On Counterfactual Metrics for Social Welfare: Incentives, Ranking, and Information Asymmetry

Serena Wang
UC Berkeley

Stephen Bates
MIT

P. M. Aronow
Yale University

Michael I. Jordan
UC Berkeley

Abstract

From the social sciences to machine learning, it is well documented that metrics do not always align with social welfare. In healthcare, Dranove et al. (2003) showed that publishing surgery mortality metrics actually harmed sicker patients by increasing provider selection behavior. Using a principal-agent model, we analyze the incentive misalignments that arise from such *average treated outcome* metrics, and show that the incentives driving treatment decisions would align with maximizing total patient welfare if the metrics *(i)* accounted for counterfactual untreated outcomes and *(ii)* considered total welfare instead of averaging over treated patients. Operationalizing this, we show how counterfactual metrics can be modified to behave reasonably in patient-facing ranking systems. Extending to realistic settings when providers observe more about patients than the regulatory agencies do, we bound the decay in performance by the degree of information asymmetry between principal and agent. In doing so, our model connects principal-agent information asymmetry with unobserved heterogeneity in causal inference.

1 INTRODUCTION

As machine learning (ML) is increasingly deployed in dynamic social systems with asymmetries in power and information between stakeholders, a core challenge is that the *metrics* that are optimized do not always align with social welfare. From education to healthcare to recommender systems, it has been repeatedly shown that the impact of the ability of modern ML to optimize arbitrarily complex objectives is often limited by the difficulty in choosing *what* to optimize (Liu et al., 2023; McNee et al., 2006; Obermeyer et al., 2019).

One particularly consequential example of this gap is the incentive misalignment that occurs when *average treated outcome* measures are used as accountability and ranking metrics. This has led to measurable societal harm when those being ranked may selectively choose whom to treat. As an illustrative example, Dranove et al. (2003) showed that the publication of hospital mortality rate metrics by the New York Health Department led to dramatically worse outcomes for severely ill patients. Specifically, hospitals had an incentive to selectively treat the healthiest patients, rather than those who would benefit most from treatment. Today, the Centers for Medicare and Medicaid Services (CMS) continues to invest billions of dollars in the development of quality metrics (Wadhera et al., 2020; Casalino et al., 2016). In addition to determining direct provider compensation (Institute, 2022), these metrics also feed into large scale *ranking systems* such as the US News and World Report and the LeapFrog Hospital Safety Score (Rosenberg, 2013; Health and Agency, 2022). At the same time, studies have continued to question the relationship between these metrics and patient outcomes (see, e.g., Glance et al., 2021; Gonzalez and Ghaferi, 2014; Ryan et al., 2009; Hwang et al., 2014; Jha et al., 2008; Smith et al., 2017).

In this work, we directly study the incentive misalignment when *average treated outcomes* are used as quality metrics. To mitigate this misalignment, we propose alternative metrics that have a foundation in causal inference. Specifically, given counterfactual estimates of patient outcomes, we outline effective usage of these estimates and discuss prevailing limitations when providers may engage in strategic behavior.

More generally, our analysis applies to any environment where metrics are learned from data over which an agent controls treatment selection. We refer to healthcare as a running example where these effects are well documented. However, other domains subject

to this phenomenon include education, where student outcome metrics affect school rankings, funding, and accreditation (Koretz, 2017); and online ranking platforms for commercial businesses like restaurants that may exercise some screening over their customers.

To study the welfare effects that arise from this dynamic interaction between quality metrics and hospitals, we employ a *principal-agent model* where the *principal* chooses a quality metric as a reward function, and the *agent* responds by optimizing this reward function to the best of their private abilities and information. To analyze the welfare effects of metric choices, we apply a causal framework similar to that of policy learning, where the goal is to allocate treatments that maximize the total positive effects over a population relative to treating no one (Manski, 2008). Our key ingredient is that the principal only has indirect control of the implemented policy—they may design a metric that shapes an agent's reward and hence behavior but cannot directly choose an agent's policy.

Contributions. We show that *average treated outcome* metrics incur unbounded regret by this definition of welfare, and show that regret can be reduced by *(i)* accounting for counterfactual untreated outcomes and *(ii)* considering total welfare instead of average welfare among treated patients. Applying these two simple insights yields an optimal functional form for a quality measure that achieves zero regret as long as the principal can learn the mean conditional untreated potential outcomes. We refer to this as rewarding the *total treatment effect.* Connecting the proposed counterfactual metric to practice, we discuss two issues that arise when operationalizing the metric in real applications. First, we study the complications that arise when the total treatment effect is used to *rank* different agents that might serve different treatment populations. Second, we consider practical issues of *information asymmetry*, where the agent might observe more features about each patient than the principal. *Even an unbiased estimate of the counterfactual untreated outcome is not sufficient to maximize patient welfare when information asymmetries remain.* In addition to giving theoretical regret bounds, we also empirically show that it is not always better for a principal to condition on all known features, as this can *amplify regret.* Our model yields new connections between information asymmetry in the principal-agent model and unobserved heterogeneity in causal inference.

1.1 Related work

Our work combines technical structure from policy learning and contract theory to analyze misalignments in accountability metrics, which have long been critiqued in the social sciences.

Policy learning. The evaluation of treatment policies by their causal effects is well established in policy evaluation and policy learning (Manski, 2008; Hirano and Porter, 2009; Stoye, 2009; Athey and Wager, 2021). We directly apply the same measures of utility and regret in this work. Building on these measures, we consider a setting where a principal (regulatory agency) can only indirectly affect the policy through measuring an agent (hospital). We also note that our formulation is not the only way in which strategic behavior may arise. For example, Sahoo and Wager (2022) model patients' strategic responses when the hospital learns a policy with limited treatment capacity. Combining this model with ours would produce an interesting pipeline analysis.

Contract theory. The principal-agent model is well established in economics as a way to model incentives and equilibrium dynamics when a *principal* sets a contract with rewards as a function of actions, and an *agent* decides which action to take based on private information and costs (Laffont and Martimort, 2009; Gibbons et al., 2013; Milgrom and Roberts, 1992). Contract theory provides a structure to analyze moral hazard, where the structure of the contract and information asymmetry may lead to misalignment between the principal and agents' incentives (Arrow, 1963; Holmstrom and Milgrom, 1991). More recently, there is a growing recognition of the importance of the algorithmic and statistical aspects of contract theory (e.g., Carroll, 2015; Tetenov, 2016; Spiess, 2018; Dütting et al., 2019; Dütting et al., 2020; Bates et al., 2022; Alon et al., 2022). Our work contributes to this line of thought, linking moral hazard from contract theory with causal inference, and showing how incentive and statistical considerations jointly guide the choice of accountability metrics.

Specifically, we use this framework to analyze misalignments under a particular form of action and information asymmetry where the agent controls treatment selection, and might know more about each treatment unit than the principal. Lazzarini et al. (2022) applied contract theory to analyze *why* regulatory agencies might lean towards outcome-based contracts (such as the average treated outcome) rather than counterfactual assessment. Our model differs in both goal and setup: while they model the agent's effort levels, our model considers welfare effects when the agent has general control over all treatment assignments.

Strategic classification. The framework of *performative prediction* describes the distribution shifts and distortion in validity that result from agent strategy in response to ML-driven decision making (Perdomo et al., 2020; Hardt et al., 2016). Related lines of work in strategic classification have considered whether clas-

sifiers incentivize agent improvement (Ahmadi et al., 2022; Kleinberg and Raghavan, 2020; Bechavod et al., 2020; Haghtalab et al., 2020), drawing connections to causality (Miller et al., 2020; Shavit et al., 2020). We consider a specific strategic structure to improve metrics when agents manipulate treatment policies.

Accountability, auditing, and measurement. Historically, the mismatch between accountability metrics and welfare has been well documented across domains, from monetary policy to education (Goodhart, 1984; Campbell, 1979; Muller, 2019; Koretz, 2017; Mau, 2019). Extensive work in the social sciences has critically examined accountability and auditing practices (Power, 1994; Strathern, 1997; Hoskin, 1996; Rothstein, 2008). Our work builds on these qualitative insights by modeling a particular misalignment that occurs when the measured party has selection power over treatment allocations, also known as "creaming" (Lacireno-Paquet et al., 2002). Measurement theory has also given both qualitative and statistical tools for understanding the validity of measurements (Bandalos, 2018), with recent extensions to fair ML (Jacobs and Wallach, 2021). Guerdan et al. (2023) apply counterfactual modeling to estimate the measurement error when proxy outcomes are used to guide treatment decisions. Complementing this focus on measurement validity, we model the treatment incentives induced when measurements are used to reward agents, and study the resulting welfare effects.

2 PRINCIPAL-AGENT MODEL

To analyze the incentives and welfare effects that arise from quality metrics, we define a principal-agent model where we will refer to the organization that collects the metrics and pays or ranks the providers as the *principal*, and the healthcare providers as *agents*. First, the principal specifies a function for rewarding the agents based on their actions and the observed outcomes. In response, the agent allocates treatment decisions with knowledge of this reward function. Our focus is on designing reward functions with high utility for the principal, which corresponds to social welfare.

Formal model. We suppose that a single agent has access to independently and identically distributed samples of characteristics $X_i \in \mathcal{X}$ for $i \in \{1, ..., n\}$ treatment units (referred to generally as the *treatment population*). The agent assigns a binary treatment according to treatment rule $\pi : \mathcal{X} \to [0, 1]$, and we use $T_i^\pi \in \{0, 1\}$ to denote a Bernoulli random variable indicating the realizations of the treatment rule: $\pi(X_i) = P(T_i^\pi = 1 | X_i)$. This framework allows for T^π to be either stochastic or deterministic.

To model patient outcomes from treatment, we apply

the potential outcomes framework (Neyman, 1923; Rubin, 1974). Let $Y_i(t)$ be the potential outcome if the patient had received treatment $t \in \{0, 1\}$. Let $Y_i \in \mathbb{R}$ denote the observed outcome under treatment assignments $Y_i = Y_i(T_i^\pi)$ under the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980), which implies the consistency and non-interference assumptions. Let $\tau(X) = E[Y_i(1) - Y_i(0)|X]$ denote the conditional average treatment effect given covariates X, and let $\mu_t(X) = E[Y_i(t)|X]$ denote the conditional mean of the potential outcome under treatment t.

We suppose that the principal observes $\{X_i, T_i^\pi\}_{i=1}^n$ (also denoted $\mathbf{X}, \mathbf{T}^\pi$) for all units, and Y_i for all units for which $T_i^\pi = 1$ (denoted $\mathbf{Y}$). The principal must then choose a reward function $w : \mathcal{X}^n \times \{0, 1\}^n \times \mathbb{R}^n \to \mathbb{R}$ with which to reward the agent.

Turning to our behavioral assumption, we consider an agent that is risk neutral and maximizes their expected reward. That is, the agent chooses the treatment rule π from their set of possible treatment rules Π that maximizes $E[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$. Let $\pi^w \in \arg\max_{\pi \in \Pi} E[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$ denote this best response. In maximizing this expected reward, we assume that the agent knows $\mu_t(x)$ for all $x \in \mathcal{X}$ and $t \in \{0, 1\}$. Note that there is no explicit model of the agent's cost here to keep the focus on incentives induced by accountability metrics alone, and all budget constraints are contained in the agent's feasible set of treatment rules Π. We further discuss these assumptions and possible extensions in the Appendix.

Total welfare and regret. Following existing work in policy evaluation (Manski, 2008), for a given treatment rule π, we define the total effect of treatment on welfare as $V(\pi) = E[Y_i(T_i^\pi) - Y_i(0)]$. $V(\pi)$ is the utility of treatment rule π relative to the alternative outcomes under no treatment (Manski, 2008; Athey and Wager, 2021). Thus, maximizing $V(\pi)$ also maximizes total welfare $E[Y]$ compared to treating no one. As also done in policy learning (Athey and Wager, 2021), to evaluate different quality measures chosen by the principal, we define the regret for a given policy π compared to the best feasible policy in Π to be $R(\pi) = \max_{\tilde{\pi} \in \Pi} V(\tilde{\pi}) - V(\pi)$. We compare different choices of quality measures w by analyzing the effect on total welfare for the induced treatment rule π. In other words, the principal's goal is to choose a reward function w that leads the agent to best respond with a treatment rule with minimal regret, $R(\pi^w)$.

3 COMPARISONS OF QUALITY METRICS

Using a principal-agent model, we formally compare different choices of quality metrics w by analyzing the regret for the induced treatment rule π given by the agent's best response. Starting with the *status quo* of rewarding the average treated outcome, we show that this incurs unbounded regret. We reduce this to two main problems with the metric: *(i)* lack of accounting for untreated outcomes, and *(ii)* rewarding an average effect instead of a total effect. Addressing each of these in turn, we show that the regret for rewarding the *average treatment effect on the treated (ATT)* is bounded but can still be high, and that rewarding the *total treatment effect* finally achieves zero regret.

Status quo: average treated outcome. We begin by analyzing regret under the current common quality measure that rewards the average treated outcome, as done by the mortality measures in the New York and Philidelphia health departments in the 1990s analyzed by (Dranove et al., 2003), and in many CMS quality measures (Institute, 2022). Lazzarini et al. (2022) also refers to these as "outcome contracts." The reward function for the *average treated outcome (ATO)* takes the following form:

Reward Function 1 (ATO):
$$w_{\text{ATO}}(\mathbf{X}, \mathbf{T}^{\pi}, \mathbf{Y}) = \begin{cases} \frac{\sum_{i=1}^{n} Y_i T_i^{\pi}}{\sum_{i=1}^{n} T_i^{\pi}} & \sum_{i=1}^{n} T_i^{\pi} > 0, \\ 0 & \text{otherwise.} \end{cases}$$
The agent's unconstrained best response is
$\pi^{w_{\text{ATO}}}(x) = \mathbb{1}(x \in \arg\max_x \mu_1(x) \text{ and } \mu_1(x) > 0)$.

Proposition 1 (ATO Regret)**.** *If the conditional mean untreated potential outcomes $\mu_0(x)$ are unbounded, then the regret for the reward function w_{ATO} may be arbitrarily large.*

Intuitively, there are at least two failure modes that can lead to this unbounded regret. First, this reward leads the agent to ignore higher treatment effects of the patients with a lower treated outcome, such as sicker patients with higher mortality probability but more benefit from surgery. This matches the findings from Dranove et al. (2003). Second, w_{ATO} rewards agents that treat the patients with a higher treated outcome, even though treatment actually harms those patients, such as healthier patients who might incur more risks or side effects from treatment.

