq 1$

[10] In line with prior literature on CATE estimation (Kennedy, 2023), we call this a plug-in approach because it involves directly substituting in empirical estimates for the nuisance functions.

[11] The term "nuisance function" refers to a learned function of the data which differs from our target quantity of interest.

[12] We suppress dependence on y, t, d and present analysis of the upper bound to ease notation. An analogous argument holds for $\underline{\hat{v}}$.

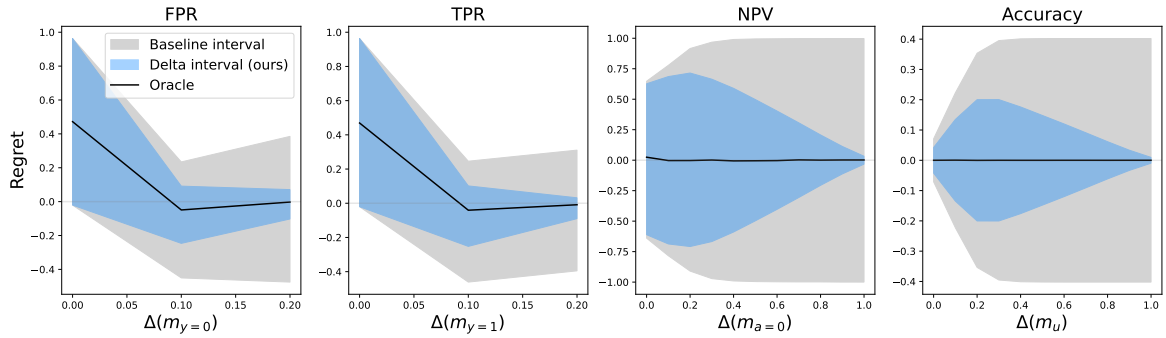


Figure 3. Improvement in bounds offered by the δ-regret interval over the baseline interval. We systematically vary the relative size of v-statistics and plot bounds as a function of interval improvement $\Delta(m) = I(m) - I_\delta(m)$ characterized by Theorem 4.3.

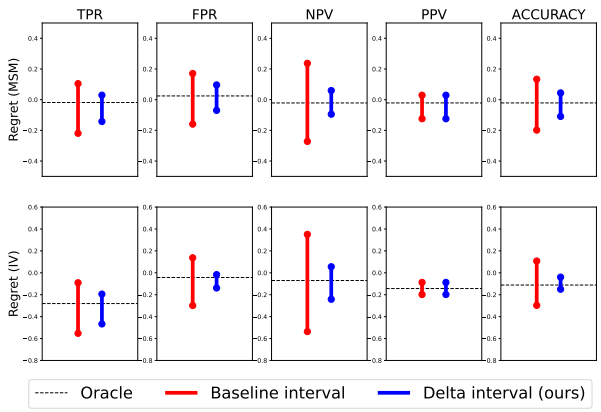


Figure 4. Comparison between δ-regret and baseline regret interval end-points averaged over $N = 20$ trials and $N_s = 20,000$ samples. First row leverages an MSM identification assumption, while the second leverages an IV assumption.

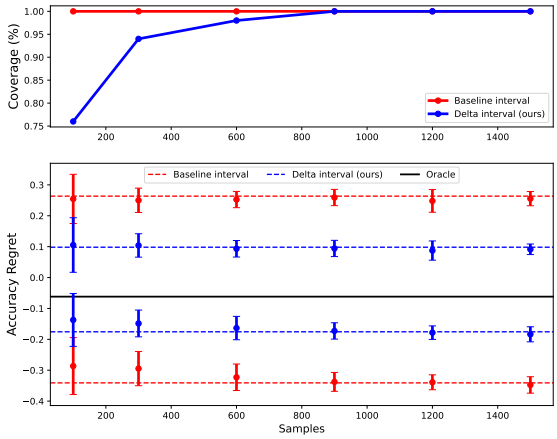


Figure 5. Top: Coverage of accuracy regret interval estimates as a function of total sample size. Bottom: 95% bootstrap confidence intervals around upper and lower regret bounds over $N = 25$ trials. Solid line indicates the oracle regret.

such that the odds ratio $\pi_0(X,U))/(1-\pi_0(X,U)) \cdot (1 - e(X))/e(X)$ lies within $[\Lambda^{-1}, \Lambda]$ (See B.3). We define the doubly robust estimator over $\mathcal{O}_k$ as

$$\hat{\bar{v}}_{DR} = \frac{1}{|\mathcal{O}_k|} \sum_{x_i \in \mathcal{O}_k} \phi(O_i; \hat{e}, \pi, \hat{\mu}_1) \tag{2}$$

where $\phi(O; e, \pi, \mu_1) = D \cdot Y\pi(X) \cdot \Lambda + (T - \pi(X)) \cdot e(X) \cdot \mu_1(X) \cdot \Lambda$ and $\mu_1(x) = \mathbb{E}[Y(1) \mid D = 1, X = x]$.

Algorithm 1 can be leveraged for doubly-robust estimation of policy regret bounds by substituting e.q. 2 for estimation of $\mathcal{H}(\hat{v}_{1,k}(t,0); \hat{\tau})$. We show in Appendix C that this estimator has second order error in nuisance estimation error.

7. Numerical Experiments

We now show numerical experiments comparing decision policies under a known data-generating mechanism.

Regret interval characterization. We first characterize the uncertainty reduction offered by the δ-regret interval. We simulate a range of v-statistic decompositions by randomly sampling $\mathbf{v}$ satisfying $\sum_{y,t,d} v_y(t,d) = 1$ and $v_0(t,d) + v_1(t,d) = \rho_{td}, \forall t, d$. This yields a collection of decompositions corresponding to valid observational distributions. We compute the analytic improvement in bounds defined in Theorem 4.3 for each $\mathbf{v}$ and plot the results in Figure 3. We observe a monotone improvement in bounds as a function of $\Delta(m, \mathcal{V})$, with no improvement when $\Delta(m, \mathcal{V}) = 0$ and significant improvement when $\Delta(m, \mathcal{V}) = 0$ is large.

Synthetic data experiment. We next validate our framework by simulating data consistent with two common assumptions for off-policy evaluation: the MSM (B.3) and IV (B.5). We draw v covariates $X_i \sim \mathcal{N}(0, I_v)$, and u confounders $U_i \sim \mathcal{N}(0, I_u)$, and let $V_i = (X_i, U_i)$. We parameterize policy, outcome, and instrument probability functions with coefficients $W_{\pi_0} \in \mathbb{R}^{v+u}$, $W_\pi \in \mathbb{R}^v$, $W_{\mu_1} \in \mathbb{R}^{v+u}$, $W_z \in \mathbb{R}^{v \times z}$, respectively, with each drawn

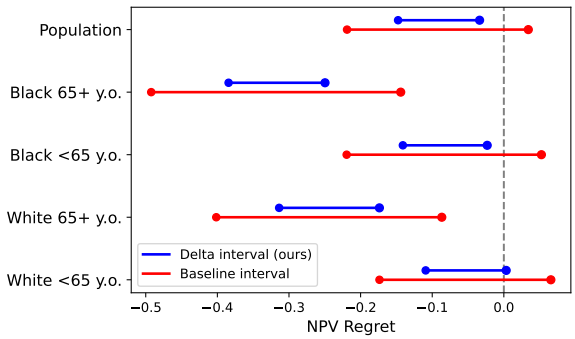


Figure 6. Comparison of the δ-regret interval against the baseline for the NPV ($m_{a=0}$). The top row indicates full population bounds, while lower rows subpopulation bounds.

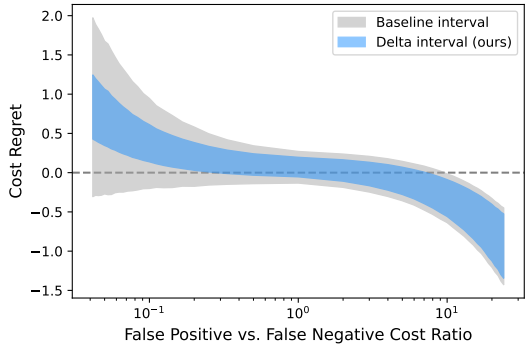


Figure 7. Expected cost regret as a function of false positive to false negative cost ratio. Left hand side (10^{-1}) indicates false negative cost ten times greater than false positive cost.

from a uniform distribution. We sample data from the probability functions

$$\gamma(X_i) := \sigma'(X_i \times W_z), \pi(X_i) := \sigma(X_i \times W_\pi)$$
$$\pi_0(V_i, Z_i) := \sigma(V_i \times W_{\pi_0} + \beta_0 \cdot Z_i)$$
$$\mu_1(V_i, Z_i) := \sigma(V_i \times W_{\mu_1} + \beta_1 \cdot Z_i)$$
$$\mu_0(V_i, Z_i) := \begin{cases} \Lambda^* \cdot \mu_1(V_i), \Lambda^* \in U(\Lambda^{-1}, \Lambda) & \text{(MSM)} \\ \sigma(V_i \times W_{\mu_0} + \beta_1 \cdot Z_i) & \text{(IV)} \end{cases}$$
$$\mu(V_i, Z_i) := \mu_1(V_i, Z_i)\pi_1(V_i, Z_i) + \mu_0(V_i, Z_i)\pi_0(V_i, Z_i)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ and $\sigma'(x)$ is the softmax function. We use Algorithm 1 with the plug-in estimator to learn estimates estimates of the δ-regret interval. We compare against the baseline regret interval by applying bounds provided in Appendix A over the same $\hat{\mathcal{V}}(p, \hat{\tau})$ used to estimate the δ-regret. See Appendix D for additional setup details.

Figure 4 provides a comparison of the δ-regret interval and baseline regret interval across five policy performance measures. In line with our theoretical results, we obtain tighter bounds from the δ-regret across all policy performance measures apart from the positive predictive value (PPV). We show the oracle regret evaluated via $Y(1)$ in black. In Figure 5, we plot coverage of the estimated δ-regret interval evaluated against the true population value, denoted by dashed lines. The bottom panel shows 95% bootstrap confidence intervals around the true upper and lower δ-regret and baseline regret intervals. These results show that estimates concentrate around the true regret interval as the number of samples increases. We observe that the baseline interval yields better coverage of the oracle regret in small sample settings because it is more conservative and thus more tolerant to bias around estimates of the asymptotic interval end points. Appendix D contains additional experimental results stress-testing coverage under causal assumption violations.

8. Real-World Application: Comparing Healthcare Enrollment Policies

We now illustrate how our framework can be used to compare alternative healthcare enrollment policies. In medical settings, providers routinely screen patients for diseases and enroll high-risk individuals in preventative care programs. However, it is challenging to assess the performance improvement of a proposed policy because outcomes are only observed among patients enrolled under the existing policy (Daysal et al., 2022). Confounding is a challenge in this setting because physicians often make decisions using unobserved information (Mullainathan & Obermeyer, 2022).

We leverage data released by Obermeyer et al. (2019) to construct an enrollment policy comparison task. This dataset contains $\approx 48,000$ records, where each entry consists of a patient evaluated for enrollment in a high-risk care management program.[13] We let π_0 be the historical enrollment policy consisting of physician decisions informed by an algorithmic risk score and interactions with the patient. We take π to be the algorithmic policy which makes decisions by thresholding predictions of patient cost from the clinical decision support tool. The goal in this task is to assess whether the algorithm-only policy would improve upon the status quo human+algorithm policy used to collect data. We detail our setup in Appendix D.0.4.

Results. Because we do not have access to physician identifiers which can be used as an instrument, we leverage the MSM assumption for identification. We compare the baseline and δ-regret intervals for the NPV in Figure 6. The top row shows the population regret, while the bottom rows

[13]We use a synthetic version of the original dataset, which was released by Obermeyer et al. (2019) to protect patient confidentiality. This dataset preserves the means and covariances of the original data.

show a breakdown across subgroups.[14] The δ-regret interval excludes zero across all subpopulations apart from White patients under 65. As a result, an analysis conducted via the δ-regret interval supports an interpretation that the algorithmic policy reduces the NPV in comparison to the status quo human+algorithm policy.[15] **Because the baseline regret interval contains zero among all patient populations except those aged 65 and over, this bounding approach supports weaker claims regarding the relative performance of decision policies.**

In Figure 7, we plot the δ-regret and baseline intervals as a function of the ratio between false positive and false negative costs. The left hand region of this figure ($\approx 10^{-1}$) reflects a setting in which false negatives are ten times more costly than false positives. This regime is realistic in our healthcare enrollment scenario when turning a sick patient away from the program incurs more harm than enrolling a healthy patient. We observe the greatest improvement from our approach in the high cost false negative setting because the $v_0(0,0)$ term which cancels under our approach is heavily weighted in the expected cost calculation. Because the δ-regret interval excludes zero in this regime, this supports an interpretation that the proposed policy has a higher expected cost than the status quo.

