Interpolated Stochastic Interventions based on Propensity Scores, Target Policies and Treatment-Specific Costs

Johan de Aguas[1,2]

[1]Department of Mathematics, University of Oslo
[2]Integreat – Norwegian Centre for Knowledge-driven Machine Learning
johanmd@uio.no

Abstract

We introduce families of stochastic interventions for discrete treatments that connect causal modeling to cost-sensitive decision making. The interventions arise from a cost-penalized information projection of the independent product of the organic propensity and a user-specified target, yielding closed-form Boltzmann–Gibbs couplings. The induced marginals define modified stochastic policies that interpolate smoothly, via a single tilt parameter, from the organic law or from the target distribution toward a product-of-experts limit when all destination costs are strictly positive. One of these families recovers and extends incremental propensity score interventions, retaining identification without global positivity. For inference, we derive efficient influence functions under a nonparametric model for the expected outcomes after these policies and construct one-step estimators with uniform confidence bands. In simulations, the proposed estimators improve stability and robustness to nuisance misspecification relative to plug-in baselines. The framework can operationalize graded scientific hypotheses under realistic constraints: because inputs are modular, analysts can sweep feasible policy spaces, prototype candidates, and align interventions with budgets and logistics before committing experimental resources. This could help close the loop between observational evidence and resource-aware experimental design.

Introduction

Evaluating the causal effects after actions, interventions, and treatment regimes is fundamental to the design and optimization of decision-making systems. Classical causal inference tasks are typically framed via *hard interventions*: hypothetical scenarios where the exposure is deterministically set to a fixed value for all units in a population (Pearl 2009; Imbens and Rubin 2015). Although foundational, hard interventions can be overly rigid for many real-world domains, such as healthcare and economic policy, where key information needed to take action is often incomplete, treatment allocation is constrained by limited resources, or when inference becomes unstable because the data-generating process (DGP) assigns zero probability to certain treatments in specific subpopulations (Díaz and van der Laan 2013).

Stochastic/soft interventions offer an expressive alternative, capturing hypothetical scenarios in which the treatment assignment mechanism is altered through probabilistic or functional shifts, typically as a function of observed covariates, while the rest of the DGP is left unchanged (Correa and Bareinboim 2020a,b). This formulation enables researchers to define, analyze, and estimate causal effects under interventions that are more adaptive, flexible and aligned with the practical constraints of real-world policy implementation (Haneuse and Rotnitzky 2013; Sarvet et al. 2023).

Among the broad class of stochastic interventions, *incremental propensity score interventions* (IPI) (Kennedy 2018) have gained particular appeal. IPIs tilt the organic propensity score by a tunable parameter, yielding controlled and interpretable odds ratio modifications that allow analysis across a continuum of intensities (Bonvini et al. 2023). They integrate naturally with modern causal workflows and align closely with real-world policy design. Applications include dropout and censoring (Kim, Kennedy, and Naimi 2021), time-fixed and time-varying treatments (Naimi et al. 2021; Rudolph et al. 2022), resource-constrained decision making (Sarvet et al. 2023), interventional mediation (Díaz and Hejazi 2020; Hejazi et al. 2023), fairness assessment (McClean et al. 2024; Opacic, Wei, and Zhou 2025), and general sensitivity analysis (Levis et al. 2024), among others.

Mechanistically, IPIs differ from other soft interventions such as *general stochastic interventions* (GSI) (Correa and Bareinboim 2020a,b) and *modified treatment policies* (MTP) (Haneuse and Rotnitzky 2013; Díaz et al. 2023), which include *shift intervention policies* (SIP) (Sani, Lee, and Shpitser 2020). GSIs allow ample changes to the exposure mechanism, for example, reducing the parent set or introducing auxiliary noise, and they may alter the σ-algebra of its conditional distribution. MTPs set the counterfactual treatment deterministically as a function of the observed treatment and covariates, with SIPs restricting the rule to depend only on the observed treatment. By contrast, IPIs operate solely by tilting the organic propensity score, preserve the underlying measurable structure of the exposure law, and define counterfactual treatments by sampling from the tilted distribution (Kennedy 2018; Bonvini et al. 2023). Recently, *generalized policies* (GPs) have been proposed, defined as a marginal of an optimal transport map, matching the distribution induced by a reallocation mechanism conditional on co-

variates, observed treatment, and auxiliary noise. They have been shown to offer favorable properties for partial identification under latent confounding (Levis et al. 2024).