More broadly, there are two main problems with the construction of the ATO metric that lead to this unbounded regret. First, the lack of accounting for counterfactual outcomes leads to the two failure modes above. Second, the measure of an average outcome instead of a total outcome means that the agent will only treat the *single* patient with the covariate value x that *maximizes* μ_1. We next analyze several reasonable modifications to reward functions that address each of these problems.

Accounting for counterfactuals. When the principal operates with full information of the agent's selection covariates X_i, then regret can be reduced if they also have access to an unbiased estimator of the mean conditional untreated potential outcome.

Assumption 1. *The principal accesses an estimator $\hat{\mu}_0(x)$ which is unbiased: $E[\hat{\mu}_0(x)] = \mu_0(x) \;\; \forall x \in \mathcal{X}$.*

In general, obtaining an unbiased estimator $\hat{\mu}_0(x)$ can be difficult, but circumstances under which causal inference can be reliably conducted are well understood (Hernan and Robins, 2020). As a concrete example, suppose the principal has access to an auxiliary data source $\{X'_j, T'_j, Y'_j\}_{j=1}^{m}$, with outcomes for untreated patients with $T'_j = 0$, collected from clinical trials or observational data. In addition to standard assumptions for identification of $\mu'_0(x) = E[Y'_j(0)|X'_j = x]$, the principal may use this data if the distribution of the conditional untreated potential outcome is also the same, $\mu'_0(x) = \mu_0(x)$, and the support of X'_j covers the support of X_i. The difficulty of obtaining such a dataset is lessened by the fact that the principal does not require identification of the *treated* potential outcome $\mu'_1(x)$ or the conditional average treatment effect $\tau'(x)$, so the treatment need not be the same. Access to additional scientific knowledge (in the form of, e.g., a more intricate structural model or functional form assumptions) can also aid in the estimation of $\mu_0(x)$. Medical research continues to develop patient risk scores using combinations of such methods (Sullivan et al., 2004; Jones and Cossart, 1999) and evaluate their validity (Kaafarani et al., 2011; Janssens, 2019).

Under the perhaps optimistic assumption of access to an unbiased estimator $\hat{\mu}_0(x)$, this work focuses on effective ways for the principal to apply this estimator. It turns out that even given this unbiased estimate, principal-agent incentive misalignment can still occur, and there are still pitfalls with its downstream usage. We begin by showing how to effectively incorporate this estimator into the reward function. In later sections, we discuss effective usage in a ranking context and regret bounds under information asymmetry.

Given $\hat{\mu}_0(x)$, the principal can modify w_{ATO} by directly subtracting an estimate of the untreated potential outcomes, and thus reward the *average treatment effect on the treated (ATT)*:

Reward Function 2 (ATT):
$$w_{\text{ATT}}(\mathbf{X}, \mathbf{T}^{\pi}, \mathbf{Y}) = \begin{cases} \frac{\sum_{i=1}^{n} (Y_i - \hat{\mu}_0(X_i)) T_i^{\pi}}{\sum_{i=1}^{n} T_i^{\pi}} & \sum_{i=1}^{n} T_i^{\pi} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The agent's unconstrained best response is
$\pi^{w_{\mathrm{ATT}}}(x) = \mathbb{1}(x \in \arg\max_x \tau(x) \text{ and } \tau(x) > 0)$.
The resulting regret is bounded, but still not zero.

Proposition 2 (ATT Regret). If $\hat{\mu}_0(x)$ is unbiased and π is unconstrained, then the regret for the reward function w_{ATT} is upper bounded as
$R(\pi^{w_{\mathrm{ATT}}}) \leq \max_{\pi \in \Pi} V(\pi)$.

Now that the reward function accounts for the counterfactual untreated outcome, Proposition 2 shows that the regret cannot exceed the maximum utility. This is notably not true of the w_{ATO}, where the regret can be arbitrarily high due to the agent sometimes treating those with a negative treatment effect. Still, while accounting for untreated potential outcomes avoids treating those with a negative treatment effect, the ATT as a reward function still suffers from misalignment with total welfare due to the fact that it rewards the *average* effect rather than the *total* effect. This means that in the best response, the agent only treats patients with the single value x with maximum treatment effect $\tau(x)$.

Rewarding total effects. To expand the agent's treatments to cover *all* individuals who would benefit, we modify the above reward function by simply removing the denominator, thus rewarding a *total* effect instead of an *average* effect. This yields a reward function for the *total treatment effect (TT)*.

Reward Function 3 (TT):
$w_{\mathrm{TT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y}) = \sum_{i=1}^n (Y_i - \hat{\mu}_0(X_i)) T_i^\pi$.

The agent's unconstrained best response is
$\pi^{w_{\mathrm{TT}}}(x) = \mathbb{1}(\tau(x) > 0)$, which yields zero regret.

Proposition 3 (TT Regret). If $\hat{\mu}_0(x)$ is unbiased, then the regret is $R(\pi^{w_{\mathrm{TT}}}) = 0$.

The regret is zero regardless of the feasible set Π. Thus, with two modifications to the status quo w_{ATO}, a quality measure w_{TT} can be constructed that is aligned with total welfare, as long as the principal has access to an unbiased estimator $\hat{\mu}_0(x)$.

4 RANKING WITH MULTIPLE AGENTS

In Section 3 we've shown that rewarding the total treatment effect leads the agent to maximize total welfare. While this theory applies cleanly in isolation, in real systems, quality measures are often further employed to *rank* hospitals (Rosenberg, 2013; Health and Agency, 2022; Smith et al., 2017). It turns out that even w_{TT} can exhibit problems as a ranking measure when different hospitals have different treatment population sizes and distributions. To apply reward

functions as ranking measures, we show that the total treatment effect can be modified to a more general form that allows for reweighting by covariates X while still preserving incentive alignment. By reweighting the reward function relative to a reference covariate distribution, we show that the resulting quality measure leads to better hospitals receiving better rankings.

Notationally, discussion of ranking requires extending our setting to account for multiple agents. Suppose each agent $k \in \{1, ..., K\}$ observes its own sample of n_k patients with covariates drawn i.i.d. from distribution P_{X^k} with the same support $\mathcal{X}$. Each agent has different treatment effects, denoted by $\mu_t^k(x)$ and $\tau^k(x)$. For rankings to be meaningful, we assume that the untreated potential outcome is the same for all k: $\mu_0^k(x) = \mu_0(x)$ for all k. In short, one provider not treating a patient is equivalent to another provider not treating the same patient. Extending the principal's action space to the multi-agent ranking setting, the principal publishes score functions $\{w_k\}_{k=1}^K$, and each agent k best responds individually with their own treatment policy π_k. The agents are then ranked from highest to lowest score function values. We expand this notation in the Appendix.

Defining desirable ranking properties. For any regulatory agency or potential patient that would utilize these rankings, a clear desirable property would be that *better hospitals should be ranked higher*. From the agents' perspectives, this property may also make the scores feel more "fair." We formally define this property with two different degrees of strictness for the meaning of "better."

First, we define "better" as an agent having *uniformly* higher treatment effects for all possible covariate values x, such that any patient would be better off being treated by this agent.

Definition 1 (Uniform Rank Preservation). A set of score functions $\{w_k\}_{k=1}^K$ preserves treatment effect ordering uniformly over $\mathcal{X}$ if for all $j, k \in \{1, ..., K\}$,
$\tau^j(x) \geq \tau^k(x) \; \forall x \in \mathcal{X} \implies$
$\max_{\pi_j \in \Pi_j} E[w_j(\mathbf{X}^k, \mathbf{T}^{\pi_j}, \mathbf{Y}^j)] \geq \max_{\pi_k \in \Pi_k} E[w_k(\mathbf{X}^k, \mathbf{T}^{\pi_k}, \mathbf{Y}^k)]$

A more relaxed version of the uniform rank preservation requirement is one where an agent is "better" if it has higher treatment effects on *average* over a reference covariate population P_{X_0}.

Definition 2 (Relative Rank Preservation). A set of score functions $\{w_k\}_{k=1}^K$ preserves treatment effect ordering relative to a reference population P_{X_0} with support $\mathcal{X}$ if for all $j, k \in \{1, ..., K\}$,
$E[\tau^j(X_0)] \geq E[\tau^k(X_0)] \implies$
$\max_{\pi_j \in \Pi_j} E[w_j(\mathbf{X}^k, \mathbf{T}^{\pi_j}, \mathbf{Y}^j)] \geq \max_{\pi_k \in \Pi_k} E[w_k(\mathbf{X}^k, \mathbf{T}^{\pi_k}, \mathbf{Y}^k)]$

This relative definition requires explicitly defining a reference population, which calls for careful consideration of policy goals and societal needs. Any set of scores will implicitly prioritize some populations, and calling attention to this as an explicit part of the ranking properties induced by quality measures can help policymakers more intentionally align their choices with policy goals.

Satisfying desirable ranking properties. We now formally show that w_{TT} as written does not directly satisfy these ranking properties.

Proposition 4. If w_k is directly given by w_{TT} for each agent k, then both ranking properties in Definitions 1 and 2 will be violated.

Intuitively, this breaks down because w_{TT} is subject to two auxiliary effects on top of the treatment effects. First, agents with a larger treatment population n_k (e.g., larger hospitals) will have higher rankings even with the same conditional average treatment effects. Second, agents with different distributions of covariates P_{X^k} but the same conditional average treatment effects will also end up with different rankings if some covariate values are "easier" to treat than others.

To mitigate these auxiliary effects, we show that there exists a general modular form of w_k that preserves the zero regret property for individual agent best responses. This general form can then be tailored to satisfy desirable ranking properties. In particular, consider the following weighted total treatment effect reward function:

Reward Function 4 (Weighted TT):
$w_{\text{TT}}^g(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y}) = \sum_{i=1}^n (Y_i - \hat{\mu}_0(X_i)) T_i^\pi g(X_i).$

Any reward function in this family induces the desired agent best response.

Theorem 1 (Incentive Alignment). *Suppose $\hat{\mu}_0(x)$ is unbiased, and π is unconstrained. For any function $g : \mathcal{X} \to \mathbb{R}^+$, w_{TT}^g yields an agent best response with regret $R(\pi^{w_{TT}^g}) = 0$.*

Theorem 1 shows that reweighting the reward function by any function of the covariates X does not hurt incentive alignment. Thus, the principal may choose functions g_k for each agent to achieve desirable ranking properties. Specifically, setting g_k to reweight each agent's covariate distribution to the reference distribution P_{X_0} satisfies both ranking properties in Definitions 1 and 2.

Theorem 2 (Ranking Desiderata Satisfied). *Let P_{X^k} be absolutely continuous with respect to P_{X_0}, and let $g_k = \frac{1}{n_k} \frac{dP_{X_0}}{dP_{X^k}}$ be the normalized Radon–Nikodym derivative of the reference distribution P_{X_0} with respect*

to agent k's covariate distribution P_{X^k}. Then setting w_k to be $w_{TT}^{g_k}$ for agent k's treatment population satisfies both ranking properties in Definitions 1 and 2 as long as Π_k is unconstrained and treatment effects are nonnegative, $\tau^k(x) \geq 0$, for all $k \in \{1, ..., K\}$.

Theorem 2 shows that a simple distributional reweighting can achieve the desirable ranking properties that preserve treatment effect ordering both uniformly over all $\mathcal{X}$ and relatively to some reference population P_{X_0}. In practice, if the exact Radon–Nikodym derivatives are not known, the importance sampling literature contains many techniques for estimating expectations with distributional reweighting (Owen, 2013). Overall, this simple reweighting modification of the total treatment effect score function addresses important policy considerations when quality measures are used for ranking.

5 INFORMATION ASYMMETRY

So far, the incentive alignment and ranking properties have relied on the assumption that both principal and agent operate with the same covariate information X. In practice, Dranove et al. (2003) remark that "providers may be able to improve their ranking by selecting patients on the basis of characteristics that are unobservable to the analysts but predictive of good outcomes." Under such information asymmetry, we show that even the optimistic assumption of an unbiased estimator $\hat{\mu}_0(x)$ is not enough to guarantee zero regret. We both upper and lower bound regret in terms of the additional *heterogeneity* observed by the agent.

Suppose the agent observes additional covariates $U_i \in \mathcal{U}$, and selects a treatment rule $\pi : \mathcal{X}, \mathcal{U} \to [0, 1]$. Suppose the principal still observes only $\{X_i, T_i^\pi, Y_i\}_{i=1}^n$, and chooses a reward function $w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})$ that does not depend on U_i. Let $\mu_t(X, U) = E[Y_i(t)|X, U]$ and $\tau(X, U) = E[Y_i(1) - Y_i(0)|X, U]$. The utility and regret are still defined as in Section 2.

Applying the optimal reward function from the full information setting in Section 3, suppose the principal rewards the agent with the total treatment effect w_{TT}. As in Section 3, suppose the principal applies an unbiased estimator $\hat{\mu}_0(X)$ of the untreated potential outcome conditional on X, with $E[\hat{\mu}_0(x)] = \mu_0(x)$ $\forall x \in \mathcal{X}$. Note that the principal identifies $\mu_0(X)$, but *not* $\mu_0(X, U)$. We show that the regret is bounded if the effect of the agent's private information U_i on the untreated potential outcomes is bounded.

Assumption 2 (Bounded Heterogeneity). The effect of U_i on the conditional untreated potential outcome is bounded as $E[|\mu_0(X_i) - \mu_0(X_i, U_i)|] \leq \gamma_{\text{marg}}$.

Note that $\mu_0(X_i) = E[\mu_0(X_i, U_i)|X_i]$. Thus, the het-

erogeneity bound is akin to bounding a statistical error between the conditional untreated potential outcome $\mu_0(X_i, U_i)$ known to the agent, and "marginal" $\mu_0(X_i)$ estimated by the principal. We can both upper and lower bound the regret in terms of this error:

Theorem 3 (Regret With Information Asymmetry). *Suppose $\hat{\mu}_0(x)$ is unbiased. Under Assumption 2, the regret is upper bounded as $R(\pi^{w_{TT}}) \leq 2\gamma_{marg}$.*

This upper bound is tight up to a linear constant.

Proposition 5. $\forall \epsilon > 0$, there exist distributions of $X_i, U_i, Y_i(0), Y_i(1)$ wherein $R(\pi^{w_{TT}}) \geq \gamma_{\mathrm{marg}} - \epsilon$.