9. Conclusion

In this work, we propose the first framework supporting predictive performance comparisons of decision-making policies under confounding. Our approach is intended to support pre-deployment evaluations of proposed policies under a flexible set of causal assumptions. Our approach addresses sources of confounding-related uncertainty which impact model evaluations and, where possible, reduces this uncertainty via technically novel partially identification approaches. Our uncertainty cancellation approach may prove useful for more tightly characterizing performance differences under other uncertainty sources, such as missing protected attributes (Kallus et al., 2022) and measurement error (Fogliato et al., 2020).

Impact Statement

Our framework is designed to support pre-deployment evaluations of proposed decision policies. While our framework is intended to faithfully represent confounding-related sources of uncertainty impacting decision policy evaluations, it does not speak to broader measurement challenges (Guerdan et al., 2023) and ethical questions underpinning the introduction of an algorithmic system (Coston et al., 2023; Rittel & Webber, 1973). Given the high-stakes contexts in which some policies are deployed (e.g., lending, healthcare, education), our framework should be applied carefully as part of a multifaceted impact assessment.

[14] Intervals include 95% confidence intervals estimated around regret bound endpoints. We omit these from the plot for readability.

[15] Given the narrow scope of our analyses and limitations of synthetic data, our findings are not intended to be a conclusive assessment of the policies evaluated by Obermeyer et al. (2019).

Acknowledgements

We thank our anonymous reviewers, the attendees of the NeurIPS 2023 Workshop on Regulatable Machine Learning, and the attendees of Carnegie Mellon University's Fairness, Explainability, Accountability, and Transparency (FEAT) reading group for their helpful feedback. This work was supported by an award from the UL Research Institutes through the Center for Advancing Safety of Machine Intelligence (CASMI) at Northwestern University and the National Science Foundation Graduate Research Fellowship Program (Award No. DGE1745016).

References

Abaluck, J., Agha, L., Kabrhel, C., Raja, A., and Venkatesh, A. The determinants of productivity in medical testing: Intensity and allocation of care. *American Economic Review*, 106(12):3730–3764, 2016.

Angluin, D. and Laird, P. Learning from noisy examples. *Machine Learning*, 2(4):343–370, 1988.

Baker, R. S. and Hawn, A. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, pp. 1–41, 2021.

Balachandar, S., Garg, N., and Pierson, E. Domain constraints improve risk prediction when outcome data is missing. In *The Twelfth International Conference on Learning Representations*, 2023.

Ben-Michael, E., Greiner, D. J., Huang, M., Imai, K., Jiang, Z., and Shin, S. Does ai help humans make better decisions? a methodological framework for experimental evaluation. *arXiv preprint arXiv:2403.12108*, 2024.

Chen, J., Li, Z., and Mao, X. Learning under selective labels with data from heterogeneous decision-makers: An instrumental variable approach. *arXiv preprint arXiv:2306.07566*, 2023.

Cheng, H.-F., Stapleton, L., Kawakami, A., Sivaraman, V., Cheng, Y., Qing, D., Perer, A., Holstein, K., Wu, Z. S., and Zhu, H. How child welfare workers reduce racial disparities in algorithmic decisions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–22, 2022.

Chouldechova, A., Benavides-Prado, D., Fialko, O., and Vaithianathan, R. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency*, pp. 134–148. PMLR, 2018.

Coston, A., Ramamurthy, K. N., Wei, D., Varshney, K. R., Speakman, S., Mustahsan, Z., and Chakraborty, S. Fair transfer learning with missing protected attributes. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 91–98, 2019.

Coston, A., Rambachan, A., and Chouldechova, A. Characterizing fairness over the set of good models under selective labels. In *International Conference on Machine Learning*, pp. 2144–2155. PMLR, 2021.

Coston, A., Kawakami, A., Zhu, H., Holstein, K., and Heidari, H. A validity perspective on evaluating the justified use of data-driven decision-making algorithms. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pp. 690–704. IEEE, 2023.

Daysal, N. M., Mullainathan, S., Obermeyer, Z., Sarkar, S. K., and Trandafir, M. An economic approach to machine learning in health policy. *Univ. of Copenhagen Dept. of Economics Discussion, CEBI Working Paper*, (24), 2022.

De-Arteaga, M., Dubrawski, A., and Chouldechova, A. Learning under selective labels in the presence of expert consistency. *arXiv preprint arXiv:1807.00905*, 2018.

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

Ek, S., Zachariah, D., Johansson, F. D., and Stoica, P. Off-policy evaluation with out-of-sample guarantees. *Transactions on Machine Learning Research*, 2023.

Fogliato, R., Chouldechova, A., and G'Sell, M. Fairness evaluation in presence of biased noisy labels. In *International Conference on Artificial Intelligence and Statistics*, pp. 2325–2336. PMLR, 2020.

Gao, R. and Yin, M. Confounding-robust policy improvement with human-ai teams. *arXiv preprint arXiv:2310.08824*, 2023.

Ghassami, A., Shpitser, I., and Tchetgen, E. T. Partial identification of causal effects using proxy variables. *arXiv preprint arXiv:2304.04374*, 2023.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., and Nelson, C. Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment*, 12(1):19, 2000.

Guerdan, L., Coston, A., Wu, Z. S., and Holstein, K. Ground(less) truth: A causal framework for proxy labels in human-algorithm decision-making. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pp. 688–704, 2023.

Hatt, T., Tschernutter, D., and Feuerriegel, S. Generalizing off-policy learning under sample selection bias. In *Uncertainty in Artificial Intelligence*, pp. 769–779. PMLR, 2022.

Huber, M. An introduction to flexible methods for policy evaluation. In *Handbook of Research Methods and Applications in Empirical Microeconomics*, pp. 82–111. Edward Elgar Publishing, 2021.

Hudgens, M. G. and Halloran, M. E. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

Ishikawa, K. and He, N. Kernel conditional moment constraints for confounding robust inference. In *International Conference on Artificial Intelligence and Statistics*, pp. 650–674. PMLR, 2023.

Jiang, N. and Li, L. Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pp. 652–661. PMLR, 2016.

Johnson, R. A. and Zhang, S. What is the bureaucratic counterfactual? categorical versus algorithmic prioritization in us social policy. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1671–1682, 2022.

Jung, J., Shroff, R., Feller, A., and Goel, S. Bayesian sensitivity analysis for offline policy evaluation. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 64–70, 2020.

Kallus, N. and Zhou, A. Confounding-robust policy improvement. *Advances in neural information processing systems*, 31, 2018.

Kallus, N. and Zhou, A. Confounding-robust policy evaluation in infinite-horizon reinforcement learning. *Advances in neural information processing systems*, 33: 22293–22304, 2020.

Kallus, N. and Zhou, A. Minimax-optimal policy learning under unobserved confounding. *Management Science*, 67 (5):2870–2890, 2021.

Kallus, N., Mao, X., and Zhou, A. Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*, 68(3):1959–1981, 2022.

Kennedy, E., Balakrishnan, S., and Wasserman, L. Semiparametric counterfactual density estimation. *Biometrika*, 110(4):875–896, 2023.

Kennedy, E. H. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008–3049, 2023.

Kennedy, E. H., Balakrishnan, S., and G'Sell, M. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008–2030, 2020.

Kitagawa, T. and Tetenov, A. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1): 237–293, 2018.

Konyushova, K., Chen, Y., Paine, T., Gulcehre, C., Paduraru, C., Mankowitz, D. J., Denil, M., and de Freitas, N. Active offline policy selection. *Advances in Neural Information Processing Systems*, 34:24631–24644, 2021.

Kuzborskij, I., Vernade, C., Gyorgy, A., and Szepesvári, C. Confident off-policy evaluation and selection through self-normalized importance weighting. In *International Conference on Artificial Intelligence and Statistics*, pp. 640–648. PMLR, 2021.

Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., and Mullainathan, S. The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 275–284, 2017.

Manski, C. F. Anatomy of the selection problem. *Journal of Human Resources*, pp. 343–360, 1989.

Manski, C. F. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990.

Manski, C. F. and Pepper, J. V. Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, 68(4):997, 2000.

Miao, W., Shi, X., and Tchetgen, E. T. A confounding bridge approach for double negative control inference on causal effects. *arXiv preprint arXiv:1808.04945*, 2018.

Mullainathan, S. and Obermeyer, Z. *A machine learning approach to low-value health care: wasted tests, missed heart attacks and mis-predictions*. National Bureau of Economic Research, 2019.

Mullainathan, S. and Obermeyer, Z. Diagnosing physician error: A machine learning approach to low-value health care. *The Quarterly Journal of Economics*, 137(2):679–727, 2022.

Namkoong, H., Keramati, R., Yadlowsky, S., and Brunskill, E. Off-policy policy evaluation for sequential decisions under unobserved confounding. *Advances in Neural Information Processing Systems*, 33:18819–18831, 2020.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

Pearl, J. *Causality*. Cambridge university press, 2009.

Precup, D. Eligibility traces for off-policy policy evaluation. *Computer Science Department Faculty Publication Series*, pp. 80, 2000.

Pu, H. and Zhang, B. Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(2):318–345, 2021.

Rambachan, A., Coston, A., and Kennedy, E. Robust design and evaluation of predictive algorithms under unobserved confounding. *arXiv preprint arXiv:2212.09844*, 2022.

Rambachan, A. et al. Identifying prediction mistakes in observational data. *Harvard University*, 2021.

Rittel, H. W. and Webber, M. M. Dilemmas in a general theory of planning. *Policy Sciences*, 4(2):155–169, 1973.

Robins, J. M. [covariance adjustment in randomized experiments and observational studies]: Comment. *Statistical Science*, 17(3):309–321, 2002.

Rosenbaum, P. R. Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, 74(1):13–26, 1987.

Rosenbaum, P. R. Sensitivity analysis in observational studies. *Encyclopedia of Statistics in Behavioral Science*, 2005.

Rubin, D. B. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

Scott, C., Blanchard, G., and Handy, G. Classification with asymmetric label noise: Consistency and maximal denoising. In *Conference on Learning Theory*, pp. 489–511. PMLR, 2013.

Si, N., Zhang, F., Zhou, Z., and Blanchet, J. Distributionally robust policy evaluation and learning in offline contextual bandits. In *International Conference on Machine Learning*, pp. 8884–8894. PMLR, 2020.

Swaminathan, A. and Joachims, T. Counterfactual risk minimization: Learning from logged bandit feedback. In *International Conference on Machine Learning*, pp. 814–823. PMLR, 2015.

Tan, Z. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.

Tchetgen, E. J. T., Ying, A., Cui, Y., Shi, X., and Miao, W. An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*, 2020.

Tennenholtz, G., Shalit, U., Mannor, S., and Efroni, Y. Bandits with partially observable confounded data. In *Uncertainty in Artificial Intelligence*, pp. 430–439. PMLR, 2021.

Uehara, M., Shi, C., and Kallus, N. A review of off-policy evaluation in reinforcement learning. *arXiv preprint arXiv:2212.06355*, 2022.

Wei, D. Decision-making under selective labels: Optimal finite-domain policies and beyond. In *International Conference on Machine Learning*, pp. 11035–11046. PMLR, 2021.

Wyden, S. R. S.2892 - algorithmic accountability act of 2022.

Xia, X., Liu, T., Wang, N., Han, B., Gong, C., Niu, G., and Sugiyama, M. Are anchor points really indispensable in label-noise learning? *Advances in Neural Information Processing Systems*, 32, 2019.

Xu, L., Kanagawa, H., and Gretton, A. Deep proxy causal learning and its application to confounded bandit policy evaluation. *Advances in Neural Information Processing Systems*, 34:26264–26275, 2021.

Yang, M., Dai, B., Nachum, O., Tucker, G., and Schuurmans, D. Offline policy selection under uncertainty. In *International Conference on Artificial Intelligence and Statistics*, pp. 4376–4396. PMLR, 2022.

Zhang, B., Weiss, J., Small, D. S., and Zhao, Q. Selecting and ranking individualized treatment rules with unmeasured confounding. *Journal of the American Statistical Association*, 116(533):295–308, 2021.

Zhang, Y., Liao, Q. V., and Bellamy, R. K. Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 295–305, 2020.

A. Asymptotic Regret Bounds

In this appendix, we derive δ-regret and baseline regret intervals for policy performance measures. We prove Theorem 4.3 in § A.0.5.

A.0.1. δ-REGRET BOUNDS ON UTILITY REGRET.