The expected outcome after an IPI can be identified without the global positivity condition required for hard interventions, provided conditional ignorability holds given a backdoor admissible adjustment set. This is important because positivity is often violated in practice, especially with high-dimensional covariates or in longitudinal and dynamic treatment settings (Kennedy 2018; Bonvini et al. 2023). When identifiable, the expected outcome after an IPI can be estimated with efficient semiparametric methods that offer favorable statistical properties, including robustness to misspecification and root-n consistency. Common implementations use either *one-step* estimators (Bonvini et al. 2023) or *targeted minimum loss estimation* (TMLE) (Naimi et al. 2021). Inference can be conducted pointwise, at a fixed tilt parameter (for example, a specific odds-ratio shift), or uniformly over a range of values to recover the full intervention response curve (Kennedy 2018).

Motivation and contribution

While standard IPIs interpolate between a non-intervention and a hard intervention, many applications call for more nuanced targets. Logistical considerations may prescribe specific treatment shares across arms (for example, treatment 1 to 80% of units, treatment 2 to 10%, and no treatment to 10%), or require a binary exposure to be split 50/50 to preserve fairness constraints. In addition, with multiple treatment arms, deployment costs can vary widely, and finite budgets may limit the overall reach of particular arms.

To address this, we introduce a formulation that yields two families of stochastic interventions governed by a single tilt parameter. The first family smoothly interpolates, in a cost-aware manner, from a non-intervention to the *product of experts* (PoE) blend of the organic propensity score and a pre-specified target policy; the second interpolates from the target policy to the PoE. The PoE limit arises when all treatment options carry strictly positive costs; if some actions are costless, the limiting behavior depends on the zero-cost set. These cost-aware interpolations are derived from a *cost-penalized I-projection* and can be viewed as a limiting case of a relaxed optimal transport problem. The first family directly generalizes IPIs, accommodates cost structures and target policies, and, as with IPIs, does not require the global positivity conditions demanded by hard interventions. When the target distribution is non-degenerate, the resulting target marginal also defines a valid stochastic intervention under a positivity condition.

Under standard identification conditions for observational data, we derive efficient influence functions in a nonparametric model for the expected outcome after these interventions, develop robust one-step estimators, and construct asymptotically valid confidence band.

We believe these ideas can help make interventions in causal inference more adaptable and better aligned with practical constraints in domains such as healthcare and economics. They also suggest several directions for future work, including exploring connections with optimal transport.

Preliminaries

Hard interventions

Let $A \in \mathcal{A}$ be a discrete point-exposure variable, $Y \in \mathbb{R}$ a continuous outcome variable, $W \in \mathcal{W}$ a vector of pre-exposure covariates, and $\pi(a \,|\, w) := \mathbb{P}(A = a \,|\, W = w)$ the propensity score of treatment option $a \in \mathcal{A}$. Potential outcomes Y^a encode unit-level counterfactuals after a hard intervention $\mathrm{do}(A = a)$ (Pearl 2009). Under counterfactual consistency, positivity $\pi(a \,|\, w) \in (0, 1), \forall w \in \mathcal{W}$, and conditional ignorability / backdoor admissibility of W, one can identify the expected outcome after intervention $\mathrm{do}(A = a)$ from observational data via the g-computation / backdoor formula, as $\mathbb{E}\left[Y^a\right] = \mathbb{E}\left[Y \mid \mathrm{do}(A = a)\right] = \mathbb{E}_W\left\{Q(Z, a)\right\}$, with $Q(w, a) = \mathbb{E}[Y \,|\, W = w, A = a]$.

Incremental propensity score interventions (IPI)

For a binary point-exposure A, an IPI can be defined as a family of stochastic interventions that replace the organic propensity score $\pi(1 \,|\, w)$ by:

$$\widetilde{\pi}_\delta(1 \,|\, w) := \frac{e^\delta \pi(1 \,|\, w)}{e^\delta \pi(1 \,|\, w) + \pi(0 \,|\, w)}, \quad \delta \in \mathbb{R}, \qquad (1)$$

for all $w \in \mathcal{W}$, and where δ is a user-specified tilt parameter that governs the degree of deviation from the organic treatment assignment mechanism. Notably, δ corresponds to the log-odds ratio between the modified and the organic propensity scores, providing an interpretable parameterization of intervention intensity (Kennedy 2018).