Thus, this notion of *heterogeneity* is key to determining regret under information asymmetry. As a realistic example of cases when Assumption 2 might be satisfied, studies of cardiovascular disease risk have shown that "the magnitude of risk related to smoking is far larger than any ostensible benefit related to moderate drinking" (Mukamal, 2006). Thus, if U were some attribute for which the relative effect on top of X was small, then γ_{marg} would be small. On the other hand, Rodgers et al. (2019) report that sex hormones and diabetes have compounding effects on cardiovascular disease risk. If X and U have strong compounding effects on $Y(0)$, then γ_{marg} could be large.

Information asymmetry and confounding. Information asymmetry relates closely to the possibility of confounding bias in the estimator $\hat{\mu}_0(x)$. First, the agent's knowledge of U could mean that the data source from which the principal estimated $\hat{\mu}_0(x)$ was also confounded by U. In this case, the literature on sensitivity analysis and policy learning with unobserved confounding proposes a range of robust estimates for $\mu_0(x)$ (see, e.g., Yadlowsky et al., 2022; Kallus and Zhou, 2018). In our setting, robust estimation of $\mu_0(x)$ is not enough, since the agent's treatment rule can depend on U. Still, the minimax techniques from these works may be a useful avenue for designing future robust reward functions w. Second, information asymmetry can exacerbate confounding if the agent were able to directly affect the principal's estimator $\hat{\mu}_0(x)$, which may happen if, e.g., the principal were to estimate $\hat{\mu}_0(x)$ from the agent's untreated units with $T_i^\pi = 0$. We discuss this case in detail in the Appendix, and show that a stronger assumption yields a similar bound to Theorem 5. Most importantly, our key result is that even without confounding, information asymmetry *still* causes problems via the agent's ability to discriminate on U.

6 EXPERIMENTS

We turn to several clinical datasets to evaluate the welfare impacts of different quality metrics under different conditions of information asymmetry. We show empirically that the regret incurred by w_{ATO} can be high. We also show that under information asymmetry, regret can be *amplified* if the principal estimates $\hat{\mu}_0(x)$ conditioned on some subsets of features. With careful feature selection for x, regret may be reduced.

Horse Colic dataset. The Horse Colic dataset from the UCI repository (Dua and Graff, 2017) contains $n = 300$ horse colic cases. Horses were either treated with surgery ($T = 1$) or not ($T = 0$), with a treatment rate of 0.6. The 20 covariates include each horse's age and presenting symptoms such as abdominal distension, pulse, blood test results, etc. (see Appendix for a full list). Our outcome of interest Y is whether the horse lived ($Y = 1$) or died ($Y = -1$). To approximate the mean conditional potential outcomes, we apply a logistic model with interaction terms between the treatment and covariates: $P(Y(t) = 1 | X = x) = \sigma(\beta_0 + \beta_1 x + \beta_2 t + \beta_3 xt)$. We estimate the parameters β on the dataset using logistic regression, and take these as given to produce $\mu_0(x)$ and $\mu_1(x)$. The fitted $\mu_0(x)$ and $\mu_1(x)$ show 62 horses benefiting from surgery and 146 being better off without surgery. On the horses that would benefit, the average benefit was 0.147, which is fairly significant. The clinical validity of these estimated potential outcomes cannot be verified from this data alone, and we instead take these estimates as synthetic potential outcomes.

International Stroke Trial dataset. We also consider data from the International Stroke Trial Collaborative Group (1997), which was a randomized trial studying the effects of drug treatments in acute stroke. Kallus and Zhou (2018) studied this dataset in a different policy learning setting, and we apply a similar setup by comparing treating with high doses of heparin and aspirin ($T = 1$) with aspirin alone ($T = 0$). This leaves $n = 7264$ patients and a treatment rate of 0.33. The 20 covariates include each patient's age, sex, and clinical symptoms such as prior stroke types and complications. Like Kallus and Zhou (2018), we consider a scalarized outcome score $Y \in [-4, 3]$ that accounts for patient outcomes including death, recovery, and side effects at 14 days and 6 months after treatment (details in the Appendix). We approximate the mean conditional potential outcomes using a linear model with interaction terms between treatment and covariates: $E[Y(t)|X = x] = \beta_0 + \beta_1 x + \beta_2 t + \beta_3 xt$. We fit an OLS estimate of β, and use the resulting $\mu_0(x)$ and $\mu_1(x)$ functions as synthetic mean conditional potential outcomes. The fitted $\mu_0(x)$ and $\mu_1(x)$ showed

1360 patients benefiting from heparin and 5904 being better off without heparin. On the patients that would benefit, the average benefit of treatment was 0.025. This tracks with the study's findings that the benefit of heparin was non-significant and inconclusive.

Results and discussion. We first compare the reward functions from Section 3 by calculating the utility and regret empirically over the dataset using the synthetic potential outcomes. Table 1 shows that for both datasets, the utility for w_{TT} is positive and higher for the Horse Colic dataset than for the Stroke Trial dataset, which tracks with the clinical finding that the heparin treatment did not have significant effect. The utility for w_{ATO} is close to zero for both datasets.

Next, we consider the effect of information asymmetry on regret. The last two rows of Table 1 show the regret when the principal applies w_{TT}, but either observes no covariates ("no info") and identifies $E[Y(0)]$, or only observes age and/or sex ("demographic info") and identifies $E[Y(0)|\text{demographics}]$. While conditioning on age has no effect on the Horse Colic dataset, on the Stroke Trial dataset, conditioning on demographics actually *hurts* utility compared to using "no info." This suggests that it is not always better for the principal to condition on all known information, and thus policymakers should exercise caution in designing stratified quality metrics. For example, CMS currently measures age-specific kidney transplant rates for organ procurement organizations (OPOs) (QCOR, 2023), and our findings question the value of such incomplete stratification under OPO treatment selection. Theoretically, this finding structurally mirrors the phenomenon of *bias amplification* when estimating causal effects with unobserved confounding, where conditioning on more observed features can actually increase bias (Pearl, 2012). Here we observe *regret amplification*, and we encourage replication of similar analyses by regulatory agencies with internal data sources.

We also study a more continuous spectrum of information asymmetry by showing regret as the principal accumulates increasingly large subsets of the available features. Relating this to Section 5, we first sort the features in ascending order of feature importance, as measured by estimating γ_{marg} when the principal knows only the individual feature. Then, Figure 1 shows regret when the principal knows increasingly large feature subsets, building up starting from the most "important" feature. Regret is reduced significantly after accounting for less than half of the features. In practice, the principal would not know true values of γ_{marg}. However, approximations of γ_{marg} may serve as a reasonable heuristic for feature selection when a regulator has a large set of known covariates, but needs to prune them for interpretability or cost.
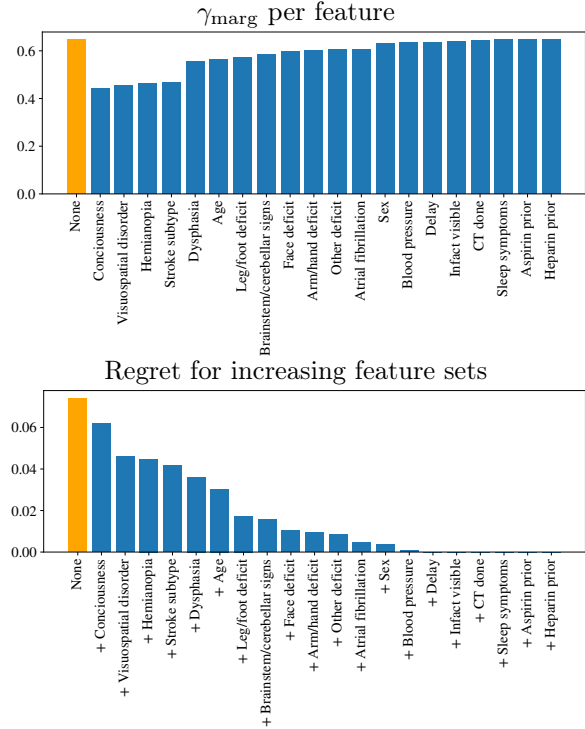

Figure 1: Regret under fine-grained information asymmetry on the Stroke Trial dataset (Horse Colic in the Appendix). The *top* plot shows γ_{marg} values if the principal only knows each individual feature. The *bottom* plot shows regret as the principal accumulates features from the left (most important).

7 CONCLUSIONS

We have studied the harm to social welfare that occurs when accountability metrics are not aligned with social utility. Even under optimistic assumptions about the availability of an unbiased counterfactual estimate, the potential for regret still exists under information asymmetry with treatment effect heterogeneity. Given the compounded difficulty of estimating causal effects on top of our consideration of treatment incentives and ranking, we recommend that designers exercise caution, humility, and vigilance in their construction of metrics. The task is difficult, but we have established the contours of one potentially fruitful approach.

7.1 Future work

There are many important avenues of future work extending from our framework. Information asymmetry presents a prevailing challenge in which we provided bounds on regret, and further analysis on how to improve these regret bounds using ideas from contract design or robust policy learning would be promising and impactful. There is also significant room for modeling

Table 1: Utility and regret comparisons for different reward functions. For each reward function w, we report utility $V(\pi^w)$, regret $R(\pi^w)$, and the realized treatment rate $P(T^{\pi^w} = 1)$.

	Horse Colic dataset			Stroke Trial dataset		
Reward function	Utility	Regret	Treat rate	Utility	Regret	Treat rate
w_{ATO}	0.00000	0.1470	0.1922	0.00004	0.0251	0.0001
w_{ATT}	0.00784	0.1391	0.0039	0.00013	0.0250	0.0001
w_{TT}	0.14695	0.000	0.2431	0.02518	0.000	0.1872
w_{TT} (no info)	0.09761	0.0493	0.6275	-0.04888	0.0741	0.4829
w_{TT} (demographic info)	0.09761	0.0493	0.6275	-0.06392	0.0891	0.5041

extensions. While we considered the simplest modeling framework that could capture the incentive effects of quality measures on treatment selection, there are many other significant factors to consider in practice, including uncertainty and variation in treatment costs, competition between agents, and the ability of treatment units (e.g., patients) to decide where to seek treatment. Analyzing end-to-end regret with these additional factors would be valuable future work.

Acknowledgements

References

30 years of uptodate: The evolution of clinical decision support and the future of evidence-based medicine. *Wolters Kluwer*, 2022. URL https://www.wolterskluwer.com/en/expert-insights/30-years-of-uptodate-evolution-of-clinical-decision-support-future-of-evidence-based-medicine.

Saba Ahmadi, Hedyeh Beyhaghi, Avrim Blum, and Keziah Naggita. On classification of strategic agents who can both game and improve. In *3rd Symposium on Foundations of Responsible Computing (FORC)*, 2022.

Tal Alon, Paul Dütting, Yingkai Li, and Inbal Talgam-Cohen. Bayesian analysis of linear contracts. *arXiv preprint arXiv:2211.06850*, 2022.

Kenneth J Arrow. Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5):941–973, 1963.

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.

Deborah L Bandalos. *Measurement Theory and Applications for the Social Sciences*. Guilford Publications, 2018.

Stephen Bates, Michael I. Jordan, Michael Sklar, and Jake A. Soloff. Principal-agent hypothesis testing. *arXiv preprint arXiv:2205.06812*, 2022.

Yahav Bechavod, Katrina Ligett, Zhiwei Steven Wu, and Juba Ziani. Causal feature discovery through strategic modification. *arXiv preprint arXiv:2002.07024*, 3, 2020.

Donald T Campbell. Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1):67–90, 1979.

Gabriel Carroll. Robustness and linear contracts. *American Economic Review*, 105(2):536–63, 2015.

Lawrence P Casalino, David Gans, Rachel Weber, Meagan Cea, Amber Tuchovsky, Tara F Bishop, Yesenia Miranda, Brittany A Frankel, Kristina B Ziehler, Meghan M Wong, et al. Us physician practices spend more than \$15.4 billion annually to report quality measures. *Health Affairs*, 35(3):401–406, 2016.

David Dranove, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. Is more information better? the effects of "report cards" on health care providers. *Journal of Political Economy*, 111(3):555–588, 2003.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

Paul Dütting, Tim Roughgarden, and Inbal Talgam-Cohen. Simple versus optimal contracts. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 369–387, 2019.

Paul Dütting, Tim Roughgarden, and Inbal-Talgam Cohen. The complexity of contracts. In *Proceedings of the 2020 ACM-SIAM Symposium on Dis-*

crete Algorithms, pages 2688–2707. SIAM, Philadelphia, PA, 2020.

Robert Gibbons, John Roberts, et al. The Handbook of Organizational Economics. Princeton University Press Princeton, NJ, 2013.

Laurent G Glance, Caroline P Thirukumaran, Changyong Feng, Stewart J Lustik, and Andrew W Dick. Association between the physician quality score in the merit-based incentive payment system and hospital performance in hospital compare in the first year of the program. JAMA Network Open, 4 (8):e2118449–e2118449, 2021.

Andrew A Gonzalez and Amir A Ghaferi. Hospital safety scores: do grades really matter? JAMA Surgery, 149(5):413–414, 2014.

Charles AE Goodhart. Problems of Monetary Management: The UK Experience. Springer, 1984.

Luke Guerdan, Amanda Coston, Kenneth Holstein, and Zhiwei Steven Wu. Counterfactual prediction under outcome measurement error. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, pages 1584–1598, 2023.

Nika Haghtalab, Nicole Immorlica, Brendan Lucier, and Jack Z Wang. Maximizing welfare with incentive-aware evaluation mechanisms. arXiv preprint arXiv:2011.01956, 2020.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In Proceedings of the 2016 ACM conference on innovations in theoretical computer science, pages 111–122, 2016.

California Health and Human Services Agency. Cal hospital compare announces honor roll hospitals. Press Release, 2022. URL https://www.chhs.ca.gov/blog/2022/08/22/cal-hospital-compare-announces-honor-roll-hospitals/.

Miguel A Hernan and James M Robins. Causal Inference: What If. CRC Press, 2020.

Keisuke Hirano and Jack R Porter. Asymptotics for statistical treatment rules. Econometrica, 77(5):1683–1701, 2009.

Bengt Holmstrom and Paul Milgrom. Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. The Journal of Law, Economics, and Organization, 7(special_issue):24–52, 1991.

Keith Hoskin. The 'awful idea of accountability': inscribing people into the measurement of objects. Accountability: Power, Ethos and the Technologies of Managing, 265, 1996.

Wenke Hwang, Jordan Derk, Michelle LaClair, and Harold Paz. Hospital patient safety grades may misrepresent hospital performance. Journal of Hospital Medicine, 9(2):111–115, 2014.