Lemma A.1 (δ_u-regret bounds)**.** *Let $m_u^*(\pi)$ be the expected utility of π given utility values $u_{ay} \geq 0$. Let $a' = 1 - a$, $\lambda_{ay} = u_{ay} - u_{a'y}$, and $\tilde{y} = \mathbb{I}\{\lambda_{11} > \lambda_{10}\}$. Then*

$$\delta_u(\mathbf{v}_0, \mathbf{v}_1) = \sum_{ay} \lambda_{ay} \cdot v_y(a, a')$$

and for all uncertainty sets $\mathcal{V}(p;\tau)$, the upper δ_u-regret bound is given by

$$\overline{R}_\delta(\pi, \pi_0; m_u, \mathcal{V}) = \delta_u(\overline{\mathbf{v}}_0, \mathbf{v}_1), \text{ where } \overline{\mathbf{v}}_0 \in \operatorname*{arg\,max}_{v_{\tilde{y}}(1,0)} \mathcal{V}(p;\tau).$$

Proof.

$$\begin{aligned} m_u^*(\pi) &= \mathbb{E}\left[\sum_{t,y} u_{ty} \cdot \mathbb{I}\{T^\pi = t, Y(1) = y\}\right] \\ &= \sum_{t,y} u_{ty} \cdot \mathbb{E}[\mathbb{I}\{T = t, Y(1) = y\}] \\ &= \sum_{t,y} u_{ty} \cdot p(T = t, Y(1) = y) \\ &= \sum_{t,y} u_{ty} \cdot (v_y(t,0) + v_y(t,1)). \end{aligned}$$

By the same argument, $m_u^*(\pi_0) = \sum_{d,y} u_{dy} \cdot (v_y(0,d) + v_y(1,d))$. Let $\lambda_{ay} = u_{ay} - u_{a'y}$ and $\tilde{y} = \mathbb{I}\{\lambda_{11} > \lambda_{10}\}$. Therefore

$$\begin{aligned} \delta_u(\mathbf{v}_0^*, \mathbf{v}_1) &= m_u^*(\pi) - m_u^*(\pi_0) \\ &= \sum_{t,y} u_{ty} \cdot (v_y(t,0) + v_y(t,1)) - \sum_{d,y} u_{dy} \cdot (v_y(0,d) + v_y(1,d)) \\ &= \sum_{a,y} (u_{ay} - u_{a'y}) \cdot v_y(a, a') \\ &= \sum_{a,y} \lambda_{ay} \cdot v_y(a, a') \end{aligned}$$

Recall that $\rho_{td} = v_1(t,0) + v_0(t,0)$ by the constraint on $p(\cdot)$. Therefore, the regret is bounded within the interval

$$\begin{aligned} R^*(\pi, \pi_0; m_y) &\leq \begin{cases} \sum_y \lambda_{0y} \cdot v_y(0,1) + \lambda_{11} \cdot \overline{v}_1(1,0) + \lambda_{10} \cdot \underline{v}_0(1,0) & \lambda_{11} > \lambda_{10} \\ \sum_y \lambda_{0y} \cdot v_y(0,1) + \lambda_{11} \cdot \underline{v}_1(1,0) + \lambda_{10} \cdot \overline{v}_0(1,0) & \lambda_{11} \leq \lambda_{10} \end{cases} \\ &= \sum_y \lambda_{0y} \cdot v_y(0,1) + \lambda_{1\tilde{y}} \cdot \overline{v}_{\tilde{y}}(1,0) + \lambda_{1,1-\tilde{y}} \cdot \overline{v}_{1-\tilde{y}}(1,0), \text{ where } \tilde{y} = \mathbb{I}\{\lambda_{11} > \lambda_{10}\} \end{aligned}$$

The lower regret bound is symmetric.

□

A.0.2. δ-REGRET BOUNDS ON POSITIVE AND NEGATIVE CLASS PERFORMANCE

Lemma A.2 (δ_y regret bounds)**.** *Suppose that* $p(Y(1) = y) > 0$*. Then* $\overline{R}_\delta(\pi, \pi_0; m_y, \mathcal{V}) = \delta_y(\overline{\mathbf{v}}_0, \mathbf{v}_1)$*, where*

$$\delta_y(\mathbf{v}_0, \mathbf{v}_1) = \frac{v_y(1,0) - v_y(0,1)}{v_y(0,0) + v_y(1,0) + v_y(0,1) + v_y(1,1)},$$

$$\overline{\mathbf{v}}_0 = \begin{cases} \arg \min_{v_y(0,0)} \max_{v_y(1,0)} \mathcal{V}(p;\tau),\ \overline{v}_y(1,0) - v_y(0,1) \geq 0 \\ \arg \max_{v_y(0,0)} \max_{v_y(1,0)} \mathcal{V}(p;\tau),\ \underline{v}_y(1,0) - v_y(0,1) < 0 \end{cases}$$

Next, we derive δ-regret bounds on m_y.

Proof.

$$\begin{aligned} \delta(\mathbf{v}_0^*, \mathbf{v}_1) &= m_y^*(\pi) - m_y^*(\pi_0) \\ &= p(T^\pi = 1 \mid Y(1) = y) - p(D^{\pi_0} = 1 \mid Y(1) = y) \\ &= \frac{p(T = 1, Y(1) = y)}{p(Y(1) = y)} - \frac{p(D = 1, Y(1) = y)}{p(Y(1) = y)} \\ &= \frac{v_y(1,0) + v_y(1,1)}{p(Y(1) = y)} - \frac{v_y(0,1) + v_y(1,1)}{p(Y(1) = y)} \\ &= \frac{v_y(1,0) - v_y(0,1)}{v_y(0,0) + v_y(1,0) + v_y(0,1) + v_y(1,1)} \end{aligned}$$

Therefore,

$$R^*(\pi, \pi_0; m_y) <= \begin{cases} \dfrac{\overline{v}_y(1,0) - v_y(0,1)}{\underline{v}_y(0,0) + \overline{v}_y(1,0) + v_y(0,1) + v_y(1,1)}, & \overline{v}_y(1,0) - v_y(0,1) > 0 \\ \dfrac{\overline{v}_y(1,0) - v_y(0,1)}{\overline{v}_y(0,0) + \overline{v}_y(1,0) + v_y(0,1) + v_y(1,1)}, & \overline{v}_y(1,0) - v_y(01) <= 0 \end{cases}$$

The lower regret bound is symmetric.

□

Remark A.3. The result above implies bounds on the FPR ($y = 0$) and TPR ($y = 1$) regret. Taking $[-\overline{R}_\delta(\pi, \pi_0; \mathbf{v}, m_y), -\underline{R}_\delta(\pi, \pi_0; \mathbf{v}, m_y)]$ recovers bounds on the TNR ($y = 0$) and FNR ($y = 1$) regret.

A.0.3. δ-REGRET BOUNDS ON POSITIVE AND NEGATIVE PREDICTIVE VALUE

Lemma A.4. *Let $m_a(\pi) = p(Y(1) = a \mid A^\pi = a)$ be the positive $(a = 1)$ or negative $(a = 0)$ predictive value of π. Let $\rho_{td} = p(T = t, D = d)$, $\sigma(a) = (1 - 2a)(\rho_{10} - \rho_{01})$ and $\psi_a(\pi) = p(A^\pi = a)$ and assume that $(D, T) \sim p^*(\cdot)$ satisfies $p(D = a) \cdot p(T = a) > 0$. Then*

$$\delta_a(\mathbf{v}_0, \mathbf{v}_1) = \frac{\sigma(a) \cdot v_a(a, a) + \psi_a(\pi_0) \cdot v_a(a, a') - \psi_a(\pi) \cdot v_a(a', a)}{\psi_a(\pi) \cdot \psi_a(\pi_0)}.$$

Additionally, for all constrained uncertainty sets $\mathcal{V}(p; \tau)$, the upper δ_a-regret bound is given by

$$\overline{R}_\delta(\pi, \pi_0; m_{a=0}, \mathcal{V}) = \delta_{a=0}(\overline{\mathbf{v}}_0, \mathbf{v}_1), \quad \textit{where } \overline{\mathbf{v}}_0 = \begin{cases} \arg\max_{v_y(0,0)} \min_{v_y(1,0)} \mathcal{V}(p; \tau), & \sigma(0) \geq 0 \\ \arg\min_{v_y(0,0)} \min_{v_y(1,0)} \mathcal{V}(p; \tau), & \sigma(0) < 0 \end{cases},$$

$$\overline{R}_\delta(\pi, \pi_0; m_{a=1}, \mathcal{V}) = \delta_{a=1}(\overline{\mathbf{v}}_0, \mathbf{v}_1), \quad \textit{where } \overline{\mathbf{v}}_0 \in \arg\max_{v_1(1,0)} \mathcal{V}(p; \tau).$$

We now prove δ-regret bounds for m_a.

Proof. Let $a' = 1 - a$. Then $\forall a \in \{0, 1\}$ we have that

$$\begin{aligned}
\delta(\mathbf{v}_0^*, v_1) &= m_a^*(\pi) - m_a^*(\pi_0) \\
&= p(Y(1) = a \mid T^\pi = a) - p(Y(1) = a \mid D^{\pi_0} = a) \\
&= \frac{p(T = a, Y(1) = a)}{p(T = a)} - \frac{p(D = a, Y(1) = a)}{p(D = a)} \\
&= \frac{v_a(a, a') + v_a(a, a)}{\rho_{aa} + \rho_{aa'}} - \frac{v_a(a, a) + v_a(a', a)}{\rho_{aa} + \rho_{a'a}} \\
&= \frac{(\rho_{a'a} - \rho_{aa'}) \cdot v_a(a, a) + \psi_a(\pi_0) \cdot v_a(a, a') - \psi_a(\pi) \cdot v_a(a', a)}{\psi_0(\pi) \cdot \psi_0(\pi_0)}
\end{aligned}$$

When $a = 1$, we have that

$$R^*(\pi, \pi_0; m_{a=1}) \leq \frac{(\rho_{01} - \rho_{10}) \cdot v_1(1, 1) + \psi_1(\pi_0) \cdot \overline{v}_1(1, 0) - \psi_1(\pi) \cdot v_1(0, 1)}{\psi_1(\pi) \cdot \psi_1(\pi_0)}$$

When $a = 0$, we have two cases.

$$R^*(\pi, \pi_0; m_{a=0}) \leq \begin{cases} \dfrac{(\rho_{10} - \rho_{01}) \cdot \overline{v}_0(0, 0) + \psi_0(\pi_0) \cdot v_0(0, 1) - \psi_0(\pi) \cdot \underline{v}_0(1, 0)}{\psi_0(\pi) \cdot \psi_0(\pi_0)}, & \rho_{10} > \rho_{01} \\ \dfrac{(\rho_{10} - \rho_{01}) \cdot \underline{v}_0(0, 0) + \psi_0(\pi_0) \cdot v_0(0, 1) - \psi_0(\pi) \cdot \underline{v}_0(1, 0)}{\psi_0(\pi) \cdot \psi_0(\pi_0)}, & \rho_{10} \leq \rho_{01} \end{cases}$$

The lower regret bound is symmetric.

□

A.0.4. BASELINE REGRET BOUNDS

We now provide baseline regret bounds for policy performance measures.

Proposition A.5 (Baseline bounds on m_u)**.** *Let $m_u(\pi)$ be the expected utility of π. Then the baseline upper bound on regret is given by*

$$\overline{R}(\pi, \pi_0; m_u, \mathcal{V}) = \sum_y u_{0y} \cdot (\overline{v}_y(0,0) + v_y(0,1) - \underline{v}_y(0,0) - \underline{v}_y(1,0)) + u_{1y} \cdot (\overline{v}_y(1,0) - v_y(0,1)),$$

where the lower bound $\underline{R}(\pi, \pi_0; m_u, \mathcal{V})$ is symmetric.