IPIs provide a smooth interpolation between a non-intervention and a hard intervention. Specifically, when $\delta = 0$, the intervention leaves the propensity score unchanged, i.e., $\widetilde{\pi}_\delta(a \,|\, w) = \pi(a \,|\, w)$. As $\delta \to \infty$, the modified mechanism approaches a deterministic assignment to treatment value $a = 1$, i.e., $\widetilde{\pi}_\delta(a \,|\, w) \to \mathbb{I}(a = 1)$; conversely, as $\delta \to -\infty$, the intervention converges to always assigning $a = 0$, i.e., $\widetilde{\pi}_\delta(a \,|\, w) \to \mathbb{I}(a = 0)$.

The expected outcome after a stochastic intervention, such as an IPI with modified exposure mechanism $A \sim \widetilde{\pi}_\delta$, can be denoted using notation $\mathbb{E}[Y \wr \sigma_A = \widetilde{\pi}_\delta]$ (Correa and Bareinboim 2020a,b), by $\mathbb{E}[Y^{\widetilde{\pi}_\delta}]$ (Naimi et al. 2021), or by $\mathbb{E}\left[Y \mid \mathrm{do}(A \sim \widetilde{\pi}_\delta)\right]$. Under conditional ignorability / backdoor admissibility of W, it is identified as:

$$\mathbb{E}[Y^{\widetilde{\pi}_\delta}] = \sum_{a \in \mathcal{A}} \mathbb{E}_W\left\{\widetilde{\pi}_\delta(a \,|\, W)\, Q(W, a)\right\}, \qquad (2)$$

Notably, unlike the case of a hard intervention, this causal query remains identifiable without requiring a global positivity condition (Kennedy 2018; Bonvini et al. 2023).

Cost-penalized I-projection

While exponentially tilted distributions have been characterized as standard I-projections under a mean constraint (Levis et al. 2024), and IPIs have been used to model interventions under limited resource and treatment costs (Sarvet et al. 2023), a broader connection between IPIs and limiting cases of relaxed optimal transport problems that accommodate cost structures and target distributions has not been described. In this work, we provide that characterization.

Given two input probability measures π (*source*) and ν (*target*) over an action set $\mathcal{A}$, a summable / integrable cost function on pairs $c : \mathcal{A}^2 \to [0, \infty)$, and a penalization parameter $\delta \geq 0$, we define the *cost-penalized I-projection* (CPIP) of the independent product $\pi \otimes \nu$ as the joint distribution $\gamma \in \mathcal{M}_+^1(\mathcal{A}^2)$ (in the class of distributions over the Cartesian product $\mathcal{A}^2 = \{(A', A'') : A', A'' \in \mathcal{A}\}$) that solves:

$$\inf_{\gamma \in \mathcal{M}_+^1(\mathcal{A}^2)} \mathbb{D}_{\mathrm{KL}}(\gamma \,|\, \pi \otimes \nu) + \delta\, \mathbb{E}_\gamma \left\{ c(A', A'') \right\}, \quad (3)$$

where $\mathbb{D}_{\mathrm{KL}}$ represents the Kullback–Leibler divergence.

This problem is closely related to *unconstrained* or *limiting-case* variants of entropic optimal transport and Schrödinger bridge problems (Léonard 2014; Frogner et al. 2015; Chizat et al. 2018a; Peyré and Cuturi 2019). Yet, while entropic optimal transport problems typically require iterative solvers such as the Sinkhorn–Knopp algorithm, the CPIP problem admits a closed-form solution thanks to the strong convexity and smoothness of its objective. Its unique minimizer is given by the Boltzmann–Gibbs kernel:

$$\gamma_\delta^\star(a', a'') \propto \pi(a')\, \nu(a'')\, e^{-\delta c(a', a'')}. \quad (4)$$