Physicians Advocacy Institute. 2022 merit-based incentive payment system (mips) scoring overview. Medicare Quality Payment Program (QPP) Physician Education Initiative, 2022. URL http://www.physiciansadvocacyinstitute.org/Portals/0/assets/docs/MIPS-Pathway/MIPS%20Scoring%20Overview.pdf.

International Stroke Trial Collaborative Group. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. The Lancet, 349(9065):1569–1581, 1997.

Abigail Z Jacobs and Hanna Wallach. Measurement and fairness. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pages 375–385, 2021.

A. Cecile J.W. Janssens. Validity of polygenic risk scores: are we measuring what we think we are? Human molecular genetics, 28(R2):R143–R150, 2019.

Ashish K Jha, E John Orav, Abigail B Ridgway, Jie Zheng, and Arnold M Epstein. Does the leapfrog program help identify high-quality hospitals? The Joint Commission Journal on Quality and Patient Safety, 34(6):318–325, 2008.

HJS Jones and L de Cossart. Risk scoring in surgical patients. British Journal of Surgery, 86(2):149–157, 1999.

Haytham MA Kaafarani, Ann M Borzecki, Kamal MF Itani, Susan Loveland, Hillary J Mull, Kathleen Hickson, Sally MacDonald, Marlena Shin, and Amy K Rosen. Validity of selected patient safety indicators: opportunities and concerns. Journal of the American College of Surgeons, 212(6):924–934, 2011.

Nathan Kallus and Angela Zhou. Confounding-robust policy improvement. Advances in neural information processing systems, 31, 2018.

Edward H Kennedy. Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint arXiv:2203.06469, 2022.

Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? ACM Transactions on Economics and Computation (TEAC), 8(4):1–23, 2020.

Daniel Koretz. The Testing Charade: Pretending to Make Schools Better. The University of Chicago Press, 2017.

Natalie Lacireno-Paquet, Thomas T Holyoke, Michele Moser, and Jeffrey R Henig. Creaming versus cropping: Charter school enrollment practices in response to market incentives. *Educational Evaluation and Policy Analysis*, 24(2):145–158, 2002.

Jean-Jacques Laffont and David Martimort. *The Theory of Incentives*. Princeton University Press, 2009.

Sergio G Lazzarini, Sandro Cabral, Sergio Firpo, and Thomaz Teodorovicz. Why are counterfactual assessment methods not widespread in outcome-based contracts? a formal model approach. *Journal of Public Administration Research and Theory*, 32(3): 509–523, 2022.

Lydia T Liu, Serena Wang, Tolani Britton, and Rediet Abebe. Reimagining the machine learning life cycle to improve educational outcomes of students. *Proceedings of the National Academy of Sciences*, 120 (9):e2204781120, 2023.

Charles F Manski. *Identification for Prediction and Decision*. Harvard University Press, 2008.

Steffen Mau. *The Metric Society: On the Quantification of the Social*. John Wiley & Sons, 2019.

Sean M McNee, John Riedl, and Joseph A Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1097–1101, 2006.

Paul Robert Milgrom and John Roberts. *Economics, Organization, and Management*. Prentice Hall, 1992.

John Miller, Smitha Milli, and Moritz Hardt. Strategic classification is causal modeling in disguise. In *International Conference on Machine Learning*, pages 6917–6926. PMLR, 2020.

Kenneth J Mukamal. The effects of smoking and drinking on cardiovascular disease and risk factors. *Alcohol Research & Health*, 29(3):199, 2006.

Jerry Z Muller. *The Tyranny of Metrics*. Princeton University Press, 2019.

Jersey Neyman. Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, 10(1):1–51, 1923.

Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

Art B. Owen. *Monte Carlo theory, methods and examples*. 2013.

Judea Pearl. On a class of bias-amplifying variables that endanger effect estimates. *arXiv preprint arXiv:1203.3503*, 2012.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830, 2011.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.

Michael Power. *The audit explosion*. Number 7. Demos, 1994.

QCOR. 2023 OPO Interim Annual Public Aggregated Performance Report: Methodology for Donation and Transplant Measure Calculations. *Survey and Certification Quality, Certification and Oversight Reports (S&C QCOR)*, 2023. URL https://qcor.cms.gov/documents/2023%20OPO%20Interim%20Annual%20Public%20Aggregated%20Report.xlsx.

Jennifer L Rodgers, Jarrod Jones, Samuel I Bolleddu, Sahit Vanthenapalli, Lydia E Rodgers, Kinjal Shah, Krishna Karia, and Siva K Panguluri. Cardiovascular risks associated with gender and aging. *Journal of cardiovascular development and disease*, 6(2):19, 2019.

Tina Rosenberg. To make hospitals less deadly, a dose of data. *The New York Times*, 2013. URL https://archive.nytimes.com/opinionator.blogs.nytimes.com/2013/12/04/to-make-hospitals-less-deadly-a-dose-of-data/.

Richard Rothstein. Holding accountability to account: How scholarship and experience in other fields inform exploration of performance incentives in education. working paper 2008-04. *National Center on Performance Incentives*, 2008.

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

Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593, 1980.

Andrew M Ryan, James F Burgess Jr, Christopher P Tompkins, and Stanley S Wallack. The relationship between medicare's process of care quality measures and mortality. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 46 (3):274–290, 2009.

Roshni Sahoo and Stefan Wager. Policy learning with competing agents. *arXiv preprint arXiv:2204.01884*, 2022.

Yonadav Shavit, Benjamin Edelman, and Brian Axelrod. Learning from strategic agents: Accuracy, improvement, and causality. In *International Conference on Machine Learning*. PMLR, 2020.

Shawna N Smith, Heidi Reichert, Jessica Ameling, and Jennifer Meddings. Dissecting leapfrog: how well do leapfrog safe practices scores correlate with hospital compare ratings and penalties, and how much do they matter? *Medical Care*, 55(6):606, 2017.

Jann Spiess. Optimal estimation when researcher and social preferences are misaligned. Working paper, 2018. URL https://gsb-faculty.stanford.edu/jann-spiess/files/2021/01/alignedestimation.pdf.

Jörg Stoye. Minimax regret treatment choice with finite samples. *Journal of Econometrics*, 151(1):70–81, 2009.

Marilyn Strathern. 'improving ratings': audit in the british university system. *European Review*, 5(3):305–321, 1997.

Lisa M Sullivan, Joseph M Massaro, and Ralph B D'Agostino Sr. Presentation of multivariate data for clinical use: The framingham study risk score functions. *Statistics in medicine*, 23(10):1631–1660, 2004.

Aleksey Tetenov. An economic theory of statistical testing. Technical report, CeMMAP working paper, 2016.

Rishi K Wadhera, Jose F Figueroa, Karen E Joynt Maddox, Lisa S Rosenbaum, Dhruv S Kazi, and Robert W Yeh. Quality measure development and associated spending by the centers for medicare & medicaid services. *JAMA*, 323(16):1614–1616, 2020.

Steve Yadlowsky, Hongseok Namkoong, Sanjay Basu, John Duchi, and Lu Tian. Bounds on the conditional and average treatment effect with unobserved confounding factors. *The Annals of Statistics*, 50(5):2587–2615, 2022.

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes, in supplemental]

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [Yes]

   (b) Complete proofs of all theoretical results. [Yes]

   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes, in supplemental]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Yes]

   (b) The license information of the assets, if applicable. [Yes]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Materials for Operationalizing Counterfactual Metrics: Incentives, Ranking, and Information Asymmetry

A ESTIMATING THE UNTREATED POTENTIAL OUTCOME

Our main paper considers incentive problems under the assumption that the principal has access to measures of patient risk through an unbiased estimator of the untreated potential outcome, $\hat{\mu}_0(x)$ with $E[\hat{\mu}_0(x)] = \mu_0(x)$.

Obtaining such an unbiased estimator can be difficult. Still, the conditions under which causal estimation can be done are well understood. In this section, we give some examples of sufficient conditions for identification of $\mu_0(x)$. This is not an exhaustive list, but rather a starting point for analysis of existing data sources.

A.1 Examples of Data Sources

We list two examples of data sources that may contribute to building the estimator $\hat{\mu}_0$, along with sufficient conditions for identification. These may not always be realistic, and are also not the only possible conditions for identification. In reality, practitioners may also be able to leverage knowledge of a more detailed structure causal model, or functional form assumptions (see Hernan and Robins (2020) for a more thorough coverage of methodologies and assumptions).

1. **Auxiliary untreated data.** Suppose the principal has access to an auxiliary dataset $\{X'_j, T'_j, Y'_j\}_{j=1}^m$. Then the principal may produce an unbiased estimator $\hat{\mu}_0(x)$ for $\mu_0(x)$ from this auxiliary dataset if:

    (a) $\mu'_0(x) = E[Y'_j(0)|X'_j = x]$ is identifiable from this dataset. A set of sufficient conditions for this would be if ignorability was satisfied, $\{Y'_j(0), Y'_j(1)\} \perp\!\!\!\perp T'_j|X'_j$, and $P(T'_j = 0|X_j = x) > 0 \ \forall x \in \mathcal{X}$, and SUTVA was satisfied, $Y'_j = Y'_j(T'_j)$ (encompassing both consistency and non-interference).

    (b) The relationship between the untreated potential outcome $Y'_j(0)$ and the covariates X'_j is the same as in the treatment population between $Y_i(0)$ and X_i. That is, $E[Y'_j(0)|X'_j = x] = E[Y_i(0)|X_i = x]$.

    (c) The support of X'_j covers the support of X_i. Formally, the distribution of X_i is absolutely continuous with respect to the distribution of X'_j.

2. **Untreated patients in the treatment population.** In the main paper, we have assumed that the principal only observed outcomes Y_i for treated units with $T_i^\pi = 1$. If the principal also observed a fraction of outcomes for untreated units, with $T_i^\pi = 0$, then these may also be incorporated into the estimate for $\hat{\mu}_0(x)$. Sufficient conditions for identifying $\mu_0(x)$ from a dataset of the agent's untreated units are:

    (a) No information asymmetry: the agent's treatment policy π only depends on X_i. Under this assumption, we have no confounding: $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i^\pi|X_i$.

    (b) Positivity: $P(T_i^\pi = 0|X_i = x) > 0$ for all $x \in \mathcal{X}$, and Y_i is observed for a nonzero proportion of untreated units for all x.

    The positivity assumption is particularly tricky. In the principal-agent game, there is no guarantee that the agent's best response policy π would satisfy positivity. Positivity may be enforced by restricting the treatment rule class Π to only include treatment rules where $\pi(x) < 1$, but this may not always be possible or ethical from a policy standpoint. In practice, it may be possible to create a composite dataset that combines data from both sources. That way, data from untreated patients in the treatment population could supplement auxiliary untreated data to produce a better estimator $\hat{\mu}_0(x)$ than using auxiliary untreated data alone.

A.2 Necessity of Estimating the Untreated Potential Outcome

Here we address the necessity of including an estimate of the untreated potential outcome in the reward function w. We show that if w is completely unable to differentiate between different distributions over the untreated potential outcome, perhaps through incorporating some estimate of some function of $Y_i(0)$ or through other constraints (e.g. co-monotonicity of $\mu_1(x)$ and $\mu_0(x)$), then regret is necessarily unbounded.

Formally, let $\mathcal{D}$ denote a distribution over $X_i, Y_i(1), Y_i(0)$. Let $\mathcal{D}'$ denote an alternate distribution such that $E_{\mathcal{D}'}[Y_i(0)|X_i = x] = \mu_0'(x)$, and the joint distribution $X_i, Y_i(1)$ under $\mathcal{D}'$ remains unchanged. Let Y_i continue to denote the observed outcome under treatment assignments $Y_i = Y_i(T_i^\pi)$, and note that the distribution of Y_i changes under $\mathcal{D}'$ as well. We show that if the reward function w cannot differentiate between worlds $\mathcal{D}$ and $\mathcal{D}'$, then regret will necessarily be unbounded.

Assumption 3 (Non-degenerate w). Assume that w does not always induce a degenerate best response: that is, there exists a distribution $\mathcal{D}$ such that $\arg\max_{\pi \in \Pi} E_{\mathcal{D}}[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$ contains π^w with $E_{\mathcal{D}}[\pi^w(X_i)] > 0$ (in other words, there exists a non-measure-zero set $\bar{\mathcal{X}}$ with $\pi^w(\bar{x}) > 0 \ \forall \bar{x} \in \bar{\mathcal{X}}$).

Theorem 4. *Suppose w satisfies Assumption 3. Let $\mathcal{S}$ be the set of all pairs of distributions $\mathcal{D}, \mathcal{D}'$ such that the distributions of $X_i, Y_i(1)$ remain unchanged between $\mathcal{D}$ and $\mathcal{D}'$. If for all pairs of distributions $\mathcal{D}, \mathcal{D}' \in \mathcal{S}$, $E_{\mathcal{D}}[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = E_{\mathcal{D}'}[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$ for all π, then regret is unbounded.*

Proof. Let $\pi' \in \arg\max_{\pi \in \Pi} E_{\mathcal{D}'}[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$, with $\pi'(\bar{x}) = \delta > 0$ for all $\bar{x} \in \bar{\mathcal{X}}$, with $\bar{\mathcal{X}}$ being a set of measure $\rho > 0$ (which exists by Assumption 3). Let $\mu_1(\bar{x}) = E[Y_i(1)|X = \bar{x}] = \alpha$ for all $\bar{x} \in \bar{\mathcal{X}}$. Let $\mathcal{D}$ be distribution such that $\mu_0(\bar{x}) = E_{\mathcal{D}}[Y_i(0)|X = \bar{x}] = \beta$ for all $\bar{x} \in \bar{\mathcal{X}}$. Since $E_{\mathcal{D}}[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = E_{\mathcal{D}'}[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$ for all π, we have $\pi' \in \arg\max_{\pi \in \Pi} E_{\mathcal{D}}[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$ as a candidate best response to w under distribution $\mathcal{D}$.

Denote the welfare maximizing policy for distribution $\mathcal{D}$ as

$$\pi^* \in \arg\max_{\pi \in \Pi} E_{\mathcal{D}}[\pi(X_i)(\mu_1(X_i) - \mu_0(X_i))].$$

Choose $\beta > \alpha$. Then the set of welfare maximizing policies contains a π^* with $\pi^*(\bar{x}) = 0$ for all $\bar{x} \in \bar{\mathcal{X}}$.