Proof. By the same argument provided in the proof of Theorem A.1, we have that

$$m_u^*(\pi) = \sum_{t,y} u_{ty} \cdot (v_y(t,0) + v_y(t,1)), \quad m_u^*(\pi_0) = \sum_{d,y} u_{dy} \cdot (v_y(0,d) + v_y(1,d))$$

This implies

$$\overline{m}_u(\pi; \mathcal{V}) = \sum_y u_{0y} \cdot (\overline{v}_y(0,0) + v_y(0,1)) + u_{1y} \cdot (\overline{v}_y(1,0) + v_y(1,1))$$
$$\underline{m}_u(\pi_0; \mathcal{V}) = \sum_y u_{0y} \cdot (\underline{v}_y(0,0) + \underline{v}_y(1,0)) + u_{1y} \cdot (v_y(0,1) + v_y(1,1))$$

Simplifying yields the result

$$\begin{aligned}\overline{R}(\pi, \pi_0; m_u, \mathcal{V}) &= \overline{m}_u(\pi; \mathcal{V}) - \underline{m}_u(\pi_0; \mathcal{V}) \\ &= \sum_y u_{0y} \cdot (\overline{v}_y(0,0) + v_y(0,1) - \underline{v}_y(0,0) - \underline{v}_y(1,0)) + u_{1y} \cdot (\overline{v}_y(1,0) - v_y(0,1))\end{aligned}$$

□

Proposition A.6 (Baseline bounds on m_y)**.** *Let $m_y^*(\pi) = p(A^\pi = 1 \mid Y(1) = y)$ be the positive $(y = 1)$ or negative $(y = 0)$ class predictive performance of π. Then the baseline upper regret bound on m_y is given by*

$$\overline{R}(\pi, \pi_0; m_y, \mathcal{V}) = \left(\frac{\overline{v}_y(1,0) + v_y(1,1)}{\underline{v}_y(0,0) + v_y(0,1) + \overline{v}_y(1,0) + v_y(1,1)}\right) - \left(\frac{v_y(0,1) + v_y(1,1)}{\overline{v}_y(0,0) + v_y(0,1) + \overline{v}_y(1,0) + v_y(1,1)}\right)$$

where the lower bound $\underline{R}(\pi, \pi_0; m_y, \mathcal{V})$ is symmetric.

Proof. We have that

$$m_y^*(\pi) = p(T^\pi = 1 \mid Y(1) = y) = \frac{p(T^\pi = 1, Y(1) = y)}{p(Y(1) = y)} = \frac{v_y(1,0) + v_y(1,1)}{v_y(0,0) + v_y(0,1) + v_y(1,0) + v_y(1,1)}$$

And similarly,

$$m_y^*(\pi_0) = p(D^{\pi_0} = 1 \mid Y(1) = y) = \frac{p(D^{\pi_0} = 1, Y(1) = y)}{p(Y(1) = y)} = \frac{v_y(0,1) + v_y(1,1)}{v_y(0,0) + v_y(0,1) + v_y(1,0) + v_y(1,1)}$$

Applying the definition of baseline regret yields the result:

$$\begin{aligned}\overline{R}(\pi, \pi_0; \mathbf{v}, m_y) &= \overline{m}_y(\pi; \mathbf{v}) - \underline{m}_y(\pi_0; \mathbf{v}) \\ &= \left(\frac{\overline{v}_y(1,0) + v_y(1,1)}{\underline{v}_y(0,0) + v_y(0,1) + \overline{v}_y(1,0) + v_y(1,1)}\right) - \left(\frac{v_y(0,1) + v_y(1,1)}{\overline{v}_y(0,0) + v_y(0,1) + \overline{v}_y(1,0) + v_y(1,1)}\right)\end{aligned}$$

The lower bound follows from applying the same decomposition with $\underline{R}(\pi, \pi_0; \mathbf{v}, m_y) = \underline{m}_y(\pi; \mathbf{v}) - \overline{m}_y(\pi_0; \mathbf{v})$. □

Proposition A.7 (Baseline bounds on m_a)**.** *Let $m_a^*(\pi) = p(Y(1) = a \mid A^\pi = a)$ be the positive $(y = 1)$ or negative $(y = 0)$ predictive value of π. Let $\psi_a(\pi) = p(A^\pi = a)$. Then the baseline regret is upper bounded by*

$$\overline{R}(\pi, \pi_0; m_{a=1}, \mathcal{V}) = \frac{\overline{v}_1(1,0) + v_1(1,1)}{\psi_1(\pi)} - \frac{v_1(0,1) + v_1(1,1)}{\psi_1(\pi_0)}$$
$$\overline{R}(\pi, \pi_0; m_{a=0}, \mathcal{V}) = \frac{\overline{v}_0(0,0) + v_0(0,1)}{\psi_0(\pi)} - \frac{\underline{v}_0(0,0) + \underline{v}_0(1,0)}{\psi_0(\pi_0)}$$

where the lower bounds $\underline{R}(\pi, \pi_0; m_{a=0}, \mathcal{V})$, $\underline{R}(\pi, \pi_0; m_{a=1}, \mathcal{V})$ are symmetric.

Proof. We begin by showing $\overline{R}(\pi, \pi_0; m_{a=1}, \mathcal{V})$.

$$m_{a=1}^*(\pi) = p(Y(1) = 1 \mid T = 1) = \frac{p(Y(1) = 1, T = 1)}{p(T = 1)} = \frac{v_1(1,0) + v_1(1,1)}{\psi_1(\pi)}$$

$$m_{a=1}^*(\pi_0) = p(Y(1) = 1 \mid D = 1) = \frac{p(Y(1) = 1, D = 1)}{p(D = 1)} = \frac{v_1(0,1) + v_1(1,1)}{\psi_1(\pi_0)}$$

$$\begin{aligned}\overline{R}(\pi, \pi_0; m_{a=1}, \mathcal{V}) &= \overline{m}_{a=1}(\pi; \mathcal{V}) - \underline{m}_{a=1}(\pi_0; \mathcal{V}) \\ &= \frac{\overline{v}_1(1,0) + v_1(1,1)}{\psi_1(\pi)} - \frac{v_1(0,1) + v_1(1,1)}{\psi_1(\pi_0)}\end{aligned}$$

The lower bound $\underline{R}(\pi, \pi_0; m_{a=1}, \mathcal{V})$ is symmetric. Similarly, for the upper bound on the negative predictive value $\overline{R}(\pi, \pi_0; m_{a=0}, \mathcal{V})$,

$$m_{a=0}^*(\pi) = p(Y(1) = 0 \mid T = 0) = \frac{p(Y(1) = 0, T = 0)}{p(T = 0)} = \frac{v_0(0,0) + v_0(0,1)}{\psi_0(\pi)}$$

$$m_{a=0}^*(\pi_0) = p(Y(1) = 0 \mid D = 0) = \frac{p(Y(1) = 0, D = 0)}{p(D = 0)} = \frac{v_0(0,0) + v_0(1,0)}{\psi_0(\pi_0)}$$

$$\begin{aligned}\overline{R}(\pi, \pi_0; m_{a=0}, \mathcal{V}) &= \overline{m}_{a=0}(\pi; \mathcal{V}) - \underline{m}_{a=0}(\pi_0; \mathcal{V}) \\ &= \frac{\overline{v}_0(0,0) + v_0(0,1)}{\psi_0(\pi)} - \frac{\underline{v}_0(0,0) + \underline{v}_0(1,0)}{\psi_0(\pi_0)}\end{aligned}$$

The lower bound $\underline{R}(\pi, \pi_0; m_{a=0}, \mathcal{V})$ is symmetric.

□

A.0.5. PROOF OF THEOREM 4.3

Proof. We will bound $\overline{\Delta}(m, \mathcal{V}) = \overline{R}(\pi, \pi_0; m, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m, \mathcal{V})$. The result over the full interval follows by taking $\Delta(m, \mathcal{V}) = 2 \cdot \overline{\Delta}(m, \mathcal{V})$ because $\overline{R}(\pi, \pi_0; m, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m, \mathcal{V}) = \underline{R}_\delta(\pi, \pi_0; m, \mathcal{V}) - \underline{R}(\pi, \pi_0; m, \mathcal{V})$. We begin for showing the result for the positive and negative class predictive performance.

Case 1: $\overline{v}_{10} > v_{01}$.

$$\begin{aligned}
\overline{\Delta}(m_y, \mathcal{V}) &= \overline{R}(\pi, \pi_0; m_y, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m_y, \mathcal{V}) \\
&= \left(\frac{v_{11} + \overline{v}_{10}}{\underline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}} - \frac{v_{11} + v_{01}}{\overline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}} \right) - \frac{\overline{v}_{10} - v_{01}}{\underline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}} \\
&= \frac{(v_{11} + v_{01}) \cdot (\overline{v}_{00} - \underline{v}_{00})}{(\underline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}) \cdot (\overline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11})} \\
&\geq \frac{\alpha \cdot v_{11}}{(\overline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}) \cdot (\overline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11})} \\
&= \frac{\alpha \cdot v_{11}}{(\overline{\gamma})^2}
\end{aligned}$$

Case 2: $\overline{v}_{10} \leq v_{01}$.

$$\begin{aligned}
\overline{\Delta}(m_y, \mathcal{V}) &= \overline{R}(\pi, \pi_0; m_y, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m_y, \mathcal{V}) \\
&= \left(\frac{v_{11} + \overline{v}_{10}}{\underline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}} - \frac{v_{11} + v_{01}}{\overline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}} \right) - \frac{\overline{v}_{10} - v_{01}}{\overline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}} \\
&= \frac{(v_{11} + \overline{v}_{10}) \cdot (\overline{v}_{00} - \underline{v}_{00})}{(\underline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11}) \cdot (\overline{v}_{00} + \overline{v}_{10} + v_{01} + v_{11})} \\
&\geq \frac{\alpha \cdot v_{11}}{(\overline{\gamma})^2}
\end{aligned}$$

Next we will bound $\overline{\Delta}(m_{a=0}, \mathcal{V})$. Case 1: $\rho_{10} - \rho_{11} > 0$. Note that $\psi_0(\pi) = \rho_{01} + \rho_{00}$ and $\psi_0(\pi_0) = \rho_{10} + \rho_{00}$. We have that

$$\begin{aligned}
\overline{\Delta}(m_{a=0}, \mathcal{V}) &= \overline{R}(\pi, \pi_0; m_{a=0}, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m_{a=0}, \mathcal{V}) \\
&= \left(\frac{\overline{v}_0(0,0) + v_0(0,1)}{\rho_{01} + \rho_{00}} - \frac{\underline{v}_0(0,0) + \underline{v}_0(1,0)}{\rho_{10} + \rho_{00}} \right) \\
&\quad - \frac{(\rho_{10} - \rho_{01}) \cdot \overline{v}_0(0,0) + (\rho_{10} + \rho_{00}) \cdot v_0(0,1) - (\rho_{01} + \rho_{00}) \cdot \underline{v}_0(1,0)}{(\rho_{01} + \rho_{00}) \cdot (\rho_{01} + \rho_{00})} \\
&= \frac{(\overline{v}_0(0,0) - \underline{v}_0(0,0))}{(\rho_{01} + \rho_{00})} \\
&= \frac{\alpha}{\psi_0(\pi)}.
\end{aligned}$$

The third equality follows from finding a common denominator and simplifying. Case 2: $\rho_{10} - \rho_{11} \leq 0$. Following the same argument, we have that $\overline{\Delta}(m_{a=0}, \mathcal{V}) = \frac{\alpha}{\psi_0(\pi_0)}$. Thus

$$\overline{\Delta}(m_{a=0}, \mathcal{V}) \geq \min\{\frac{\alpha}{\psi_0(\pi)}, \frac{\alpha}{\psi_0(\pi_0)}\} = \frac{\alpha}{\max\{\psi_0(\pi), \psi_0(\pi_0)\}}.$$

Next we will bound $\overline{\Delta}(m_u, \mathcal{V})$. Let $\lambda_{ay} = u_{ay} - u_{a'y}$.

Case 1: $\lambda_{11} > \lambda_{10}$.

$$\begin{aligned}
\overline{\Delta}(m_u, \mathcal{V}) &= \overline{R}(\pi, \pi_0; m_u, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m_u, \mathcal{V}) \\
&= (u_{00} + u_{01})(\overline{v}_0(0,0) - \underline{v}_0(0,0)) + (u_{00} + u_{11})(\overline{v}_0(1,0) - \underline{v}_0(1,0))
\end{aligned}$$

Case 2: $\lambda_{11} \leq \lambda_{10}$.

$$\begin{aligned}\overline{\Delta}(m_u, \mathcal{V}) &= \overline{R}(\pi, \pi_0; m_u, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m_u, \mathcal{V}) \\ &= (u_{00} + u_{01})(\overline{v}_0(0,0) - \underline{v}_0(0,0)) + (u_{10} + u_{01})(\overline{v}_0(1,0) - \underline{v}_0(1,0))\end{aligned}$$

Combining cases yields the result:

$$\overline{\Delta}(m_u, \mathcal{V}) \leq (u_{00} + u_{01}) \cdot (\overline{v}_0(0,0) - \underline{v}_0(0,0))$$

The result that $\overline{\Delta}(m_{a=1}, \mathcal{V}) = \overline{R}(\pi, \pi_0; m_{a=1}, \mathcal{V}) - \overline{R}_\delta(\pi, \pi_0; m_{a=1}, \mathcal{V}) = 0$ follows directly from plugging in definitions of baseline upper regret bound and δ regret bound and simplifying.