Tilted marginal distributions

Let A be a categorical point-exposure variable with domain $\mathcal{A} = \{\alpha_1, \ldots, \alpha_K\}$, where the K treatment options may potentially include a placebo or null-treatment. Let ν be a target marginal distribution over $\mathcal{A}$ and let $\pi(a \,|\, w)$ be the propensity score of treatment $A = a$ given pre-exposure covariate profile $W = w$. Let $c(a', a'')$ be the cost of reallocating a unit from treatment $A = a'$ to treatment $A = a''$, which does not depend on the profile w. Then, for all $a \in \mathcal{A}$ and $w \in \mathcal{W}$,

$$\pi_\delta^\star(a \,|\, w) := \frac{\pi(a \,|\, w) \sum_{a'' \in \mathcal{A}} \nu(a'')\, e^{-\delta c(a, a'')}}{\sum_{a', a'' \in \mathcal{A}} \pi(a' \,|\, w)\, \nu(a'')\, e^{-\delta c(a', a'')}}, \quad (5)$$

$$\nu_\delta^\star(a \,|\, w) := \frac{\nu(a) \sum_{a' \in \mathcal{A}} \pi(a' \,|\, w)\, e^{-\delta c(a', a)}}{\sum_{a', a'' \in \mathcal{A}} \pi(a' \,|\, w)\, \nu(a'')\, e^{-\delta c(a', a'')}} \quad (6)$$

are the marginal distributions of the CPIP solution in (4) when the source is $\pi(\cdot \,|\, w)$ and the target is ν. We refer to these as the *tilted source / target marginal distributions*. Derivation is provided in the technical appendix A1.

Remark 1. *Let:*

1. *$A \in \mathcal{A} = \{0, 1\}$ be a binary point-exposure,*
2. *the target marginal ν be the degenerate distribution that always assigns treatment, $\nu(a) = \mathbb{I}(a = 1)$,*
3. *$c(a', a'') = \mathbb{I}(a' \neq a'')$ be the Hamming cost.*

Then, the tilted source marginal of the CPIP solution with penalization parameter δ coincides with an IPI with tilt parameter δ, and thus $\widetilde{\pi}_\delta(1 \,|\, w) = \pi_\delta^\star(1 \,|\, w)$ for all $w \in \mathcal{W}$.

We provide a proof in the technical appendix A2.

This remark provides an interpretation of IPIs in the binary exposure setting with $\delta > 0$. Suppose the target intervention assigns treatment to all units, and the reallocation cost is 1 for any switch between $A = 0$ and $A = 1$ while maintaining the same treatment status has no cost. Then an IPI with tilt parameter δ coincides with the first marginal of a joint distribution that minimizes the expected reallocation cost, subject to a penalty (with coefficient $1/\delta$) on its divergence from the independent product distribution $\pi(a' \,|\, w)\, \mathbb{I}(a'' = 1)$. Such product places zero mass on pairs $(a', 0)$ and assigns mass to pairs $(a', 1)$ equivalent to the propensity score $\pi(a' \,|\, w)$.

The CPIP objective is strictly convex and well posed for $\delta > 0$, yielding a unique solution. For $\delta < 0$, the joint distribution $\gamma_\delta^\star$ in (4) is not the optimizer of the CPIP program; however, when the action set $\mathcal{A}$ is finite and costs are bounded, the associated Boltzmann–Gibbs kernel is summable, so $\gamma_\delta^\star$ remains normalizable and produces smooth and closed-form *parametric extensions* for $\pi_\delta^\star$ and $\nu_\delta^\star$. In this extrapolative regime, the induced tilted marginals shift mass toward higher-cost and repulsive pairings and are pushed away from the independence law $\pi(\cdot \,|\, w) \otimes \nu$, favoring more adversarial couplings. For continuous spaces with potentially unbounded costs, additional integrability or tail conditions are required for $\delta < 0$ to deliver well posed stochastic policies.

Notably, within the setting of remark 1, the tilted target marginal coincides with the input target distribution, so $\nu_\delta^\star(a \,|\, w) = \nu(a) = \mathbb{I}(a = 1)$ for all $\delta > 0$ and all $w \in \mathcal{W}$. Consequently, it does not provide a basis for a differentiated stochastic policy. In the next remark, we introduce a generalization of IPIs based on stochastic plans that incorporates treatment-specific costs and a flexible target policy.