The regret is then

$$R(\pi') = E_{\mathcal{D}}[(\pi^*(X_i) - \pi'(X_i))(\mu_1(X_i) - \mu_0(X_i))] \geq -\delta(\alpha - \beta)\rho.$$

Fixing $\delta > 0, \rho > 0, \alpha < \infty$ and choosing β to be arbitrarily high results in arbitrarily high regret. $\square$

B PROOFS FROM SECTION 3 ON REWARD FUNCTION COMPARISONS

This section provides proofs and formal statements for the best responses and regrets corresponding to each reward function in Section 3.

B.1 Notation

Let $\int f(Z)dP_Z$ denote integration of the function $f(Z)$ with respect to the probability measure for the random variable Z. Let the bold variable $\mathbf{Z}$ denote the vector of i.i.d. random variables $\{Z_i\}_{i=1}^n$. Let $\mathbb{1}(\cdot)$ denote an indicator function with

$$\mathbb{1}(x \in S) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise.} \end{cases}$$

Let π^w denote the agent's best response for the reward function w:

$$\pi^w = \arg\max_{\pi \in \Pi} E[w(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})].$$

B.2 Additional Reward Function

A simple extension from w_{ATO} would be to measure the *total treated outcome (TO)*:

Reward Function 5 (TO):

$$w_{\text{TO}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y}) = \sum_{i=1}^n Y_i T_i^\pi. \tag{1}$$

The unconstrained best response is $\pi^{w_{\text{TO}}}(x) = \mathbb{1}(\mu_1(x) > 0)$. Like w_{ATO}, this incurs unbounded regret; however, this rule treats more people due to treating all individuals for whom the treated outcome is positive. Thus, there is no longer the "creaming" Muller (2019) issue of maximizing the *average* effect at the expense of the *total* effect.

B.3 Formal Statements and Proofs for Agent Best Responses

We give formal statements and proofs for the agent best responses to the different reward functions stated in Section 3.

Proposition 6 (ATO Best Response). Suppose Π is the set of all functions $\pi : \mathcal{X} \to [0, 1]$, and suppose X is a discrete random variable supported on $\mathcal{X}$. Then the agent's best response to w_{ATO} is:

$$\pi^{w_{\text{ATO}}}(x) = \begin{cases} 1 & \text{if } x \in \arg\max_x \mu_1(x) \text{ and } \mu_1(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. By the tower property,

$$E[w_{\text{ATO}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = E[E[Y_i | T_i^\pi = 1, X_i] | T_i^\pi = 1] = \int E[Y_i | T_i^\pi = 1, X_i] dP_{X_i | T_i^\pi = 1}.$$

Since T_i^π only depends on X_i by construction, we have the *strong ignorability* property that $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i^\pi | X_i$. Therefore,

$$\int E[Y_i | T_i^\pi = 1, X_i] dP_{X_i | T_i^\pi = 1} = \int \mu_1(X_i) dP_{X_i | T_i^\pi = 1}.$$

For discrete X_i, this is equal to

$$\sum_{x \in \mathcal{X}} \mu_1(x) P(X_i = x | T_i^\pi = 1) = \sum_{x \in \mathcal{X}} \mu_1(x) \frac{\pi(x) P(X_i = x)}{\sum_{z \in \mathcal{X}} \pi(z) P(X_i = z)}.$$

The π function that maximizes this is exactly that given in Proposition 6. $\qquad\square$

As a tool to prove further results, we give the following lemma for the optimal treatment rule.

Lemma 1 (Optimal Treatment Rule). If the agent's treatment rule π depends only on X_i, then the treatment rule π^* that maximizes $V(\pi)$ is

$$\pi^*(x) = \begin{cases} 1 & \text{if } \tau(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof.

$$\begin{aligned} V(\pi) &= E[Y_i(T_i^\pi) - Y_i(0)] \\ &= E[Y_i(T_i^\pi) - Y_i(0) | T_i^\pi = 1]P(T_i^\pi = 1) + E[Y_i(T_i^\pi) - Y_i(0) | T_i^\pi = 0]P(T_i^\pi = 0) \\ &= E[Y_i(1) - Y_i(0) | T_i^\pi = 1]P(T_i^\pi = 1). \end{aligned}$$

By the tower property,

$$E[Y_i(1) - Y_i(0) | T_i^\pi = 1] = E[E[Y_i(1) - Y_i(0) | T_i^\pi = 1, X_i] | T_i^\pi = 1].$$

Since we have ignorability in X_i,

$$E[Y_i(1) - Y_i(0) | T_i^\pi = 1, X_i] = E[Y_i(1) - Y_i(0) | X_i] = \tau(X_i).$$

Thus,

$$V(\pi) = E[\tau(X_i)|T_i^\pi = 1]P(T_i^\pi = 1) = \int \tau(X_i)P(T_i^\pi = 1)dP_{X_i|T_i^\pi=1}.$$

By Bayes' theorem,

$$\int \tau(X_i)P(T_i^\pi = 1)dP_{X_i|T_i^\pi=1} = \int \tau(X_i)P(T_i^\pi = 1|X_i)dP_{X_i} = \int \tau(X_i)\pi(X_i)dP_{X_i}.$$

The π function that maximizes this is exactly as given in Lemma 1. □

Proposition 7 (ATT Best Response). Suppose Π is the set of all functions $\pi : \mathcal{X} \to [0,1]$, and suppose X is a discrete random variable supported on $\mathcal{X}$. Then the agent's best response to w_{ATT} is:

$$\pi^{w_{\mathrm{ATT}}}(x) = \begin{cases} 1 & \text{if } x \in \arg\max_x \tau(x) \text{ and } \tau(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof.

$$E[w_{\mathrm{ATT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = E[Y_i - \hat{\mu}_0(X_i)|T_i^\pi = 1]$$
$$= E[Y_i|T_i^\pi = 1] - E[E[\hat{\mu}_0(X_i)|T_i^\pi = 1, X_i]|T_i^\pi = 1].$$

Since $E[\hat{\mu}_0(x)] = \mu_0(x)$ for all $x \in \mathcal{X}$, we have $E[\hat{\mu}_0(X_i)|T_i^\pi = 1, X_i] = \mu_0(X_i)$. Thus,

$$E[w_{\mathrm{ATT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = E[Y_i|T_i^\pi = 1] - E[\mu_0(X_i)|T_i^\pi = 1].$$

Since we have ignorability in X_i,

$$E[Y_i|T_i^\pi = 1] = E[E[Y_i|T_i^\pi = 1, X_i]|T_i^\pi = 1] = E[E[Y_i(1)|X_i]|T_i^\pi = 1] = E[\mu_1(X_i)|T_i^\pi = 1].$$

Combining these,

$$E[w_{\mathrm{ATT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = E[\mu_1(X_i) - \mu_0(X_i)|T_i^\pi = 1] = E[\tau(X_i)|T_i^\pi = 1].$$

For discrete X_i, this is equal to

$$\sum_{x \in \mathcal{X}} \tau(x)P(X_i = x|T_i^\pi = 1) = \sum_{x \in \mathcal{X}} \tau(x)\frac{\pi(x)P(X_i = x)}{\sum_{z \in \mathcal{X}} \pi(z)P(X_i = z)}.$$

The π function that maximizes this is exactly that given in Proposition 7. □

Proposition 8 (TO Best Response). Suppose Π is the set of all functions $\pi : \mathcal{X} \to [0,1]$. Then the agent's best response to w_{TO} is:

$$\pi^{w_{\mathrm{TO}}}(x) = \begin{cases} 1 & \text{if } \mu_1(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof.

$$E[w_{\mathrm{TO}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = nE[Y_i|T_i^\pi = 1]P(T_i^\pi = 1).$$

Since we have ignorability in X_i, we have $E[Y_i|T_i^\pi = 1] = E[\mu_1(X_i)|T_i^\pi = 1]$ (shown in more detail in the above proofs). Thus,

$$E[w_{\mathrm{TO}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = nE[\mu_1(X_i)|T_i^\pi = 1]P(T_i^\pi = 1) = n\int \mu_1(X_i)P(T_i^\pi = 1)dP_{X_i|T_i^\pi=1}.$$

By Bayes' theorem,

$$n\int \mu_1(X_i)P(T_i^\pi = 1)dP_{X_i|T_i^\pi=1} = n\int \mu_1(X_i)P(T_i^\pi = 1|X_i)dP_{X_i} = n\int \mu_1(X_i)\pi(X_i)dP_{X_i}.$$

The π function that maximizes this is exactly that given in Proposition 8. □

Proposition 9 (TT Best Response). Suppose Π is the set of all functions $\pi : \mathcal{X} \to [0,1]$. Then the agent's best response to w_{TT} is:

$$\pi^{w_{\mathrm{TT}}}(x) = \begin{cases} 1 & \text{if } \tau(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof.

$$E[w_{\mathrm{TT}}(\mathbf{X}, \mathbf{T}^{\pi}, \mathbf{Y})] = nE[Y_i - \hat{\mu}_0(X_i) | T_i^{\pi} = 1]P(T_i^{\pi} = 1).$$

Since we have ignorability in X_i, we have $E[Y_i | T_i^{\pi} = 1] = E[\mu_1(X_i) | T_i^{\pi} = 1]$. (shown in more detail in the above proofs). Since $E[\hat{\mu}_0(x)] = \mu_0(x)$ for all $x \in \mathcal{X}$, we have $E[\hat{\mu}_0(X_i) | T_i^{\pi} = 1] = E[\mu_0(X_i) | T_i^{\pi} = 1]$. Thus,

$$E[w_{\mathrm{TT}}(\mathbf{X}, \mathbf{T}^{\pi}, \mathbf{Y})] = nE[\tau(X_i) | T_i^{\pi} = 1]P(T_i^{\pi} = 1) = nV(\pi).$$

Applying Lemma 1, the π function that maximizes this is exactly that given in Proposition 9. $\qquad\square$

B.4 Regret proofs

We give proofs for the regret bounds stated for each reward function in Section 3.

Proposition 1 (ATO Regret). *If the conditional mean untreated potential outcomes $\mu_0(x)$ are unbounded, then the regret for the reward function w_{ATO} may be arbitrarily large.*

Proof. We construct an family of distributions for which the regret is unbounded. Let $\mathcal{X} = \{0,1\}$ with $P(X_i = 1) = p$, and let $\mu_1(0) = 0$, $\mu_1(1) = 1$. Let $\mu_0(0) = \alpha$, and $\mu_0(1) = \beta$. Suppose Π is the set of all functions $\pi : \mathcal{X} \to [0,1]$.

For w_{ATO}, the agent's best response is $\pi^{w_{\mathrm{ATO}}}(0) = 0$ and $\pi^{w_{\mathrm{ATO}}}(1) = 1$.

We illustrate two failure modes for the reward function w_{ATO}. First, suppose $\alpha < 0$, and $\beta = 0$. Then the regret is given by

$$R(\pi^w) = \max_{\pi \in \Pi} V(\pi) - V(\pi^w) = -\alpha(1-p) + p - p = -\alpha(1-p).$$

This is unbounded for unbounded α. Intuitively, this illustrates an example where the agent ignores the higher treatment effect of the patients with a lower treated outcome, such as sicker patients with higher mortality probability but more benefit from surgery. Since the agent's best response does not account for the patient's untreated potential outcome, the agent thus ignores the fact that sicker patients would otherwise have very poor outcomes without treatment. This matches the findings from Dranove et al. (2003).

A second failure mode would be if $\alpha = -1$, and $\beta > 1$. Then the regret is given by

$$R(\pi^w) = (1-p) - p(\beta - 1) = 1 - p\beta.$$

This is also unbounded for unbounded β. This illustrates an example where the agent treats the patients with a higher treated outcome, even though treatment actually harms those patients, such as healthier patients who might incur more risks or side effects from treatment. $\qquad\square$

Proposition 2 (ATT Regret). *Suppose Π is the set of all functions $\pi : \mathcal{X} \to [0,1]$. Then the regret for the reward function w_{ATT} is upper bounded by*

$$R(\pi^{w_{ATT}}) \le \max_{\pi \in \Pi} V(\pi).$$

Proof. Let $\pi^* = \arg\max_{\pi \in \Pi} V(\pi)$. Then

$$\pi^*(x) = \begin{cases} 1 & \text{if } \tau(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$
\begin{aligned}
R(\pi^{w_{\mathrm{ATT}}}) &= V(\pi^*) - V(\pi^{w_{\mathrm{ATT}}}) \\
&= E[\pi^*(X_i)\tau(X_i)] - E[\pi^{w_{\mathrm{ATT}}}(X_i)\tau(X_i)] \\
&= E[\tau(X_i)\,\mathbb{1}(\tau(X_i) > 0)] - E[\tau(X_i)\,\mathbb{1}(\tau(X_i) > 0 \cap X_i \in \arg\max_{x \in \mathcal{X}} \tau(x))] \\
&= E[\tau(X_i)\,\mathbb{1}(\tau(X_i) > 0 \cap X_i \notin \arg\max_{x \in \mathcal{X}} \tau(x))]
\end{aligned}
$$

$\square$

Proposition 3 (TT Regret). *If $E[\hat{\mu}_0(X)] = \mu_0(X)$, then the regret from applying the reward function w_{TT} is $R(\pi^{w_{TT}}) = 0$.*

Proof. As shown in Proposition 9,

$$
E[w_{\mathrm{TT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = nE[\tau(X_i)|T_i^\pi = 1]P(T_i^\pi = 1) = nV(\pi).
$$

Therefore, the agent's best response $\pi^{w_{\mathrm{TT}}}$ also maximizes $V(\pi)$ over the same feasible set, Π. $\square$

C PROOFS FROM SECTION 4 ON RANKING

We give expanded notation and proofs for the results in Section 4 on modifications of w_{TT} to satisfy ranking desiderata.

C.1 Detailed notation for ranking with multiple agents

We expand the notation for ranking with multiple agents, as applied in Section 4. Suppose there are K agents, where each agent k observes its own sample of n_k patients with covariates $\mathbf{X}^{(k)} = \{X_i^{(k)}\}_{i=1}^{n_k}$ drawn i.i.d. from distribution $P_{X^{(k)}}$ with support $\mathcal{X}$. Let $Y_i^{(k)}(t)$ denote the potential outcomes when agent k treats the patient with treatment t. Let $\mu_t^{(k)}(x) = E[Y_i^{(k)}(t)|X_i^{(k)} = x]$, and $\tau^{(k)}(x) = E[Y_i^{(k)}(1) - Y_i^{(k)}(0)|X_i^{(k)} = x]$.