□

B. Assumption Mapping Extensions and Proofs

In this appendix, we discuss additional causal assumptions which imply uncertainty sets over partially-identified v-statistics. Rosenbaum's Γ-sensitivity model (Rosenbaum, 2005), the proximal identification framework (Ghassami et al., 2023), and Manski-style no assumptions bounds (Manski & Pepper, 2000) imply bounding functions $\tau(\cdot)$, which can be used to construct $\mathcal{V}(p;\tau)$ by invoking Lemma 5.2.

B.1. Rosenbaum's Γ-Sensitivity Model

Rosenbaum's Γ-*sensitivity analysis model* bounds the influence of unobserved confounders on the odds of being treated versus untreated (Rosenbaum, 1987). Namkoong et al. (2020) leverage a sequential adaptation of this model to partially identify the value function of a new policy π given confounded off-policy data, while Zhang et al. (2020) leverage this model to rank individualized treatment rules under confounding.

Assumption B.1 (Γ-sensitivity)**.** For some $\Gamma \geq 1$, $(D, T, X, U, Y(1)) \sim p^*(\cdot)$ satisfies

$$\Gamma^{-1} \leq \frac{P(D=1 \mid X, U=u)}{P(D=0 \mid X, U=u)} \frac{P(D=0 \mid X, U=\tilde{u})}{P(D=1 \mid X, U=\tilde{u})} \leq \Gamma \tag{3}$$

for all $u, \tilde{u} \in \mathcal{U}$ and $X \in \mathcal{X}$ with probability one.

Lemma B.2 (Rambachan et al. (2022))**.** *Suppose that* $(D, T, X, U, Y(1)) \sim p^*(\cdot)$ *satisfies Assumption B.1 for some* $\Gamma \geq 1$. *Then*

$$\Gamma^{-1} \cdot \mu_1(x) \leq \mu_0(x) \leq \Gamma \cdot \mu_1(x), \quad \forall x \in X.$$

This result follows from Proposition 8.2 of Rambachan et al. (2022), which shows that Rosenbaum's Γ-sensitivity model implies a marginal sensitivity model in binary outcome settings. Because Rosenbaum's Γ sensitivity model does not imply *sharp* bounds under the MSM, an uncertainty set constructed using $\underline{\tau}(x) = \Gamma^{-1} \cdot \mu_1(x)$, $\overline{\tau}(x) = \Gamma \cdot \mu_1(x)$ does not guarantee we recover the tightest regret interval attainable under Assumption B.1.

B.2. Marginal Sensitivity Model

The marginal sensitivity model (MSM) restricts the extent to which unobserved confounders impact the odds of treatment under the status quo policy (Tan, 2006). In particular, this model holds that a confounder $U \in \mathbb{R}^k$ exists such that decisions would be conditionally randomized if we controlled for both X and U. As pointed out by Robins (2002), it suffices to assign $U = Y(1)$.

Assumption B.3 (Marginal Sensitivity Model)**.** For some $\Lambda \geq 1$, $(X, D, Y(1)) \sim p^*(\cdot)$ satisfies

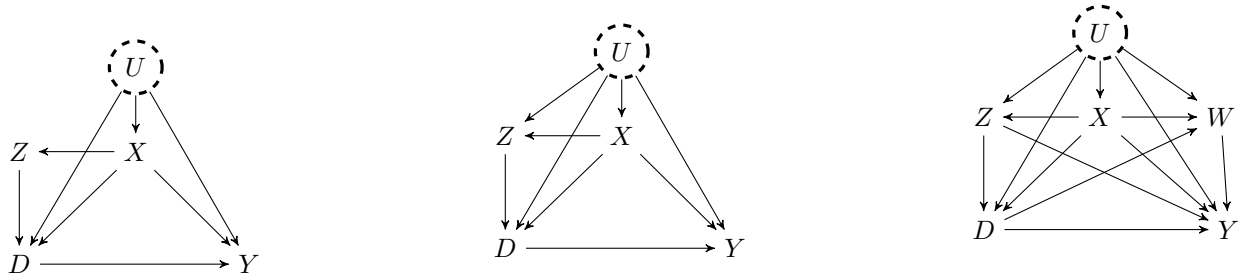
$$\Lambda^{-1} \leq \frac{p(D=1 \mid X, Y(1))}{p(D=0 \mid X, Y(1))} \cdot \frac{p(D=1 \mid X)}{p(D=0 \mid X)} \leq \Lambda \tag{4}$$

As in the IV setting, Assumption B.3 implies bounds on the unobserved regression $\mu_0(x)$, which in turn can be used to construct $\tau(\cdot)$.

Lemma B.4 (Rambachan et al. (2022))**.** *Suppose Assumptions 3.1 and 3.2 hold and that the MSM (B.3) is satisfied for some* $\Lambda \geq 1$. *Then*

$$\Lambda^{-1} \cdot \mu_1(x) \leq \mu_0(x) \leq \Lambda \cdot \mu_1(x), \ \forall x \in X.$$

Lemma B.4 follows from Bayes' rule, and is a direct consequence of Proposition 8.1 in (Rambachan et al., 2022). Moreover, because Rosenbaum's Γ model implies bounds under the MSM (Rambachan et al., 2022), we can similarly construct bounding functions under this model by taking $\underline{\tau}(x) = \Gamma^{-1} \cdot \mu_1(x)$, $\overline{\tau}(x) = \Gamma \cdot \mu_1(x)$.



(a) A DAG containing an instrumental variable Z.

(b) A DAG containing a treatment confounding proxy Z.

(c) A DAG containing a treatment confounding proxy Z and outcome confounding proxy W.

Figure 8. Three sets of structural assumptions on $p^*(\cdot)$ which imply bounding functions $\tau(\cdot)$.

B.3. Instrumental Variable

Instrumental variables can be used for identification when the status quo policy was influenced by a random source of selection rate heterogeneity (Lakkaraju et al., 2017; Chen et al., 2023; Rambachan et al., 2021; Kleinberg et al., 2018). For example, in medical testing contexts, an instrument is available when patients are randomly assigned to physicians with heterogeneous testing rates (Abaluck et al., 2016).

Assumption B.5 (Instrumental Variable)**.** Let Z be a finite-valued instrument satisfying

1. Relevance: $Z \not\perp\!\!\!\perp D \mid X$
2. IV independence: $Z \perp\!\!\!\perp U$
3. Exclusion restriction: $Z \perp\!\!\!\perp Y \mid D, X, U$

In the following Lemma, we show that Assumption B.5 implies sharp bounds on the unobserved outcome regression.

Lemma B.6 (Rambachan et al. (2022; 2021))**.** *Suppose Assumptions 3.1, 3.2, and B.5 hold. Let $\mu(x) = \mathbb{E}[Y(1) \mid X = x]$ and $\mu_d(x) = \mathbb{E}[Y(1) \mid D = d, X = x]$ be outcome regressions and let $e_d(x) = p(D = d \mid X = x)$ be the propensity function. Then $\forall x \in X$, $\mu_0(x)$ is bounded by*

$$\frac{\underline{\mu}(x) - \mu_1(x) \cdot e_1(x)}{e_0(x)} \leq \mu_0(x) \leq \frac{\overline{\mu}(x) - \mu_1(x) \cdot e_1(x)}{e_0(x)},$$

where $\underline{\mu}(x) = \max_{\tilde{z} \in \mathcal{Z}} \{\mu_1(x, \tilde{z}) \cdot e_1(x, \tilde{z})\}$, $\bar{\mu}(x) = \min_{\tilde{z} \in \mathcal{Z}} \{e_0(x, \tilde{z}) + \mu_1(x, \tilde{z}) \cdot e_1(x, \tilde{z})\}$.

Proof. By iterated expectations, $\mu(x) = \mu_1(x) \cdot e_1(x) + \mu_0(x) \cdot e_0(x)$. Therefore

$$\mu_1(x) \cdot e_1(x) \leq \mu(x) \leq e_0(x) + \mu_1(x) \cdot e_1(x) \tag{5}$$

$$\mu_1(x, z) \cdot e_1(x, z) \leq \mu(x) \leq e_0(x, z) + \mu_1(x, z) \cdot e_1(x, z) \tag{6}$$

$$\max_{\tilde{z} \in \mathcal{Z}} \{\mu_1(x, \tilde{z}) \cdot e_1(x, \tilde{z})\} \leq \mu(x) \leq \min_{\tilde{z} \in \mathcal{Z}} \{e_0(x, \tilde{z}) + \mu_1(x, \tilde{z}) \cdot e_1(x, \tilde{z})\} \tag{7}$$

where (5) implies sharp bounds on $\mu_0(x)$ (Manski & Pepper (2000)) and (6) follows from Assumption B.5. Solving for $\mu_0(x)$ yields

$$\underline{\mu}_0(x) = \frac{\underline{\mu}(x) - \mu_1(x) \cdot e_1(x)}{e_0(x)}, \quad \overline{\mu}_0(x) = \frac{\overline{\mu}(x) - \mu_1(x) \cdot e_1(x)}{e_0(x)}$$

with $\underline{\mu}(x), \overline{\mu}(x)$ defined as in (7). Note that both terms are defined because $e_0(x) > 0, \forall x \in X$ by Assumption 3.2. □

B.4. Proximal Variable

The proximal causal inference framework relaxes the IV unconfoundedness condition imposed by the IV framework (Tchetgen et al., 2020). Ghassami et al. (2023) extend this framework to support partial-identification of the average treatment effect. We show that this framework also implies point-wise bounding functions on the unobserved outcome regression. We discuss two versions of this framework – the treatment confounded proxy (Fig 8.b) and treatment/outcome confounded proxy (Fig 8.c) – which are most realistic in our setting.

Assumption B.7. (Treatment confounding proxy) Let Z be a treatment confounding proxy variable such that $(Z, Y, D, X, U) \sim p^*(\cdot)$ satisfies $Z \perp\!\!\!\perp Y \mid D, X, U$.

Assumption B.7 is identical to the exclusion restriction in the IV setting, but allows for a relaxation of IV independence. Therefore, this condition can be reasonable in settings where IV independence is violated (e.g., confounded assignment of instances to decision-makers).

Assumption B.8 (Treatment confounding bridge)**.** There exists a non-negative bridge function m such that almost surely

$$\mathbb{E}[m(Z, D, X) \mid D, X, U] = \frac{p(U \mid 1 - D, X)}{p(U \mid D, X)}.$$

Note that this condition only stipulates that m exists and does not require its point identification.

Lemma B.9. *(Ghassami et al. (2023)) Let Assumptions 3.2, B.7 and B.8 hold. Let* $\mu_d(x, z) = \mathbb{E}[Y(1) \mid D = d, Z = z, X = x]$. *Then* $\forall x \in X$

$$\min_{\tilde{z} \in Z}\{\mu_1(x, \tilde{z})\} \leq \mu_0(x) \leq \max_{\tilde{z} \in Z}\{\mu_1(x, \tilde{z})\} \tag{8}$$

Assumption B.10. (Treatment and outcome confounding proxy) Let Z be a treatment confounding proxy and W be an outcome confounding proxy such that $(Z, W, Y, D, X, U) \sim p^*(\cdot)$ satisfies $Z \perp\!\!\!\perp W \mid D, X, U$.

Assumption B.10 enables a relaxation of the IV unconfoundedness condition. This assumption also obviates the exclusion restriction required by the IV (B.5) and treatment confounding proxy (B.7) assumptions. Therefore, Assumption B.10 applies given a confounded treatment assignment variable (e.g., the identity of a decision-maker Z) and a temporally-lagged outcome variable W. Temporally-lagged outcomes are ubiquitous in multi-step decision-making processes, where the final outcome of interest is often preceded by a series of intermediary decisions.

Assumption B.11. (Outcome confounding bridge) There exists a non-negative bridge function h such that almost surely $\mathbb{E}[Y \mid D, X, U] = \mathbb{E}[h(W, D, X) \mid D, X, U]$

Intuitively, this assumption requires that W is sufficiently informative of confounders such that there exists a function of W which recovers the potential outcome mean as a function of U (Miao et al., 2018; Ghassami et al., 2023). Similarly to Assumption B.8, this condition does not require identification of h.

Lemma B.12 (Ghassami et al. (2023))**.** *Let Assumptions 3.2, B.8, B.10, and B.11 hold. Let* $\mu_1(x) = \mathbb{E}[Y(1) \mid D = 1, X = x]$, $\eta_1(w) = p(W = w \mid D = 1, X = x)$, $\eta_1(z) = p(Z = z \mid D = 1, X = x)$, $\eta_1(w, z) = p(W = w, Z = z \mid D = 1, X = x)$. *Then*

$$\mu_1(x) \cdot \min_{w,z} \frac{\eta_1(w, z)}{\eta_1(w) \cdot \eta_1(z)} \leq \mu_0(x) \leq \mu_1(x) \cdot \max_{w,z} \frac{\eta_1(w, z)}{\eta_1(w) \cdot \eta_1(z)}, \quad \forall x \in X.$$

B.5. Manski Style No-Assumptions Bounds

Manski-style no assumptions bounds on partially identified policy comparison terms (Manski, 1989; Manski & Pepper, 2000) follow by setting $\underline{\tau}(x) = 0, \overline{\tau}(x) = 1$. Invoking Lemma 5.2 yields $\mathcal{H}(v_y(t, 0); \tau) = [0, p(D = 0, T = 1)]$, which is the same interval recovered by the *unconstrained* uncertainty set $\mathcal{V}(p; \tau)$.