Remark 2. *Let:*

1. *$A \in \mathcal{A} = \{\alpha_1, \ldots, \alpha_K\}$ be a categorical point-exposure variable with K treatment options,*
2. *The target marginal ν be any valid probability distribution over $\mathcal{A}$,*
3. *the reallocation cost from $A = \alpha_j$ to $A = \alpha_k \neq \alpha_j$ be a value that is specific for the received treatment α_k and constant over profiles $W = w$, i.e., $c(\alpha_j, \alpha_k) = c(\alpha_k)\, \mathbb{I}(\alpha_j \neq \alpha_k)$, with $0 \leq c(a) < \infty$ for all $a \in \mathcal{A}$.*

Then, for each $w \in \mathcal{W}$, the tilted marginals of the CPIP solution with parameter δ are:

$$\pi_\delta^\star(a \,|\, w) := \frac{(\zeta_\delta + \xi_\delta(a))\, \pi(a \,|\, w)}{\sum_{a' \in \mathcal{A}} (\zeta_\delta + \xi_\delta(a'))\, \pi(a' \,|\, w)}, \quad (7)$$

$$\nu_\delta^\star(a \,|\, w) := \frac{\nu(a) - \xi_\delta(a)(1 - \pi(a \,|\, w))}{\sum_{a' \in \mathcal{A}} (\zeta_\delta + \xi_\delta(a'))\, \pi(a' \,|\, w)}, \quad (8)$$

where:

$$\xi_\delta(a) := \nu(a) \left(1 - e^{-\delta c(a)}\right), \quad (9)$$

$$\zeta_\delta := \sum_{a' \in \mathcal{A}} \nu(a')\, e^{-\delta c(a')}. \quad (10)$$

We provide a derivation in the technical appendix A3.

Note that $\pi_0^\star(a \,|\, w) = \pi(a \,|\, w)$ and $\nu_0^\star(a \,|\, w) = \nu(a)$ for all $a \in \mathcal{A}$ and $w \in \mathcal{W}$. In other words, setting $\delta = 0$ results in no modification of the input distributions. Furthermore, denote $\mathcal{A}_0 = \{a \in \mathcal{A} : c(a) = 0\}$, $\mathcal{A}_+ = \{a \in \mathcal{A} : c(a) > 0\}$, $\nu^\dagger(a) := \nu(a)\, \mathbb{I}(a \in \mathcal{A}_+) + \sum_{a \in \mathcal{A}_0} \nu(a)$ and

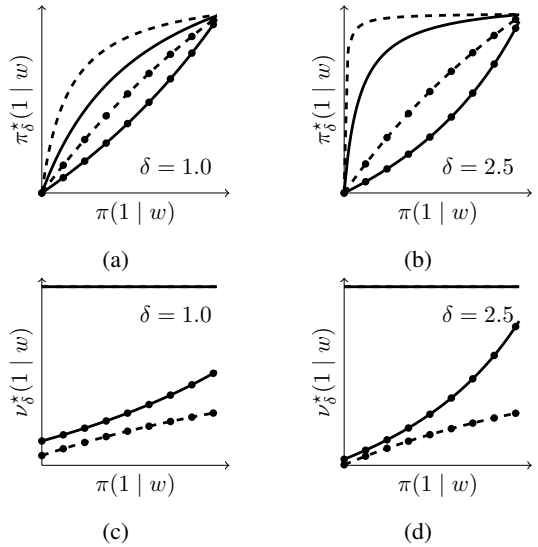


Figure 1: Tilted source distributions $\pi^\star_\delta$ in panels (a) and (b), and tilted target distributions $\nu^\star_\delta$ in panels (c) and (d), for a binary exposure. Curves show pointwise transformation of the propensity score $\pi(1 \mid w)$ for $\delta \in \{1.0, 2.5\}$. Target configuration is encoded by dot pattern: $\nu = (0, 1)$ is non-dotted and $\nu = (0.7, 0.3)$ is dotted. Cost structure is encoded by line style: solid for $c = (1, 1)$ and dashed for $c = (0, 2)$. Vector components are ordered as $(A = 0, A = 1)$.

$\pi^\dagger(a \mid w) := \pi(a \mid w) + (1 - \pi(a \mid w))\, \mathbb{I}(a \in \mathcal{A}_0)$. Then, in the limit $\delta \to \infty$, one obtains:

$$\pi^\star_\infty(a \mid w) = \frac{\pi(a \mid w)\, \nu^\dagger(a)}{\sum_{a' \in \mathcal{A}} \pi(a' \mid w)\, \nu^\dagger(a')}, \tag{11}$$

$$\nu^\star_\infty(a \mid w) = \frac{\pi^\dagger(a \mid w)\, \nu(a)}{\sum_{a' \in \mathcal{A}} \pi^\dagger(a' \mid w)\, \nu(a')}. \tag{12}$$

When all treatment costs are positive, both reduce to the *product of experts* (PoE) distribution $\mathrm{PoE}(a) \propto \pi(a \mid w)\, \nu(a)$ (Hinton 1999).