For rankings to be meaningful, we assume that if the same patient with covariate value $x \in \mathcal{X}$ were to be treated by either agent j or agent k, their potential outcomes would follow each agent's respective conditional potential outcome distributions for covariate value x. Furthermore, the conditional potential untreated outcome has the same distribution for all k: for each $x \in \mathcal{X}$, the distributions $P_{Y^{(k)}(0)|X^{(k)}=x}$ are identical for all $k \in \{1, ..., K\}$.

Let $\mu_0(x)$ denote the shared mean conditional untreated potential outcome, with $\mu_0^{(k)}(x) = \mu_0(x)$ for all k. In short, one provider not treating a patient is equivalent to another provider not treating the same patient.

Suppose agent k chooses treatment policy π_k, thus producing realized treatments denoted $\mathbf{T}^{\pi_k} = \{T_i^{\pi_k}\}_{i=1}^{n_k}$ and outcomes $Y_i^{(k)} = Y_i^{(k)}(T_i^{\pi_k})$, $\mathbf{Y}^{(k)} = \{Y_i^{(k)}\}_{i=1}^{n_k}$.

Suppose the reward function w is used to rank these K agents in the following way: the principal publishes score functions $w_k : \mathcal{X}^{n_k} \times \{0,1\}^{n_k} \times \mathbb{R}^{n_k} \to \mathbb{R}$, and each agent k gets score $w_k(\mathbf{X}^{(k)}, \mathbf{T}^{\pi_k}, \mathbf{Y}^{(k)})$ after choosing their treatment policy π_k. The agents are then ranked from highest to lowest score function values.

We assume that each agent seeks to maximize their individual ranking, and make the simplifying assumption that the agents act independently: that is, each agent does not consider the potential actions of other agents when choosing their actions. This may be realistic in a setting with a large number of hospitals serving more-or-less independent populations, though more complex competitive multi-agent models may make for interesting future extensions.

C.2 Proofs

We give proofs for the proposition and theorems from Section 4.

Proposition 4. *If*

$$
w_k(\mathbf{X}^{(k)}, \mathbf{T}^{\pi_k}, \mathbf{Y}^{(k)}) = \sum_{i=1}^{n_k} (Y_i^{(k)} - \hat{\mu}_0(X_i^{(k)}))T_i^{\pi_k},
$$

then both ranking properties in Definitions 1 and 2 will be violated.

Proof. Suppose for agents j and k, Π_j and Π_k are both unconstrained. For w_k as defined above,

$$\max_{\pi_j \in \Pi_j} E[w_j(\mathbf{X}^{(j)}, \mathbf{T}^{\pi_j}, \mathbf{Y}^{(k)})] = n_j E[\tau^{(j)}(X^{(j)}) \mathbb{1}(\tau^{(j)}(X^{(j)}) > 0)],$$

$$\max_{\pi_k \in \Pi_k} E[w_k(\mathbf{X}^{(k)}, \mathbf{T}^{\pi_k}, \mathbf{Y}^{(k)})] = n_k E[\tau^{(k)}(X^{(k)}) \mathbb{1}(\tau^{(k)}(X^{(k)}) > 0)].$$

For Definition 1, suppose $\tau^{(j)}(x) \geq \tau^{(k)}(x)$ for all $x \in \mathcal{X}$, and $X^{(j)}$ and $X^{(k)}$ are identically distributed. Then,

$$E[\tau^{(j)}(X^{(j)}) \mathbb{1}(\tau^{(j)}(X^{(j)}) > 0)] \geq E[\tau^{(k)}(X^{(k)}) \mathbb{1}(\tau^{(k)}(X^{(k)}) > 0)].$$

Let n_j, n_k be a pair such that

$$n_j < n_k \frac{E[\tau^{(k)}(X^{(k)}) \mathbb{1}(\tau^{(k)}(X^{(k)}) > 0)]}{E[\tau^{(j)}(X^{(j)}) \mathbb{1}(\tau^{(j)}(X^{(j)}) > 0)]}.$$

This immediately results in Definition 1 being violated.

Definition 2 is also violated, since $\tau^{(j)}(x) \geq \tau^{(k)}(x)$ for all $x \in \mathcal{X}$ implies that $E[\tau^{(j)}(X_0)] \geq E[\tau^{(k)}(X_0)]$ for any reference population P_{X_0}. However, for the above n_j, n_k, we've shown that

$$\max_{\pi_j \in \Pi_j} E[w_j(\mathbf{X}^{(j)}, \mathbf{T}^{\pi_j}, \mathbf{Y}^{(k)})] < \max_{\pi_k \in \Pi_k} E[w_k(\mathbf{X}^{(k)}, \mathbf{T}^{\pi_k}, \mathbf{Y}^{(k)})].$$

$\square$

Theorem 1 (Incentive Alignment). *Suppose $E[\hat{\mu}_0(x)] = \mu_0(x)$, and suppose Π is the set of all functions $\pi : \mathcal{X} \to [0,1]$. For any function $g : \mathcal{X} \to \mathbb{R}^+$, w_{TT}^g yields an agent best response with zero regret.*

Proof. By the tower property,

$$E[w_{\mathrm{TT}}^g(\mathbf{Y}, \mathbf{T}^\pi, \mathbf{X})] = nP(T_i^\pi = 1)E[E[(Y_i - \hat{\mu}_0(X_i))g(X_i)|T_i^\pi = 1, X_i]|T_i^\pi = 1]$$
$$= nP(T_i^\pi = 1)E[g(X_i)E[Y_i - \hat{\mu}_0(X_i)|T_i^\pi = 1, X_i]|T_i^\pi = 1].$$

Since $E[\hat{\mu}_0(x)] = \mu_0(x)$ and we have ignorability in X_i,

$$E[Y_i - \hat{\mu}_0(X_i)|T_i^\pi = 1, X_i] = \tau(X_i).$$

This is shown in more detail in the proof for Proposition 9.

Combining this with the above, we have

$$E[w_{\mathrm{TT}}^g(\mathbf{Y}, \mathbf{T}^\pi, \mathbf{X})] = nP(T_i^\pi = 1)E[g(X_i)\tau(X_i)|T_i^\pi = 1].$$

Applying Bayes' theorem as in Lemma 1,

$$P(T_i^\pi = 1)E[g(X_i)\tau(X_i)|T_i^\pi = 1] = \int g(X_i)\tau(X_i)P(T_i^\pi = 1)dP_{X_i|T_i^\pi=1}$$
$$= \int g(X_i)\tau(X_i)P(T_i^\pi = 1|X_i)dP_{X_i}$$
$$= \int g(X_i)\tau(X_i)\pi(X_i)dP_{X_i}.$$

Therefore,

$$E[w_{\mathrm{TT}}^g(\mathbf{Y}, \mathbf{T}^\pi, \mathbf{X})] = nE[g(X_i)\tau(X_i)\pi(X_i)].$$

Since $g(X_i) > 0$, the treatment rule π that maximizes this is the same as π^* from Lemma 1. Therefore, the regret is zero. $\square$

Theorem 2 (Ranking Desiderata Satisfied). *Let $P_{X^{(k)}}$ be absolutely continuous with respect to P_{X_0}, and let $g_k = \frac{1}{n_k}\frac{dP_{X_0}}{dP_{X^{(k)}}}$ be the normalized Radon–Nikodym derivative of the reference distribution P_{X_0} with respect to agent k's covariate distribution $P_{X^{(k)}}$. Then*

$$w_k(\mathbf{X}^{(k)}, \mathbf{T}^{\pi_k}, \mathbf{Y}^{(k)}) = \sum_{i=1}^{n_k}(Y_i^{(k)} - \hat{\mu}_0(X_i^{(k)}))T_i^{\pi_k}g_k(X^{(k)}) \tag{2}$$

satisfies both ranking properties in Definitions 1 and 2 as long as Π_k is unconstrained and treatment effects are nonnegative, $\tau^{(k)}(x) \geq 0$, for all $k \in \{1, ..., K\}$.

Proof. As shown in the proof of Theorem 1,

$$E[w_k(\mathbf{X}^{(k)}, \mathbf{T}^{\pi_k}, \mathbf{Y}^{(k)})] = n_k E[\tau^{(k)}(X_i^{(k)})\pi_k(X_i^{(k)})g_k(X_i^{(k)})].$$

With $g_k = \frac{1}{n_k}\frac{dP_{X_0}}{dP_{X^{(k)}}}$, we have

$$n_k E[\tau^{(k)}(X_i^{(k)})\pi_k(X_i^{(k)})g_k(X_i^{(k)})] = E[\tau^{(k)}(X_0)\pi_k(X_0)].$$

Let Π_j, Π_k both be unconstrained. Then

$$\max_{\pi_j \in \Pi_j} E[w_j(\mathbf{X}^{(j)}, \mathbf{T}^{\pi_j}, \mathbf{Y}^{(k)})] = E[\tau^{(j)}(X_0)\,\mathbb{1}(\tau^{(j)}(X_0) > 0)],$$

$$\max_{\pi_k \in \Pi_k} E[w_k(\mathbf{X}^{(k)}, \mathbf{T}^{\pi_k}, \mathbf{Y}^{(k)})] = E[\tau^{(k)}(X_0)\,\mathbb{1}(\tau^{(k)}(X_0) > 0)].$$

Definition 1 is immediately satisfied, since $\tau^{(j)}(x) \geq \tau^{(k)}(x)$ for all $x \in \mathcal{X}$ implies $E[\tau^{(k)}(X_0)\,\mathbb{1}(\tau^{(k)}(X_0) > 0)] \geq E[\tau^{(k)}(X_0)\,\mathbb{1}(\tau^{(k)}(X_0) > 0)]$.

If $\tau^{(k)}(x) \geq 0$ for all $x \in \mathcal{X}$ and all $k \in \{1, ..., K\}$, then Definition 2 is satisfied, since

$$E[\tau^{(j)}(X_0)] \geq E[\tau^{(k)}(X_0)] \implies E[\tau^{(j)}(X_0)\,\mathbb{1}(\tau^{(j)}(X_0) > 0)] \geq E[\tau^{(k)}(X_0)\,\mathbb{1}(\tau^{(k)}(X_0) > 0)].$$

$\square$

D   PROOFS AND ADDITIONAL RESULTS FOR SECTION 5 ON INFORMATION ASYMMETRY

In this section, we give proofs and additional regret bound results for Section 5 on information asymmetry. First, we prove the regret bounds in Section 5 when $\hat{\mu}_0(x)$ is estimated from auxiliary data where the mean untreated potential outcome conditioned on X, $\mu_0(x)$, is identifiable.

Second, we prove similar results when $\hat{\mu}_0(x)$ is estimated from the untreated units in the treatment population, where $T^\pi = 0$. Under information asymmetry, the mean untreated potential outcome conditioned on X, $\mu_0(X)$, is no longer identifiable. Therefore, $\hat{\mu}_0(x)$ will be subject to confounding bias. Still, if we apply a similar but stronger assumption than Assumption 2, we can get similar regret bounds to Theorem 3.

D.1   Detailed notation for information asymmetry

We model information asymmetry in our principal agent game as follows: suppose the agent observes additional covariates per patient $U_i \in \mathcal{U}$, and selects a treatment rule $\pi : \mathcal{X}, \mathcal{U} \to [0, 1]$ from a feasible set of treatment rules Π, with $\pi(X, U) = P(T_i^\pi = 1|X, U)$. Suppose the principal still observes only $\{X_i, T_i^\pi, Y_i\}_{i=1}^n$, and chooses a reward function $w : \mathcal{X}^n \times \{0, 1\}^n \times \mathbb{R}^n \to \mathbb{R}$ with which to reward the agent. Notably, the principal's reward function w cannot depend on U. Let $\mu_t(X, U) = E[Y_i(t)|X, U]$ and $\tau(X, U) = E[Y_i(1) - Y_i(0)|X, U]$.

The utility is still defined as in Section 2, and with the additional U variable can be rewritten as

$$V(\pi) = E[\pi(X_i, U_i)\tau(X_i, U_i)].$$

The regret is also still defined as in Section 2.

D.2 Proofs for regret bounds with unbiased counterfactual estimate

Suppose the principal estimates the mean conditional untreated potential outcome from auxiliary data $\{X'_j, T'_j, Y'_j\}_{j=1}^m$ drawn i.i.d. from auxiliary dataset $\mathcal{Q}$, denoted $\hat{\mu}_0^{\mathcal{Q}}(x)$. Suppose we have a "best case scenario" where the relationship between X'_j and $Y'_j(0)$ is the same in the auxiliary data as the relationship between X_i and $Y_i(0)$ in the treatment population, and the mean untreated potential outcome conditional on X'_j is identifiable from $\mathcal{Q}$, such that

$$E[\hat{\mu}_0^{\mathcal{Q}}(x)] = E[Y'_j(0)|X'_j = x] = E[Y_i(0)|X_i = x].$$

Outside of this ideal setting, any distribution shift or problems with identifiability in the $\mathcal{Q}$ dataset would increase the regret. Proposition 10 below is an intermediate result that does not actually rely on identifiability of $\mu_0(x)$, and may provide a good starting point for future analyses of distribution shift or non-identifiability. To simplify the proofs below, we first give Lemma 2.