B.6. Assumption Mapping Proofs

B.6.1. Proof of Lemma 5.2.

Proof. For the upper bound $\overline{\tau}(x)$ we have that $\forall x \in X$

$$\begin{aligned} \mathbb{E}[Y(1) \mid D=0, X=x] &\leq \overline{\tau}(x) \\ \mathbb{E}[Y(1) \mid D=0, T=t, X=x] &\leq \overline{\tau}(x) \\ \mathbb{E}[Y(1) \mid D=0, T=t, X=x] \cdot p(D=0, T=t, X=x) &\leq p(D=0, T=t, X=x) \cdot \overline{\tau}(x) \end{aligned} \tag{9}$$

where (9) holds because $T \perp\!\!\!\perp \{D, Y(1)\} \mid X$. Simplifying the right hand side yields

$$\begin{aligned} &\mathbb{E}[Y(1) \mid D=0, T=t, X=x] \cdot p(D=0, T=t, X=x) \\ &\leq p(D=0 \mid T=t, X=x) \cdot p(T=t \mid X=x) \cdot p(X=x) \cdot \overline{\tau}(x) \\ &\leq p(D=0 \mid X=x) \cdot p(T=t \mid X=x) \cdot p(X=x) \cdot \overline{\tau}(x) \end{aligned}$$

The result follows from marginalizing over X

$$\begin{aligned} v_1(t,0) &\leq \sum_x e_0(x) \cdot \pi_t(x) \cdot \overline{\tau}(x) \cdot p(X=x) \\ &= \mathbb{E}[\overline{\tau}(x) \cdot e_0(x) \cdot \pi_t(x)] \end{aligned}$$

The upper bound is sharp because equality holds when $\overline{\tau}(x) = \mathbb{E}[Y(1) \mid D=0, X=x]\ \forall x \in X$. The lower bound on $v_t(t,0)$ follows by the same argument. □

B.6.2. Proof of Theorem 5.3.

Importantly, Theorem 5.3 does *not* imply that $\mathcal{V}(p;\tau)$ guarantees the tightest regret interval attainable from a specific causal assumption. In some cases, it may be possible to contract regret intervals further by exploiting additional information provided by a causal assumption. We discuss an example involving Rosenbaum's Γ model above. Nevertheless, the flexibility of our assumption mapping approach is important for the real-world applicability of our framework.

Proof. $\mathcal{V}(p;\tau)$ is minimal if (1) it is consistent with $p(X, D, T, Y)$ and $\underline{\tau}, \overline{\tau}$ and (2) no consistent set $\mathcal{V}^*(p;\tau)$ exists such that $\lambda(\mathcal{V}^*(p;\tau)) < \lambda(\mathcal{V}(p;\tau))$. The first condition holds if $\forall \mathbf{v}_0 \in \mathcal{V}(p;\tau)$,

$$\begin{aligned} v_1(0,0) &\in [\mathbb{E}[\underline{\tau}(x) \cdot e_0(x) \cdot \pi_0(x)], \mathbb{E}[\overline{\tau}(x) \cdot e_0(x) \cdot \pi_0(x)]] \\ v_1(1,0) &\in [\mathbb{E}[\underline{\tau}(x) \cdot e_0(x) \cdot \pi_1(x)], \mathbb{E}[\overline{\tau}(x) \cdot e_0(x) \cdot \pi_1(x)]] \end{aligned}$$

and $v_0(t,d) + v_1(t,d) = \rho_{td}, \forall t, d$. This follows by definition of $\mathcal{V}(p;\tau)$. Additionally,

$$\sum_{d,t} p(D=d, T=t) = 1 \implies \sum_{d,t} v_0(d,t) + v_1(d,t) = 1 \implies \sum_{d,t,y} v_y(d,t) = 1$$

where the first equality follows by the law of total probability and the first implication follows by the constraint. Thus

$$\sum_{d,y} v_y(t,d) = p(T=t),\ \sum_{t,y} v_y(t,d) = p(D=d),\ p(Y=y) = \sum_{t,y} v_y(t,1) \leq \sum_{d,t,d} v_y(t,d) \leq 1.$$

Therefore, the constraint $v_0(t,d) + v_1(t,d) = 1\ \forall t, d$ implies that $\mathbf{v}_0$ is consistent with other marginals of $p(X, D, T, Y)$. We will show (2) by contradiction. The Lebesgue measure over a cartesian product of intervals is given by $\lambda(\mathcal{V}) = \prod_{y,t} |\overline{v}_y(t,0) - \underline{v}_y(t,0)|$. Let $\mathcal{V}^*(p;f) \in [0,1]^4$ be an arbitrary set satisfying $\lambda(\mathcal{V}) > \lambda(\mathcal{V}^*)$. By a property of Lebesgue measures it follows that $A \supseteq B \implies \lambda(A) \leq \lambda(B)$. Thus

$$\begin{aligned} \lambda(\mathcal{V}) > \lambda(\mathcal{V}^*) &\implies \mathcal{V} \not\subseteq \mathcal{V}^* \\ &\implies \exists\ \mathbf{v}_0 \in \mathcal{V} \text{ s.t. } \mathbf{v}_0 \notin \mathcal{V}^* \\ &\implies \exists\ t \in \{0,1\} \text{ s.t. } \overline{v}_1(t,0) > \overline{v}_1^*(t,0) \text{ or } \underline{v}_1^*(t,0) > \underline{v}_1(t,0). \end{aligned}$$

Define $s(x) = p(x) \cdot e_0(x) \cdot \pi_t(x)$. Then

$$\begin{aligned}
\overline{v}_1(t,0) > \overline{v}_1^*(t,0) &\implies \sum_x s(x) \cdot \overline{\tau}(x) > \sum_x s(x) \cdot \overline{\tau}^*(x) \\
&\implies \sum_x s(x) \cdot \overline{\tau}(x) - \sum_x s(x) \cdot \overline{\tau}^*(x) > 0 \\
&\implies \sum_x s(x)(\overline{\tau}(x) - \overline{\tau}^*(x)) > 0 \\
&\implies \exists\, x \in X \text{ s.t. } \overline{\tau}^*(x) < \overline{\tau}(x) \\
&\implies \exists\, x \in X \text{ s.t. } \overline{\tau}^*(x) \leq \mathbb{E}[Y(1) \mid D = 0, X = x] \leq \overline{\tau}(x)
\end{aligned}$$

However, the final implication violates the condition imposed by the bounding function. An analogous violation occurs for the lower bound $\underline{v}_1^*(t,0) > \underline{v}_1(t,0)$, proving the contradiction.

□

C. Theoretical Estimation Results

C.0.1. PROOF OF THEOREM 6.1

Proof. Let $P_n(f)$ denote sample averages $\frac{1}{n}\sum_{i=1}^{n} f(O_i)$ on a separate fold of the data from that used to estimate $\hat{f}$. We can decompose the error term into

$$\hat{v} - v = \underbrace{(P_n - P)\, f}_{Z^*} + \underbrace{(P_n - P)\,(\widehat{f} - f)}_{T_1} + \underbrace{P(\widehat{f} - f)}_{T_2}.$$

$Z^* = O_p(1/\sqrt{n})$ by the central limit theorem. $T_1 = o_p(1/\sqrt{n})$ by consistency of $\hat{f}$ in the L_2 norm and sample splitting (Kennedy et al., 2020). Finally, we have for T_2 that

$$\begin{aligned}
P(\hat{f} - f) &= P\left(\hat{e}(X)\cdot\hat{\tau}(X) - e(X)\cdot\tau(X)\right)\\
&= P\left(\hat{e}(X)\cdot\hat{\tau}(X) + \hat{e}(X)\cdot\tau(X) - \hat{e}(X)\cdot\tau(X) - e(X)\cdot\tau(X)\right)\\
&= P\left(\hat{e}(X)(\hat{\tau}(X) - \tau(X)) + \tau(X)\cdot(\hat{e}(X) - e(X))\right)\\
&= P\left(\hat{e}(X)(\hat{\tau}(X) - \tau(X))\right) + P(\tau(X)\cdot(\hat{e}(X) - e(X)))\\
&\leq ||\hat{e}(x) - e(x)|| + ||\hat{\tau}(x) - \tau(x)||
\end{aligned}$$

where the first equality follows by adding and subtracting $\hat{e}(X)\cdot\tau(X)$ and the final inequality follows by Cauchy-Schwarz. The overall convergence rate follows by combining terms.

□

C.1. Rate of doubly robust estimator

Theorem C.1. *The doubly-robust estimator satisfies*

$$\hat{\bar{v}}_{DR} - \bar{v} = O_P\Big(\,\|e(x) - \hat{e}(x)\|\,\|\mu_1(x) - \hat{\mu}_1(x)\|\,\Big) + (P_n - P)\phi(O;\eta) + o_p(\frac{1}{\sqrt{n}})$$

when $\left\|\phi - \hat{\phi}\right\| = o_p(1)$.

This theorem demonstrates that the error of our estimator is a product of nuisance function errors. This enables us to achieve faster rate of convergence even when estimating the nuisance function at slow rates. For example, to obtain $n^{-1/2}$ rates for our estimator, it is sufficient to estimate the nuisance functions at $n^{-1/4}$, allowing us to use flexible machine learning methods to nonparametrically estimate the nuisance functions under smoothness or sparsity assumptions. The theorem's condition that ϕ converges in probability in the $L_2(P)$ norm is mild and can be satisfied by using flexible regression methods.

Proof. Let η indicate the nuisance functions (e, π, μ_1) and let $\hat{\eta}$ indicate $(\hat{e}, \pi, \hat{\mu}_1)$.

$$P_n(\phi(O;\hat{\eta})) - \bar{v} = \overbrace{P(\phi(O;\hat{\eta}) - \bar{v})}^{A} + \overbrace{(P_n - P)(\phi(O;\hat{\eta}) - \phi(O;\eta)}^{B} + \overbrace{(P - P_n)(-\phi(O;\eta))}^{C}$$

For term A, we have that

$$\begin{aligned}
P(\phi(O;\hat{\eta}) - \bar{v}) = \int &\bar{\mu}(x)\Big(e(x) - \bar{e}(x)\Big)\Big(\bar{\pi}(x) - \pi(x)\Big) + \bar{e}(x)\Big(\pi(x) - \bar{\pi}(x)\Big)\Big(\bar{\mu}(x) - \mu(x)\Big)\\
&+ \Big(\mu(x) - \bar{\mu}(x)\Big)\Big(e(x) - \bar{e}(x)\Big)\Big(\bar{\pi}(x) - \pi(x)\Big))dP\\
&= O_P\Big(\,\|e(x) - \hat{e}(x)\|\,\|\mu(x) - \hat{\mu}(x)\|\,\Big)
\end{aligned}$$

where the second line applies Cauchy-Schwarz.

For term B, since P_n is the empirical measure on an independent sample from $\hat{P}$, we can apply Lemma 2 of Kennedy et al. (2020) with our assumption that $\left\|\phi - \hat{\phi}\right\| = o_P(1)$:

$$(P_n - P)(\phi(O;\hat{\eta}) - \phi(O;\eta) = O_P\Big(\frac{\|\phi(O;\hat{\eta}) - \phi(O;\eta)\|}{\sqrt{n}}\Big) = o_P(\frac{1}{\sqrt{n}})$$

Combining yields the result. □

D. Experiment Setup Details and Further Results

D.0.1. SYNTHETIC SETUP DETAILS

We sample data from the probability functions

$$\begin{aligned}
\gamma(X_i) &:= \sigma'(X_i \times W_z), \pi(X_i) := \sigma(X_i \times W_\pi) \\
\pi_0(V_i, Z_i) &:= \sigma(V_i \times W_{\pi_0} + \beta_0 \cdot Z_i) \\
\mu_1(V_i, Z_i) &:= \sigma(V_i \times W_{\mu_1} + \beta_1 \cdot Z_i) \\
\mu_0(V_i, Z_i) &:= \begin{cases} \Lambda^* \cdot \mu_1(V_i), \Lambda^* \in U(\Lambda^{-1}, \Lambda) & \text{(MSM)} \\ \sigma(V_i \times W_{\mu_0} + \beta_1 \cdot Z_i) & \text{(IV)} \end{cases} \\
\mu(V_i, Z_i) &:= \mu_1(V_i, Z_i)\pi_1(V_i, Z_i) + \mu_0(V_i, Z_i)\pi_0(V_i, Z_i)
\end{aligned}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ and $\sigma'(x)$ is the softmax function. We then sample $Z_i \sim$ Multinomial$(\gamma(X_i))$, $D_i^{\pi_0} \sim$ Bern$(\pi_0(V_i, Z_i))$, $T_i^{\pi} \sim$ Bern$(\pi(X_i))$, and outcomes

$$Y_i(1) \sim \begin{cases} \text{Bern}(\mu_1(V_i, Z_i)), & D_i^{\pi_0} = 1 \\ \text{Bern}(\mu_0(V_i, Z_i)), & D_i^{\pi_0} = 0 \end{cases}. \tag{10}$$

Experimental Trials. We randomly sample coefficient vectors W_z, W_{π_0}, W_{μ_1} and W_{μ_0} parameterizing π_0 and π in each of the experiments reported in Section 7. As a result, the oracle regret between the status quo and updated policy varies across experimental trials depending on the sigmoidal outcome probabilities induced by our randomly sampled coefficient vectors.