Although, as noted in remark 1, typical incremental IPIs are defined using the tilted source distribution $\pi^\star_\delta$, we argue that the tilted target distribution $\nu^\star_\delta$ can also be used to define stochastic policies, particularly when the target ν is non-degenerate. For example, suppose the ideal policy prescribes treatment 1 to 80% of units, treatment 2 to 10%, and no treatment to the remaining 10%. Then, in the case with positive costs, $\nu^\star_\delta$ provides a smooth interpolation between this ideal allocation ($\delta = 0$) and the PoE law ($\delta \to \infty$).

Figure 1 presents the tilted marginal distributions for a binary exposure under varying configurations of the tilt parameter δ, cost functions c, and target distribution ν.

Pushforward distribution

Since the CPIP objective imposes no soft or hard constraints on the marginals, the optimizer $\gamma^\star_\delta$ in (4) is not a relaxed optimal transport plan from π to ν. Consequently, although $\nu^\star_\delta$ provides a smooth, cost-aware, one-parameter deformation of the input distributions, it should not be interpreted as the pushforward of π through $\gamma^\star_\delta$. One may instead enforce the source marginal to be equal or close to π while leaving the other marginal unconstrained by adding marginal penalties, as in relaxed optimal transport (Frogner et al. 2015; Chizat et al. 2018a,b), but such formulations typically do not admit closed-form solutions. As an operational alternative, one can consider the *pushforward distribution*, obtained by applying the Markov kernel induced by $\gamma^\star_\delta$ to the source law:

$$\nu^{\star\star}_\delta(a \mid w) := \sum_{a' \in \mathcal{A}} \frac{\gamma^\star_\delta(a', a \mid w)}{\pi^\star_\delta(a' \mid w)}\, \pi(a' \mid w), \tag{13}$$

which is well defined because $\pi^\star_\delta(a \mid w)$ is proportional to $\pi(a \mid w)$ with a strictly positive weight (hence $\pi(a \mid w)/\pi^\star_\delta(a \mid w) > 0$ for all a and w). Here $\nu^{\star\star}_\delta(\cdot \mid w)$ represents the distribution induced by applying the stochastic plan encoded in $\gamma^\star_\delta$ to units drawn from the organic source $\pi(\cdot \mid w)$. A systematic comparison between marginal-unconstrained, marginal-penalized, and pushforward formulations is left for future work.

Identification

Let $\mu^S_\delta \equiv \mathbb{E}\big[Y^{\pi^\star_\delta}\big]$ and $\mu^T_\delta \equiv \mathbb{E}\big[Y^{\nu^\star_\delta}\big]$ denote the expected outcomes after the stochastic interventions associated with the tilted marginals in remark 2. These functionals are identified from observational data under conditional ignorability / backdoor admissibility.

Let $W = \mathrm{pa}(A; \mathcal{G})$ denote the causal parents of A in the assumed causal graph $\mathcal{G}$ of the system. In typical structures without latent confounding between A and Y, the set W is backdoor admissible (Pearl 2009). However, W may include instrumental variables or predictors of A that are unrelated to Y; such variables do not induce bias but can reduce precision (Cinelli, Forney, and Pearl 2024). This motivates the use of a possibly smaller backdoor admissible set $Z \subseteq W$ such that $Y \perp\!\!\!\perp_d (W \setminus Z) \mid Z, A$ in $\mathcal{G}$. If such set exists, then:

$$\mu^S_\delta = \sum_{a \in \mathcal{A}} \mathbb{E}_W \left\{ \pi^\star_\delta(a \mid W)\, Q(Z, a) \right\}, \tag{14}$$

$$\mu^T_\delta = \sum_{a \in \mathcal{A}} \mathbb{E}_W \left\{ \nu^\star_\delta(a \mid W)\, Q(Z, a) \right\}, \tag{15}$$

where $Q(z, a) = \mathbb{E}[Y \mid Z = z, A = a]$ and $\mathbb{E}_W$ denotes expectation with respect to the marginal distribution of W (equivalently, of Z since $Z \subseteq W$).