Lemma 2. Under information asymmetry,

$$E[w_{\mathrm{TT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = n(E[\pi(X_i, U_i)\mu_1(X_i, U_i)] - E[\pi(X_i, U_i)\hat{\mu}_0(X_i)]).$$

Proof. We have previously shown that

$$\begin{aligned}
E[w_{\mathrm{TT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] &= nP(T_i^\pi = 1)E[Y_i - \hat{\mu}_0(X_i)|T_i^\pi = 1] \\
&= n(P(T_i^\pi = 1)E[Y_i|T_i^\pi = 1] - P(T_i^\pi = 1)E[\hat{\mu}_0(X_i)|T_i^\pi = 1]).
\end{aligned}$$

Considering the first term, since we have ignorability in X_i and U_i,

$$\begin{aligned}
P(T_i^\pi = 1)E[Y_i|T_i^\pi = 1] &= P(T_i^\pi = 1)E[E[Y_i|X_i, U_i, T_i^\pi = 1]|T_i^\pi = 1] \\
&= P(T_i^\pi = 1)E[\mu_1(X_i, U_i)|T_i^\pi = 1] \\
&= P(T_i^\pi = 1) \int \mu_1(X_i, U_i) dP_{X_i, U_i|T_i^\pi=1} \\
&= \int \mu_1(X_i, U_i) P(T_i^\pi = 1|X_i, U_i) dP_{X_i, U_i} \\
&= E[\pi(X_i, U_i)\mu_1(X_i, U_i)].
\end{aligned}$$

For the second term, we again apply Bayes' theorem:

$$\begin{aligned}
P(T_i^\pi = 1)E[\hat{\mu}_0(X_i)|T_i^\pi = 1] &= P(T_i^\pi = 1) \int \hat{\mu}_0(X_i) dP_{X_i, U_i|T_i^\pi=1} \\
&= \int \hat{\mu}_0(X_i) P(T_i^\pi = 1|X_i, U_i) dP_{X_i, U_i} \\
&= E[\pi(X_i, U_i)\hat{\mu}_0(X_i)].
\end{aligned}$$

$\square$

Proposition 10. Suppose the principal applies the reward function w_{TT} with an estimate $\hat{\mu}_0^{\mathcal{Q}}(X)$. Then the regret is bounded by the average bias in the conditional untreated potential outcome estimate.

$$R(\pi^w) \leq 2E[|\hat{\mu}_0^{\mathcal{Q}}(X_i) - \mu_0(X_i, U_i)|]. \tag{3}$$

Proof. Let $\hat{V}(\pi) = \frac{1}{n}E[w_{\mathrm{TT}}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})]$. Then $\pi^{w_{\mathrm{TT}}}$ maximizes $\hat{V}(\pi)$ as well.

$$\begin{aligned}
V(\pi^*) - V(\pi^{w_{\mathrm{TT}}}) &\leq V(\pi^*) - V(\pi^{w_{\mathrm{TT}}}) + \hat{V}(\pi^{w_{\mathrm{TT}}}) - \hat{V}(\pi^*) \\
&\leq |V(\pi^*) - \hat{V}(\pi^*)| + |\hat{V}(\pi^{w_{\mathrm{TT}}}) - V(\pi^{w_{\mathrm{TT}}})| \\
&\leq 2 \max_{\pi \in \Pi} |\hat{V}(\pi) - V(\pi)| \\
&= 2 \max_{\pi \in \Pi} |E[\pi(X_i, U_i)(\hat{\mu}_0^{\mathcal{Q}}(X_i) - \mu_0(X_i, U_i))]|,
\end{aligned}$$

where the last line follows from Lemma 2. By Jensen's inequality,

$$\max_{\pi \in \Pi} |E[\pi(X,U)(\hat{\mu}_0^{\mathcal{Q}}(X_i) - \mu_0(X_i, U_i))]| \leq \max_{\pi \in \Pi} E[\pi(X_i, U_i)|\hat{\mu}_0^{\mathcal{Q}}(X_i) - \mu_0(X_i, U_i)|]$$
$$\leq E[|\hat{\mu}_0^{\mathcal{Q}}(X_i) - \mu_0(X_i, U_i)|].$$

$\square$

Theorem 3 (Regret With Information Asymmetry)**.** *If the principal applies the reward function from w_{TT} with an unbiased estimate $\hat{\mu}_0(X)$ where $E[\hat{\mu}_0(x)] = \mu_0(x)$, then under Assumption 2, the regret is upper bounded as*

$$R(\pi^{w_{TT}}) \leq 2\gamma_{marg}.$$

Proof. This follows similarly to the proof of Proposition 10:

$$V(\pi^*) - V(\pi^{w_{TT}}) \leq V(\pi^*) - V(\pi^{w_{TT}}) + \hat{V}(\pi^{w_{TT}}) - \hat{V}(\pi^*)$$
$$\leq |V(\pi^*) - \hat{V}(\pi^*)| + |\hat{V}(\pi^{w_{TT}}) - V(\pi^{w_{TT}})|$$
$$\leq 2 \max_{\pi \in \Pi} |\hat{V}(\pi) - V(\pi)|$$
$$= 2 \max_{\pi \in \Pi} |E[\pi(X_i, U_i)(\hat{\mu}_0(X_i) - \mu_0(X_i, U_i))]|$$

By the tower property,

$$E[\pi(X,U)\hat{\mu}_0(X)] = E[E[\hat{\mu}_0(X_i)\pi(X_i, U_i)|X_i, U_i]]$$
$$= E[\pi(X_i, U_i)E[\hat{\mu}_0(X_i)|X_i, U_i]]$$
$$= E[\pi(X_i, U_i)\mu_0(X_i)].$$

Therefore,

$$E[\pi(X_i, U_i)(\hat{\mu}_0(X_i) - \mu_0(X_i, U_i))] = E[\pi(X_i, U_i)(\mu_0(X_i) - \mu_0(X_i, U_i))].$$

By Jensen's inequality,

$$\max_{\pi \in \Pi} |E[\pi(X_i, U_i)(\mu_0(X_i) - \mu_0(X_i, U_i))]| \leq \max_{\pi \in \Pi} E[\pi(X_i, U_i)|\mu_0(X_i) - \mu_0(X_i, U_i)|]$$
$$\leq E[|\mu_0(X_i) - \mu_0(X_i, U_i)|].$$

The regret bound then follows directly from Assumption 2 which says that

$$E[|\mu_0(X_i) - \mu_0(X_i, U_i)|] \leq \gamma_{\mathrm{marg}}.$$

$\square$

Proposition 5. *For all $\epsilon > 0$, there exists a distribution of $X_i, U_i, Y_i(0), Y_i(1)$ wherein $R(\pi^{w_{TT}}) \geq \gamma_{marg} - \epsilon$.*

Proof. We construct a family of distributions of $X_i, U_i, Y_i(0), Y_i(1)$ that achieves this regret bound. Let $U \in \{0, 1\}$, with $P(U = 1) = \frac{1}{2}$. Suppose X is entirely uncorrelated with $Y_i(0), Y_i(1)$, such that $E[Y(t)|X] = E[Y(t)] = \mu_t$. For $\alpha > 0, \beta > 0$, let

$$\mu_1(x,u) = \begin{cases} 0 & \text{if } u = 1 \\ \beta & \text{if } u = 0 \end{cases}, \quad \mu_0(x,u) = \begin{cases} -\alpha & \text{if } u = 1 \\ \alpha & \text{if } u = 0 \end{cases},$$

which also means that $\mu_0(x) = \mu_0 = 0$ for all x. Suppose Π includes all functions $\pi : \mathcal{X}, \mathcal{U} \to [0, 1]$. Then by Lemma 2, and assuming $\hat{\mu}_0(x)$ is unbiased, we have

$$E[w_{TT}(\mathbf{X}, \mathbf{T}^\pi, \mathbf{Y})] = n(E[\pi(X_i, U_i)(\mu_1(X_i, U_i) - \mu_0(X_i))]).$$

The resulting policy $\pi^{w_{TT}}$ that maximizes this is

$$\pi^{w_{TT}}(x,u) = \begin{cases} 0 & \text{if } u = 1 \\ 1 & \text{if } u = 0 \end{cases}.$$

Suppose $\beta < \alpha$. Then the optimal policy is

$$\pi^*(x, u) = \begin{cases} 1 & \text{if } u = 1 \\ 0 & \text{if } u = 0 \end{cases}.$$

with utility $V(\pi^*) = \frac{\alpha}{2}$. The regret $R(\pi^{w_{\text{TT}}})$ is then

$$R(\pi^{w_{\text{TT}}}) = \frac{\alpha}{2} - \frac{\beta - \alpha}{2} = \alpha - \frac{\beta}{2}.$$

Note that $\gamma_{\text{marg}} = \alpha$. For any ϵ, choosing $\beta = \epsilon$ gets $R(\pi^{w_{\text{TT}}}) = \gamma_{\text{marg}} - \frac{\epsilon}{2}$, satisfying the bound in Proposition 5. □

D.3 Additional regret bounds when estimating the counterfactual using agent data

Suppose the principal estimates $\hat{\mu}_0(x)$ from the untreated data from the agent's treatment population, i.e. those individuals for whom $T_i^\pi = 0$. As discussed in Sections 5 and A, under full information symmetry, the mean conditional untreated potential outcome is identifiable as long as the agent's treatment rule class Π maintains positivity, $\pi(x) > 0$ for all $x \in \mathcal{X}$. Under information asymmetry, positivity is no longer sufficient for identifiability, as the agent's additional information makes T_i^π depend on U_i, and thus ignorability in X_i is no longer satisfied.

Still, we can analyze what happens when the principal constructs a counterfactual estimator from the agent's untreated outcomes. This estimator depends on the agent's treatment rule π, and unlike in Section A, information asymmetry means that the agent's treatment rule π affects this estimator as well. Let $\hat{\mu}_0^\pi(X) = E[Y_i | T^\pi = 0, X]$ denote the principal's estimate of the mean untreated potential outcome from the agent's data.

As with the auxiliary data, if we assume a bound on how much U_i affects the untreated potential outcome given X_i, we can still bound the regret if the principal were to apply the reward function w_{TT} using $\hat{\mu}_0^\pi(X)$. However, the required assumption is a bit stronger.

Assumption 4. The maximum effect of the unobserved attributes U on the conditional untreated potential outcome is bounded on average. Define the maximum difference in the untreated potential outcome for a given x, u to be

$$\Delta(x, u) = \max_{\tilde{u} \in \mathcal{U}} |\mu_0(x, u) - \mu_0(x, \tilde{u})|.$$

The average difference is bounded as:

$$E[\Delta(X, U)] \leq \gamma_{\text{max}}.$$

Assumption 4 is stronger than Assumption 2 in the sense that Assumption 2 is not sufficient to bound the regret when the principal estimates the mean conditional untreated potential outcome from the agent's data using $\hat{\mu}_0^\pi(X)$. Furthermore, the bound in Assumption 4 implies that Assumption 2 is satisfied with $\gamma_{\text{marg}} \leq \gamma_{\text{max}}$. Intuitively, Assumption 4 is not sufficient to bound the regret when the principal uses $\hat{\mu}_0^\pi(X)$ because the agent can choose π to allocate treatment to single values of $\tilde{u} \in \mathcal{U}$, such that $\hat{\mu}_0^\pi(x)$ ends up matching a single value $\mu_0(x, \tilde{u})$.

More generally, Assumption 4 is closer to bounds from senstivity analysis on expected outcome functions under unobserved confounding Kennedy (2022). While many existing sensitivity analyses bound the effect of unobserved confounding on treatment in prior data Yadlowsky et al. (2022), in this case the agent's simultaneous treatment selection with data collection makes it less reasonable to bound the dependence of the treatment on U.

Theorem 5. *Suppose for all $\pi \in \Pi$, $\pi(x, u) > 0$ for all $x \in \mathcal{X}$, $u \in \mathcal{U}$. If the principal applies the reward function w_{TT} using an estimate $\hat{\mu}_0^\pi(x)$, then under Assumption 4, the regret is upper bounded as*

$$R(\pi^w) \leq 2\gamma_{max}.$$

Proof. This also follows similarly to the proof of Proposition 10:

$$
\begin{aligned}
V(\pi^*) - V(\pi^{w_{\mathrm{TT}}}) &\leq V(\pi^*) - V(\pi^{w_{\mathrm{TT}}}) + \hat{V}(\pi^{w_{\mathrm{TT}}}) - \hat{V}(\pi^*) \\
&\leq |V(\pi^*) - \hat{V}(\pi^*)| + |\hat{V}(\pi^{w_{\mathrm{TT}}}) - V(\pi^{w_{\mathrm{TT}}})| \\
&\leq 2 \max_{\pi \in \Pi} |\hat{V}(\pi) - V(\pi)| \\
&= 2 \max_{\pi \in \Pi} |E[\pi(X_i, U_i)(\hat{\mu}_0^\pi(X_i) - \mu_0(X_i, U_i))]|
\end{aligned}
$$

By Jensen's inequality,

$$
\begin{aligned}
\max_{\pi \in \Pi} |E[\pi(X, U)(\hat{\mu}_0^\pi(X) - \mu_0(X, U))]| &\leq \max_{\pi \in \Pi} E[\pi(X, U)|\hat{\mu}_0^\pi(X) - \mu_0(X, U)|] \\
&\leq \max_{\pi \in \Pi} E[|\hat{\mu}_0^\pi(X) - \mu_0(X, U)|]
\end{aligned}
$$

Define

$$
\Delta^\Pi(x, u) = \max_{\pi \in \Pi} |\hat{\mu}_0^\pi(x) - \mu_0(x, u)|.
$$

Since for all $\pi \in \Pi$, $\pi(x, u) > 0$ for all $x \in \mathcal{X}$, $u \in \mathcal{U}$, we can apply Lemma 3 below, which says that $\hat{\mu}_0^\pi(x) \in [\min_u \mu_0(x, u), \max_u \mu_0(x, u)]$. Therefore, for all x, u,

$$
\Delta^\Pi(x, u) \leq \Delta(x, u).
$$

Putting this together,

$$
\max_{\pi \in \Pi} E[|\hat{\mu}_0^\pi(X) - \mu_0(X, U)|] \leq E[\max_{\pi \in \Pi} |\hat{\mu}_0^\pi(X) - \mu_0(X, U)|] \leq E[\Delta^\Pi(X, U)] \leq E[\Delta(X, U)].
$$

Therefore, under assumption 4, $R(\pi^w) \leq 2\gamma_{\max}$. $\qquad \square$

Lemma 3. Suppose that for all $\pi \in \Pi$, $P_{U|X=x, T^\pi=0}$ is a well defined probability distribution. Let $\hat{\mu}_0^\pi(x) = E[Y|X = x, T^\pi = 0]$. Then for all π, $\hat{\mu}_0^\pi(x) \in [\min_u \mu_0(x, u), \max_u \mu_0(x, u)]$.