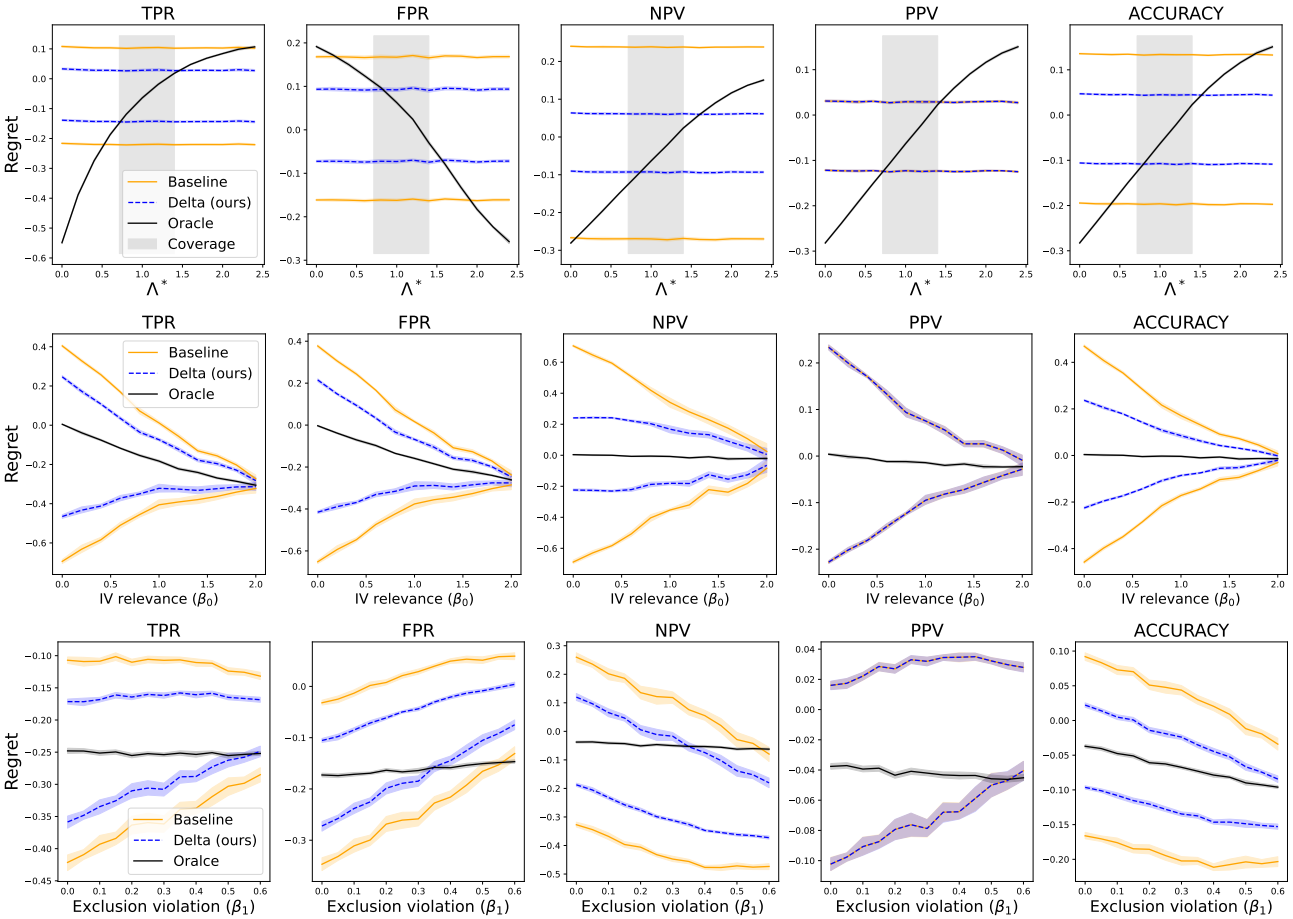


Figure 9. Top row (MSM): we fix $\Lambda = 1.4$ and vary Λ^*. The area shown in grey indicates the values of Λ^* for which the MSM assumption is satisfied. Middle row (IV): we vary β_0 controlling the relevance of Z on the status quo decision-making policy. Bottom row (IV): we vary β_1 controlling the magnitude of the exclusion restriction violation. Shaded error region shows a 95% bootstrap CI over 20 trials.

D.0.2. Assumption robustness experiments

We now leverage the same synthetic setup outlined in Section 7 to stress-test the coverage of our regret intervals to assumption violations. Observe that the IV unconfoundedness condition is satisfied because $\gamma(X_i)$ does not depend on U_i, the relevance condition is satisfied when $\beta_0 > 0$, and the exclusion restriction is satisfied when $\beta_1 = 0$. We test robustness to violation of the IV relevance and exclusion assumptions by varying β_0, β_1, respectively. We test robustness to violation of the MSM by selecting values for $\Lambda^* \notin [\Lambda^{-1}, \Lambda]$.

In Figure 9, we provide bounds for policy performance measures as we introduce violations to the MSM and IV assumptions. The top row shows regions in which the MSM assumption $\Lambda = 1.4$ is satisfied. The $\hat{R}_\delta$ interval yields valid coverage when $\Lambda^* \in [\Lambda^{-1}, \Lambda]$, but breaks when the assumption is violated. Observe that the relevance condition (middle row) controls the tightness of the regret interval, but does not introduce coverage violations. As a result, more heterogeneity across finite values of the instrument yields tighter uncertainty quantification, but does not impact coverage. This is inline with prior empirical evaluations conducted under the IV framework (Lakkaraju et al., 2017; Kleinberg et al., 2018). The bottom row shows that violation of the exclusion violation does introduce coverage violations as β_1 increases. Across settings, observe the the baseline and delta regret intervals overlap for the PPV, which is inline with the results of our theoretical analysis.

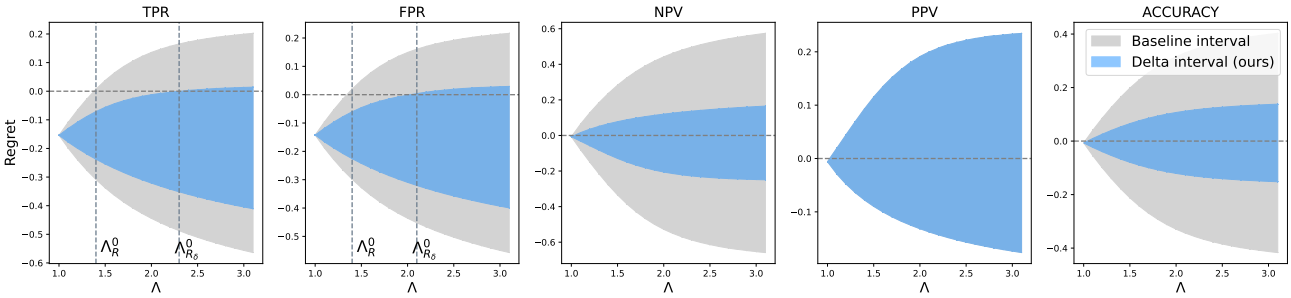


Figure 10. Design sensitivity analysis. Our δ-regret interval certifies a policy performance difference up to a larger magnitude of unmeasured confounding Λ^0 for the TPR and FPR measures.

D.0.3. Design sensitivity analysis

One common measure of robustness to confounding is the *design sensitivity*, or the value of the sensitivity parameter at which the analysis exceeds a key threshold of interest (Rosenbaum, 1987; 2005; Namkoong et al., 2020). In our setting, we are most interested in the degree of unmeasured confounding permissible before the regret interval crosses zero. We conduct a design sensitivity analysis under the MSM by varying Λ and measuring the value at which the δ-regret and baseline intervals include zero. We denote these thresholds as $\Lambda^0_{R_\delta}$ and Λ^0_R, respectively. We leverage the same synthetic setup reported above and fit $\hat{\eta}$ via parametric logistic regression models to improve computation time. Figure 10 indicates that the improvement in our δ interval enables certifying a policy difference up to a larger magnitude of unmeasured confounding than would be possible via the baseline approach. For example, the δ-interval certifies a TPR performance difference up to $\Lambda = 2.3$, a significant improvement over $\Lambda = 1.4$ yielded by the baseline interval.

D.0.4. Real-world data experiments

Setup Details. Each record of the synthetic dataset provided by (Obermeyer et al., 2019) contains patient demographics, health information measured in the prior year, and outcome variables measured in the following year. Only $\approx 1\%$ of patients were enrolled in the care management program under the status quo policy. This low selection rate yields vacuous regret intervals without strong assumptions on confounding. Therefore, for this illustration, we construct a status quo policy with an $\approx 18\%$ selection rate ($N = 2452$) by including all 452 records marked as enrolled in the program and randomly sampling 2000 unenrolled records. We fit the cost prediction model and nuisance functions using patient demographic variables, comorbidities, and prescription information measured in the preceding year. We use a linear regression to predict patient cost, and threshold predictions at the 55th percentile. This cutoff matches the threshold for physician enrollment recommendations of the deployed risk assessment (Obermeyer et al., 2019). We use 40% of the subsampled data to fit the cost prediction model and the remaining 60% as an observational sample for the policy comparison task. Because we do not have access to physician identifiers which can be used as an instrument, we leverage the MSM assumption for identification. We use a logistic regression classifier to fit nuisance functions. We use Algorithm 1 to estimate $[\underline{\hat{R}}_\delta(\pi, \pi_0; m, \hat{\mathcal{V}}), \hat{\overline{R}}_\delta(\pi, \pi_0; m, \hat{\mathcal{V}})]$ under the MSM with with $\Lambda = 1.2$. We report results over $N = 20$ trials with $K = 2$ folds.

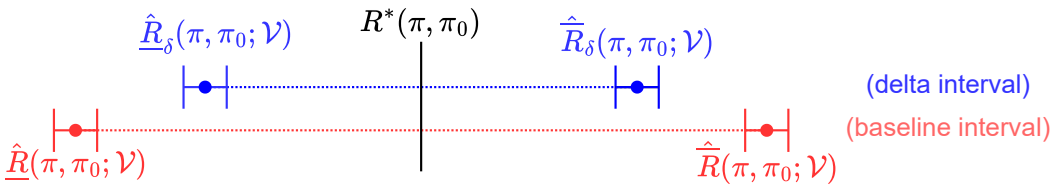


Figure 11. Two sources of uncertainty impacting policy regret bounds. (1) Confounding-related uncertainty impacts the width of the partially-identified regret interval (dotted lines). Our δ-regret interval yields tighter partially-identified regret intervals than the baseline (Theorem 4.3). (2) Finite sample error impacts estimation of upper- and- lower interval end points, as shown by solid confidence intervals. Estimators for the δ-regret interval and baseline interval have similar variance (see 95% CI of each interval reported in Figure 5).

E. Characterizing Benefits of Uncertainty Cancellation for Off-Policy Evaluation

In this Appendix, we provide additional discussion of our uncertainty cancellation approach. We begin by discussing the differential impact of **confounding-related uncertainty** and **finite sample uncertainty** on policy comparisons (§ E.1). Our δ-regret interval targets confounding-related uncertainty (Figure 11). We then provide more intuition for our uncertainty cancellation approach (§ E.2) and conclude by discussion further policy evaluation contexts which may benefit from our framework (§ E.3).