For (15) to be well defined and estimable, a global positivity condition is required:

$$\sup_{a \in \mathcal{A}} \frac{\nu^\star_\delta(a \mid W)}{\pi(a \mid W)} < \infty, \;\; P_W\text{-almost surely.} \tag{16}$$

A sufficient condition is uniform overlap on the support of ν, namely the existence of constants $\{\varepsilon_a > 0\}_{a \in \mathcal{A}}$ such that $\pi(a \mid W) \geq \varepsilon_a$ P_W-almost surely for all a with $\nu(a) > 0$. By contrast, the corresponding global positivity condition for μ^S_δ holds automatically, since $\pi^\star_\delta(a \mid W)$ is a bounded multiplicative tilt of $\pi(a \mid W)$. Hence, no additional overlap is needed for μ^S_δ, in accordance with the corresponding property of IPIs (Kennedy 2018).

Semiparametric efficient estimation

To develop estimators of the expected outcomes, we assume that the cost function c and the target distribution ν are known and fixed; they are treated as design inputs rather than estimated from data. Plug-in estimators for μ_δ^S and μ_δ^T are then obtained by estimating π and Q, applying the transformations in (7) and (8) to construct the tilted marginals, substituting these into the inner product in (14) and (15), and averaging over i.i.d. samples.

Recent work has extended semiparametric and robust estimation to queries under stochastic interventions (Díaz and van der Laan 2012; Duong, Li, and Xu 2021; Bonvini et al. 2023; Díaz et al. 2023). Semiparametric efficient approaches are prized for accommodating data-adaptive learning while attaining optimal asymptotics under flexible DGP assumptions (Díaz 2019). Many estimators enjoy double or multiple robustness, remaining consistent despite some nuisance misspecification (Daniel 2018). Core frameworks include one-step (Newton–Raphson) corrections (Pfanzagl 1982; Bickel et al. 1993; Robins, Rotnitzky, and Zhao 1995; Robins and Rotnitzky 1997), targeted minimum loss estimation (TMLE) (van der Laan and Rose 2011, 2018), and debiased machine learning (DML) (Chernozhukov et al. 2018). These methods leverage the *efficient influence function* (EIF) of the smooth functional defining the parameter (Hines et al. 2022).

Remark 3. *Let $\mathcal{S}_\delta[P]$ denote the functional that evaluates expression* (14) *at an arbitrary distribution P. Similarly, let $\mathcal{T}_\delta[P]$ represent the corresponding P-functional from expression* (15)*. Let $O_i = (W_i, A_i, Y_i)$ denote a sample drawn from the true DGP P^*. Under a nonparametric model, suitable smoothness and regularity conditions, the* uncentered *EIF of $\mathcal{S}_\delta[P]$ at P^*, evaluated at point O_i, exists and is:*

$$D_\delta^S(O_i) = D_\delta^{S,1}(O_i) + D_\delta^{S,2}(O_i), \tag{17}$$

$$D_\delta^{S,1}(O_i) = \frac{\pi_\delta^\star(A_i \mid W_i)}{\pi(A_i \mid W_i)} \left[Y_i - \sum_{a \in \mathcal{A}} \pi_\delta^\star(a \mid W_i)\, Q(Z_i, a) \right]$$

$$D_\delta^{S,2}(O_i) = \sum_{a \in \mathcal{A}} \pi_\delta^\star(a \mid W_i)\, Q(Z_i, a).$$

Analogously, under a nonparametric model and typical smoothness and regularity conditions, the uncentered *EIF of $\mathcal{T}_\delta[P]$ at P^*, evaluated at point O_i, exists and is given by:*

$$D_\delta^T(O_i) = D_\delta^{T,1}(O_i) + D_\delta^{T,2}(O_i) + D_\delta^{T,3}(O_i), \tag{18}$$

$$D_\delta^{T,1}(O_i) = \frac{\nu_\delta^\star(A_i \mid W_i)}{\pi(A_i \mid W_i)} \left[Y_i - Q(Z_i, A_i) \right]$$

$$D_\delta^{T,2}(O_i) = \left[2 - \frac{\pi_\delta^\star(A_i \mid W_i)}{\pi(A_i \mid W_i)} \right] \sum_{a \in \mathcal{A}} \nu_\delta^\star(a \mid W_i)\, Q(Z_i, a)$$

$$D_\delta^{T,3}(O_i) = \frac{\pi_\delta^\star(A_i \mid W_i)}{\pi(A_i \mid W_i)} \varrho_\delta(A_i)\, Q(Z_i, A_i) - \sum_{a \in \mathcal{A}} \pi_\delta^\star(a \mid W_i)\, \varrho_\delta(a)\, Q(Z_i, a),$$

where $\varrho_\delta(a) := \dfrac{\xi_\delta(a)}{\zeta_\delta + \xi_\delta(a)}$.