Proof. We decompose $\hat{\mu}_0^\pi(x)$ as

$$
\hat{\mu}_0^\pi(x) = E[Y|X = x, T^\pi = 0] = \int \mu_0(x, u) dP_{U|X=x, T^\pi=0}
$$

For any π, $\int \mu_0(x, u) dP_{U|X=x, T^\pi=0} = \int \mu_0(x, u) dW(u)$ for some distribution W; therefore,

$$
\left\{ \int \mu_0(x, u) dP_{U|X=x, T^\pi=0} : \pi \in \Pi \right\}
$$

$$
\subseteq \left\{ \int \mu_0(x, u) dW(u) : W \text{ is a distribution over } U, \int dW(u) = 1 \right\}.
$$

Both the maximizer and the minimizer of the smaller set are contained in the larger set. Specifically,

$$
\max_\pi \hat{\mu}_0^\pi(x) = \max_\pi \int \mu_0(x, u) dP_{U|X=x, T^\pi=0} \leq \max_{W: \int dW(u)=1} \int \mu_0(x, u) dW(u) \leq \max_u \mu_0(x, u)
$$

$$
\min_\pi \hat{\mu}_0^\pi(x) = \min_\pi \int \mu_0(x, u) dP_{U|X=x, T^\pi=0} \geq \min_{W: \int dW(u)=1} \int \mu_0(x, u) dW(u) \geq \min_u \mu_0(x, u)
$$

Therefore, for all π, $\hat{\mu}_0^\pi(x) \in [\min_u \mu_0(x, u), \max_u \mu_0(x, u)]$. $\qquad \square$

Overall, we have discussed two plausible data collection settings for the principal to estimate $\mu_0(x)$ to implement the reward function w_{TT}. In the first setting using auxiliary data where $\mu_0(x)$ is identifiable, the regret is bounded by the gap between the conditional effects of U on $Y_i(0)$ and the marginal effect of only X on $Y_i(0)$. In the second setting using untreated units from the agent's treatment population, $\mu_0(x)$ is not identifiable, and the regret can be bounded under a stronger assumption on the sensitivity of $\mu_0(X, U)$ to the agent's private information U_i.

E ADDITIONAL EXPERIMENT DETAILS AND RESULTS

We give additional experiment details and results here.

E.1 Implementation

All experiment code is available at `https://github.com/serenalwang/counterfactual_metrics`. All models were trained using the linear regression and logistic regression packages from scikit-learn (Pedregosa et al., 2011). All categorical features were one-hot encoded.

E.2 Additional dataset details

We provide additional setup and training details for each dataset.

E.2.1 Horse Colic dataset

For the Horse Colic dataset from UCI Dua and Graff (2017), we let X consist of all features observed prior to surgery, which includes 13 categorical features and 7 numerical features. All features used are listed in Figure 3. We removed all examples in which the horse was euthanized, and used the remaining "outcome" variable as Y, where we set $Y_i = 1$ if the horse lived, and $Y_i = -1$ if the horse died. We only used the main `horse-colic.data` dataset, and did not use the "test" dataset included in the UCI directory.

To simulate $\mu_t(x)$, we assume that the outcome distribution takes the parametric form,

$$P(Y(t) = 1 | X = x) = \sigma(\beta_0 + \beta_1^\top x + \beta_2 t + \beta_3^\top xt)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the standard logistic function.

We trained a logistic regression model of Y on X, T, and the interaction term XT to estimate parameters $\hat{\beta}$, and used the resulting estimate to compute $\mu_t(x) = \sigma(\hat{\beta}_0 + \hat{\beta}_1^\top x + \hat{\beta}_2 t + \hat{\beta}_3^\top xt)$. For the fitted model, the AUC was 0.9924, and the accuracy was 0.6824. When computing agent best responses and regret, we take this function $\mu_t(x)$ to be given as synthetic mean conditional potential outcomes. Note that the clinical validity of $\mu_t(x)$ as actual potential outcomes cannot be verified from the dataset alone. There may be error in both our unconfoundedness assumption with respect to X, and in the logistic parametric specification of the outcome model.

In Table 1, the "demographic information" for the Horse Colic dataset consists of only the *age* feature.

E.2.2 International Stroke Trial dataset

For the International Stroke Trial dataset International Stroke Trial Collaborative Group (1997), we let X include all clinical data prior to treatment, which includes all "Randomisation data" except for dates and times. This includes 17 categorical features and 3 numerical features. All features used are listed in Figure 1. For the outcome variable Y, we apply the negative of the scalarized composite outcome score from Kallus and Zhou (2018). Specifically,

$$\begin{aligned} Y = &- 2\,\mathbb{1}(\text{death}) - \mathbb{1}(\text{recurrent stroke}) - 0.5\,\mathbb{1}(\text{pulmonary embolism or intracranial bleeding}) \\ &- 0.5\,\mathbb{1}(\text{other side effects}) + 2\,\mathbb{1}(\text{full recovery at 6 months}) + \mathbb{1}(\text{discharge within 14 days}). \end{aligned}$$

This results in $Y \in [-4, 3]$.

To simulate $\mu_t(x)$, we assume the conditional mean of the potential outcome distribution has the linear parametric form,

$$E[Y(t)|X = x] = \beta_0 + \beta_1^\top x + \beta_2 t + \beta_3^\top xt.$$

We trained a linear regression model of Y on X, T, and the interaction term XT to estimate parameters $\hat{\beta}$, and used the resulting estimate to compute $\mu_t(x) = \hat{\beta}_0 + \hat{\beta}_1^\top x + \hat{\beta}_2 t + \hat{\beta}_3^\top xt$. The fitted RMSE was 1.34 and the R^2 was 0.26. As with the Horse Colic dataset, when computing agent best responses and regret, we take this function $\mu_t(x)$ to be given as synthetic mean conditional potential outcomes. In this dataset the unconfoundedness

assumption should hold as this data came from a randomized control trial. However, there may still be error in the linear parametric specification of the outcome model.

In Table 1, the "demographic information" for the Stroke Trial dataset consists of both the *age* feature and the *sex* feature.

E.3   Effect of better estimates of $\mu_t(x)$

Better estimates of $\mu_t(x)$ may come from doubly robust estimators. However, the contribution of this work is not to improve estimation methods for $\mu_t(x)$, nor do the proposed reward policies rely on the quality or variance of $\mu_t(x)$ estimators. The quality of $\mu_t(x)$ only affects the how well the regrets reported in experiments might approximate real regrets. This is because in experiments, we estimate $\mu_t(x)$ to simulate potential outcomes for two purposes: *(i)* simulating the agent's perception of potential outcomes and thus yielding the agent's best responses; and *(ii)* calculating an approximate regret. Better estimates of $\mu_t(x)$ would improve the alignment with both of these simulations with reality.

Our theory only assumes that the principal has an estimator $\hat{\mu}_0(x)$ which is unbiased with respect to the agent's preceived $\mu_0(x)$ function which the agent uses to calculate their best response. For the experiments, this unbiasedness assumption is true, as we assume both principal and agent have access to the same synthetic $\mu_t(x)$ values. We do not directly simulate the principal's calculation of $\hat{\mu}_0(x)$ from a subset of the data, since this would not affect the actions of the expectation maximizing agent.

The main limitation of our experiments is that in the absence of true counterfactual outcomes, they rely on parametric estimates of μ_t which we can't guarantee are well specified. An ideal dataset for evaluating true welfare impacts would have a structure that identifies P_{X_i} and $\mu_t(x)$, where $\mu_t(x)$ ideally matches the agent's *perceived* mean conditional potential outcomes. While we can't guarantee that our estimated imputed values for $\mu_t(x)$ match real providers, our experiments provide a structure by which such experiments could be run in the future if regulatory agencies have internal access to more ideal data, perhaps even through surveying providers themselves.

E.4   Calculating γ_{marg}

When simulating information asymmetry, we compute empirical estimates of γ_{marg} over the data. For each individual feature in Figures 1 and 3, let X represent the individual feature, and let U represent the set of all other features. We use the full regression result as $\mu_0(x, u)$. To calculate $\mu_0(x)$ where x represents the individual feature, we take the empirical conditional mean over the dataset:

$$\mu_0(x) = \frac{\sum_{i=1}^n \mu_0(X_i, U_i)\, \mathbb{1}(X_i = x)}{\mathbb{1}(X_i = x)}.$$

Then, γ_{marg} is calculated empirically over the dataset as

$$\gamma_{\mathrm{marg}} = \frac{1}{n} \sum_{i=1}^n |\mu_0(X_i) - \mu_0(X_i, U_i)|.$$

We take these empirical estimates as given in the absence of a closed form for the joint distribution of all features. More sophisticated Bayesian distribution estimation or smoothing may produce different γ_{marg} estimates.

E.5   Additional results

Table 2 shows the utility and regret comparisons for different reward functions, including an additional comparison to Reward Function 5 in the Appendix, w_{TO}.

Figure 2 shows the regrets when the principal knows only each individual feature. Interestingly, use of several of these individual features leads to worse regret than if the principal knew no features and just estimated the marginal expected untreated potential outcome $E[Y(0)]$. This confirms that the regret amplification effect can occur for features other than the demographic information in Table 1.

Table 2: Utility and regret comparisons for different reward functions. For each reward function w, we report utility $V(\pi^w)$, regret $R(\pi^w)$, and the realized treatment rate $P(T^{\pi^w} = 1)$.

| Reward function	Horse Colic dataset			Stroke Trial dataset		
	Utility	Regret	Treat rate	Utility	Regret	Treat rate
w_{ATO}	0.00000	0.1470	0.1922	0.00004	0.0251	0.0001
w_{ATT}	0.00784	0.1391	0.0039	0.00013	0.0250	0.0001
w_{TO}	0.08599	0.0610	0.6706	-0.08278	0.1079	0.6689
w_{TT}	0.14695	0.000	0.2431	0.02518	0.000	0.1872
w_{TT} (no info)	0.09761	0.0493	0.6274	-0.04888	0.0741	0.4829
w_{TT} (demographic info)	0.09761	0.0493	0.6274	-0.06392	0.0891	0.5041
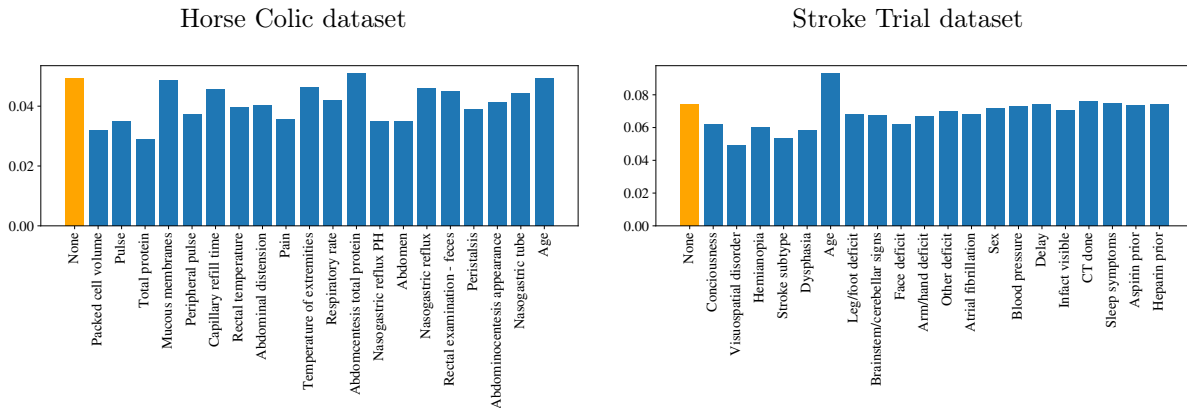
Figure 2: Regret if the principal only knows the labeled feature. Features are sorted in order of their individual γ_{marg}.

Figure 3 is the equivalent of Figure 1 for the Horse Colic dataset. For the Horse Colic dataset, the regret drops to close to zero if the principal knows only a few features. This is likely due to the fact that there is little variation in other feature values in the dataset conditioned *packed cell volume* and *pulse*, exacerbated by the fact that the dataset is so small. The value of γ_{marg} for these two features combined is small compared to the individual γ_{marg} values in Figure 3 (*left*), at 0.035.

This may change for larger datasets, where it may be more likely to observe more samples with the same *packed cell volume* and *pulse* values alongside more variation in other feature values. If in practice other features did not vary much given *packed cell volume* and *pulse*, then the regret shape in Figure 3 (*right*) would hold. More data would be needed to verify the joint distribution of X, U to confirm this.

The Stroke Trial dataset does not exhibit a similar effect since its three numerical features take significantly fewer possible values relative to the size of the dataset.

F Additional Discussion of Modeling Assumptions

We do not explicitly model the agent's costs, and all budget constraints are contained in the agent's feasible set of treatment rules Π. This may be viewed as considering the agent's incentives when there is no cost to treatment, thus decoupling our analysis of the agent's incentives to do well on healthcare report cards from the broader treatment pricing market. In practice, the compensation hospitals and doctors receive for treatment could yield a positive adjustment or sometimes a negative adjustment to their utility per treated unit. Analysis of how other external incentives pair with the quality measure incentives would be interesting future work, but for now we focus on the incentives induced solely by the accountability metrics.

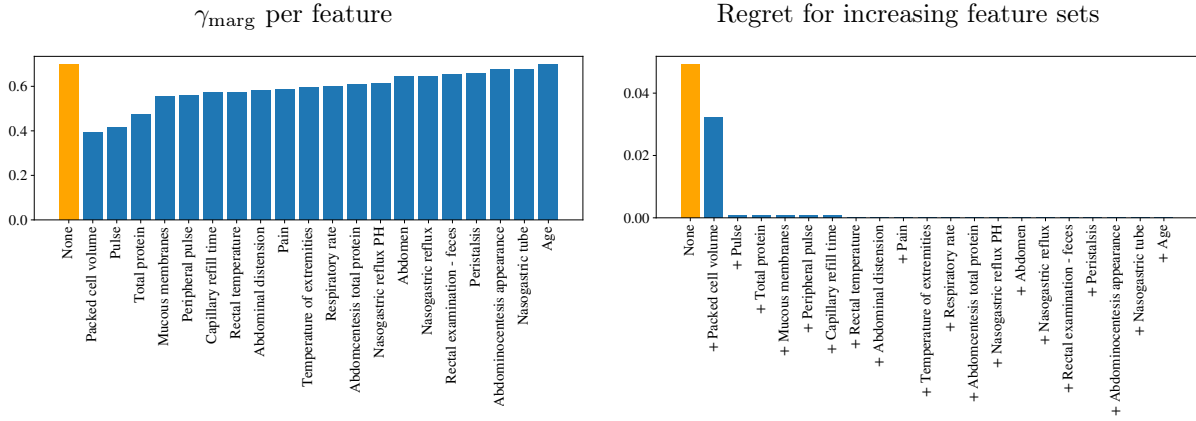Whether the information asymmetry modeled here is common or severe in the medical space has been debated

Figure 3: Regret under fine-grained information asymmetry on the Horse Colic dataset. The *left* plot shows γ_{marg} values if the principal only knows each individual feature. The *right* plot shows regret as the principal accumulates features from the left (most important).

and may change over time. Dranove et al. (2003) remark that "providers may be able to improve their ranking by selecting patients on the basis of characteristics that are unobservable to the analysts but predictive of good outcomes." However, medical treatment protocols are also generally heavily codified, with online clinical decision support tools becoming increasingly widely used upt (2022). Still, if the regulatory agency does not have full access to patents' medical records, then information asymmetry will arise.