E.1. Sources of uncertainty impacting policy comparisons

Confounding-related uncertainty: Let $\mu(x) = \mathbb{E}[Y(1) \mid X = x]$ be the target outcome regression function. We can understand the impact of confounding related uncertainty on off-policy evaluation via the target regression decomposition

$$\mu(x) = \underbrace{\mathbb{E}[Y(1) \mid D^{\pi_0} = 1, X = x]}_{\text{Identified from observational data}} \cdot p(D^{\pi_0} = 1 \mid X = x) + \underbrace{\mathbb{E}[Y(1) \mid D^{\pi_0} = 0, X = x]}_{\in[0,1],\text{Unidentified from observational data}} \cdot p(D^{\pi_0} = 0 \mid X = x).$$

Now suppose we are in the asymptotic setting with no finite sample uncertainty. Thus we know $e_d(x) = p(D^{\pi_0} = d \mid X = x)$ and the observed outcome regression

$$\mu_1(x) := \mathbb{E}[Y(1) \mid D^{\pi_0} = 1, X = x] = \mathbb{E}[Y \mid D^{\pi_0} = 1, X = x]$$

from observational data.[16] However, because $D^{\pi_0} \not\perp\!\!\!\perp Y(1) \mid X$ due to confounding, the unobserved outcome regression $\mu_0(x) := \mathbb{E}[Y(1) \mid D^{\pi_0} = 0, X = x]$ is bounded within the worst-case interval $0 \leq \mu_0(x) \leq 1, \ \forall x \in X$.

As we show in Figure 12, bounds around the target regression $\mu(x) = e_1(x) \cdot \mu_1(x) + (1 - e_1(x)) \cdot \mu_0(x)$ widen as $e_1(x)$ decreases. This is because *decreasing* the propensity score *increases* the weighting of the unidentified term $\mu_0(x)$ in the regression decomposition. Therefore, the tightness of bounds around $\mu(x)$ depend on (1) the propensity function $e_d(x)$ and (2) the tightness of bounds around $\mu_0(x)$. The bounds around partially-identified v-statistics which we use to construct regret intervals inherits the same dependence on these terms (Lemma 5.2). As a result, **asymptotic regret intervals are wider when the status quo policy has a lower selection rate, regardless of the amount of data available for estimation.**

Finite-sample uncertainty: Finite sample uncertainty also impacts our ability to estimate regret interval endpoints. When less data is available, confidence intervals around upper and lower regret intervals will tend to be wider. As a result, although both the baseline and δ intervals have valid asymptotic coverage for the true regret (Appendix A), coverage can be violated under bias in estimates of the regret interval end points. Figure 5 shows confidence intervals around regret interval end-points as a function of dataset size. Our doubly-robust estimation approach improves data-efficiency of our regret estimator under no parametric assumptions on $\pi_0(x, u)$. We show that our DR estimator attains fast $\sqrt{n}$-rates in Appendix C.

[16]The second equality follows by Assumption 3.2

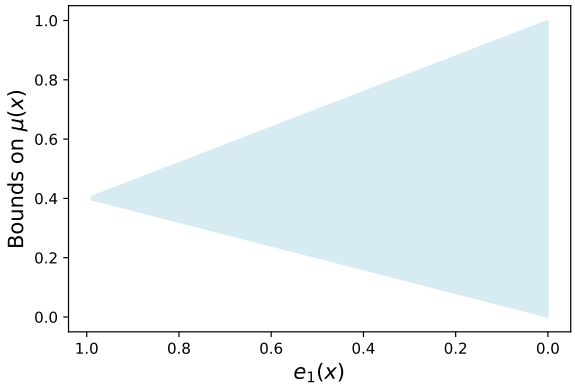


Figure 12. Bounds around $\mu(x) := \mathbb{E}[Y(1) \mid X = x]$ widen as the propensity score (i.e., the selection rate under the status quo policy) $e_1(x)$ decreases. Recall that $\mu(x) = e_1(x) \cdot \mu_1(x) + e_0(x) \cdot \mu_0(x)$. We fix $\mu_1(x) = .4$, let $\mu_0(x) \in [0, 1]$, and let $e_1(x)$ vary from 0 to 1. Bounds around the target regression increase as the weighting on the unobserved regression $\mu_0(x)$ increases.

Synthetic data example. We illustrate the relationship between confounding related uncertainty and finite sample error via a synthetic data experiment following a similar setup as the one outlined in Appendix D. We construct subgroups by defining two protected attributes with two levels each. The first correlates with X_1, while the second correlates with X_2. This yields four intersectional subgroups G1-G4 with varying sizes (ω) and selection rates under the status quo policy (γ).

Figure 13 compares the δ-regret interval and the baseline interval via worst case (WC) and instrumental variable (IV) partial identification strategies. We see that groups with lower selection rates under the status quo policy (e.g., G1, G2) have wider asymptotic bounds than those with higher selection rates (e.g., G3, G4). This indicates that the key driver of uncertainty at the asymptotic level is confounding. We can also examine statistical uncertainty in subgroup regret estimates by inspecting the size of confidence intervals. We observe that smaller subgroups (e.g., G1, G3) have larger variance in regret estimates than larger subgroups (e.g., G2, G4, full population).[17] Importantly, we observe that a small subgroup with a large selection rate (e.g., B5, $\gamma = .65$, $\omega = .08$) has tighter bounds than a larger subgroup with a lower selection rate (e.g., Population, $\gamma = 0.5$, $\omega = 1.0$). In line with our discussion above, this indicates that subgroup selection rates under π_0 are a key driver of uncertainty in asymptotic regret intervals.

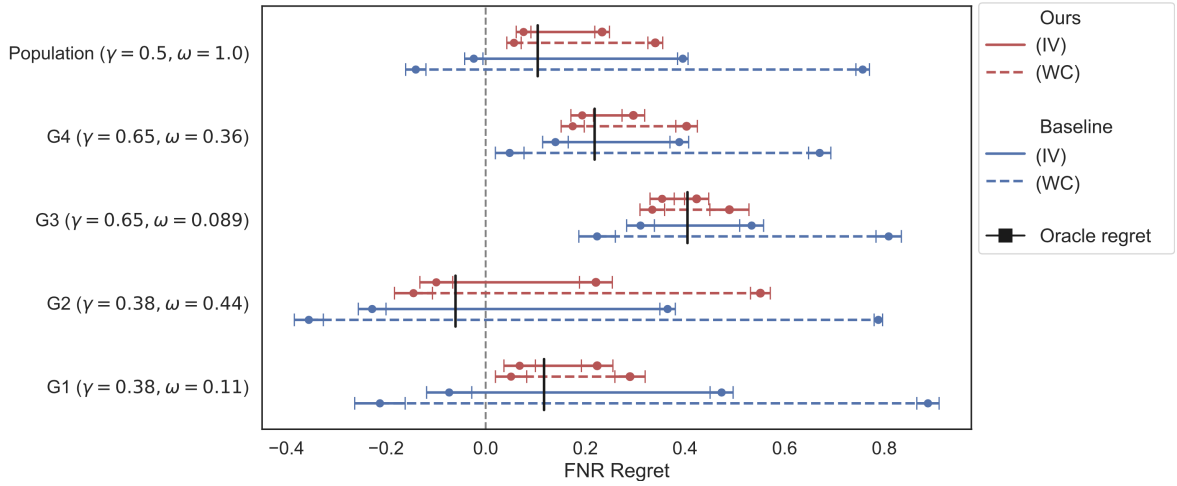


Figure 13. Top row indicates full population bounds, while lower rows show regret over subgroups of varying selection rates (γ) and sizes (ω). Horizontal bars indicate asymptotic regret bounds under confounding. We show statistical uncertainty over $N = 10$ runs by plotting 95% confidence intervals centered at each upper and lower asymptotic bound.

[17] We can isolate subgroup size (ω) as the source of this uncertainty because we fix selection rates across subgroups such that G1(γ)=G2(γ), G3(γ)=G4(γ).

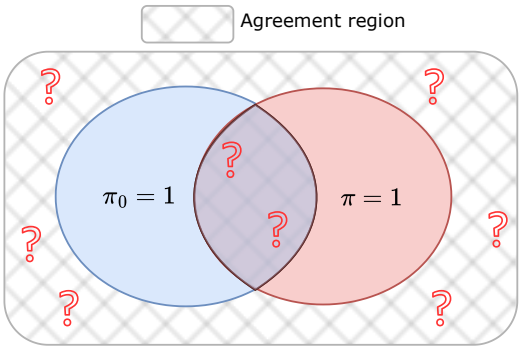


Figure 14. Uncertainty in the agreement region of the policy comparison (denoted via red question marks) cancels in comparative policy performance analyses.

E.2. Building intuition for uncertainty cancellation

The intuition for our approach is that we can safely ignore redundant uncertainty in the agreement region of the policy comparison (Figure 14). We see this concretely via the accuracy regret decomposition

$$\begin{aligned} R^*(\pi, \pi_0; m_u) &= m_u^*(\pi) - m_u^*(\pi_0) \\ &= p(T^\pi = Y(1)) - p(D^{\pi_0} = Y(1)) \\ &= (v_1(1,1) + v_1(1,0) + v_0(0,1) + v_0(0,0)) - (v_1(1,1) + v_1(0,1) + v_0(1,0) + v_0(0,0)) \\ &= v_1(1,0) + v_0(0,1) - v_1(0,1) - v_0(1,0). \end{aligned}$$

Observe that the agreement terms $v_1(1,1)$ and $v_0(0,0)$ cancel when we take the difference across policies. The agreement region ($\pi_0 = 1, \pi = 1$) at the center of the Venn Diagram *does not* add uncertainty in our context because $Y(1)$ is observed when $\pi_0 = 1$. Therefore, cancellation of $v_1(1,1)$ does not improve regret bounds. However, the agreement region in the complement space ($\pi_0 = 0, \pi = 0$) *does* contribute to uncertainty because $Y(1)$ is unobserved when $\pi_0 = 0$. Therefore cancellation of $v_0(0,0)$ is the main driver of the performance improvement. Theorem 4.3 formalizes this notion by showing that the improvement in the tightness of the δ-regret interval is proportional to the amount of uncertainty in the $v_0(0,0)$ term – i.e., $\alpha = \overline{v}_0(0,0) - \underline{v}_0(0,0)$.

We next turn to the positive predictive value regret for an example where uncertainty cancellation *does not* improve bounds

$$\begin{aligned} R^*(\pi, \pi_0; m_{a=1}) &= m_{a=1}^*(\pi) - m_{a=1}^*(\pi_0) \\ &= p(Y(1) = 1 \mid T^\pi = 1) - p(Y(1) = 1 \mid D^{\pi_0} = 1) \\ &= \frac{p(T = 1, Y(1) = 1)}{p(T = 1)} - \frac{p(D = 1, Y(1) = 1)}{p(D = 1)} \\ &= \frac{v_1(1,0) + v_1(1,1)}{\rho_{11} + \rho_{10}} - \frac{v_1(1,1) + v_1(0,1)}{\rho_{11} + \rho_{01}} \end{aligned}$$

where $\rho_{td} = p(T = t, D = d)$. Because only $v_1(1,1)$ cancels in this decomposition, we observe no improvement from the δ-regret interval for this performance measure.

More generally, we will observe a performance improvement from the δ-regret interval in any off-policy evaluation context where there is uncertainty arising from the agreement region of the policy comparison. **Importantly, this insight holds for any causal assumption (e.g., Rosenbaum's Γ, MSM, IV, Proximal) which can be expressed as pointwise bounding functions, which we detail in Appendix B**.

E.3. Other potential applications of uncertainty cancellation

While our work is concerned with policy evaluation under unmeasured confounding, other sources of uncertainty can also complicate policy comparisons. Our δ-regret interval may support partial identification under these uncertainty sources.[18]

Measurement Error and Noisy Labels. In many cases, labels available for model evaluation are observed under measurement error or label noise (Scott et al., 2013; Xia et al., 2019; Angluin & Laird, 1988; Guerdan et al., 2023). A common setup in this setting is to model the difference between the true outcome $Y^* \in \{0, 1\}$ and its proxy $Y \in \{0, 1\}$ via the false negative rate $\beta = p(Y = 0 \mid Y^* = 1)$ and false positive rate $\alpha = p(Y = 1 \mid Y^* = 0)$. However, when conducting comparative performance analyses, it may be possible to disregard measurement error in the agreement regions of the policy action space. This may tighten partial identification bounds studied in prior work (Fogliato et al., 2020).

Missing Protected Attributes. Protected attributes are sometimes unavailable for fairness assessments due to data collection or regulatory constraints (Kallus et al., 2022; Coston et al., 2019). However, we may wish to compare fairness statistics of alternative policies under missing protected attributes. If a partial identification approach is used to bound fairness characteristics under missing protected attributes (Kallus et al., 2022), it may be possible to tighten bounds by studying fairness differences over the disagreement region of the policy space.

[18] We offer these examples as illustrative of our technique and acknowledge further technical development would be required to develop these approaches in practice.