Derivation is provided in technical appendix A4.

For binary exposure A and target distribution $\nu(a) = \mathbb{I}(a = 1)$, it is straightforward to verify that $D_\delta^S(O_i)$ coincides with the uncentered EIF for the expected outcome after an IPI, as derived in prior work (Kennedy 2018; Bonvini et al. 2023).

Note that $\varrho_0(a) = 0$ and $\pi_0^\star(a \mid w) = \pi(a \mid w)$ for all $a \in \mathcal{A}, w \in \mathcal{W}$. Hence, when $\delta = 0$ and $\nu(a) = \mathbb{I}(a = a')$, we recover the standard uncentered EIF for the expected outcome after the hard intervention $\mathrm{do}(A = a')$, identified by $\mathbb{E}_Z\{Q(Z, a')\}$.

Consequently, one-step estimators of the expected outcomes in (14) and (15) are:

$$\widehat{\mu}_\delta^S = \frac{1}{n} \sum_{i=1}^n \widehat{D}_\delta^S(O_i) \quad \text{and} \quad \widehat{\mu}_\delta^T = \frac{1}{n} \sum_{i=1}^n \widehat{D}_\delta^T(O_i), \tag{19}$$

using samples $\{O_i\}_{i=1}^n \overset{iid}{\sim} P^*$, and where $\widehat{D}_\delta^S(\cdot)$ and $\widehat{D}_\delta^T(\cdot)$ are obtained by plugging the estimated nuisance components $\widehat{\pi}$ and $\widehat{Q}$ into $D_\delta^S(\cdot)$ and $D_\delta^T(\cdot)$, respectively. These estimators eliminate the first-order bias in the von Mises expansion of the corresponding functionals around the true distribution (Pfanzagl 1982; Bickel et al. 1993), even under partial misspecification of some components of the DGP.

Inference

To relax technical conditions for weak convergence, such as requiring nuisance estimators to belong to a Donsker class, we employ sample splitting and cross-fitting. In this approach, nuisance functions are estimated on one subset of the data and then plugged into the influence function evaluated on a disjoint subset; final estimates average results across folds (Chernozhukov et al. 2018).

Let $\widehat{\sigma}_\delta^S$ and $\widehat{\sigma}_\delta^T$ be consistent estimators of $\mathrm{var}\left(D_\delta^S(O)\right)^{1/2}$ and $\mathrm{var}\left(D_\delta^T(O)\right)^{1/2}$. Estimation and uncertainty quantification proceed as follows:

1. Split the sample into K folds. On each training fold k, fit $\widehat{\pi}_k$ and $\widehat{Q}_k$ using data-adaptive learners.
2. For a grid G of δ values, construct $\widehat{\pi}_{\delta,k}^\star$ and $\widehat{\nu}_{\delta,k}^\star$ via (7)–(8).
3. Evaluate $\widehat{D}_{\delta,k}^S$ and $\widehat{D}_{\delta,k}^T$ on the corresponding held-out fold, aggregate across units and folds, and form one-step estimators $\widehat{\mu}_\delta^S$ and $\widehat{\mu}_\delta^T$.
4. **Uniform confidence bands**: For $b \in \{1, \ldots, B\}$, draw i.i.d. multipliers $\{\chi_i^{(b)}\}_{i=1}^n \overset{iid}{\sim} N(0,1)$, and compute
$$\zeta_b^S = \sup_{\delta \in G} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi_i^{(b)} (\widehat{D}_\delta^S(O_i) - \widehat{\mu}_\delta^S) / \widehat{\sigma}_\delta^S \right|, \tag{20}$$
and save as ξ^S the 95% quantile of $\{\zeta_b^S\}_{b=1}^B$ for band construction around $\widehat{\mu}_\delta^S$. Analogously, construct ξ^T using $\widehat{D}_\delta^T$ and $\widehat{\sigma}_\delta^T$.
5. Report final point estimates and intervals:
$$\widehat{\mu}_\delta^S \mp \xi^S \frac{\widehat{\sigma}_\delta^S}{\sqrt{n}} \quad \text{and} \quad \widehat{\mu}_\delta^T \mp \xi^T \frac{\widehat{\sigma}_\delta^T}{\sqrt{n}}. \tag{21}$$

For *pointwise* inference at a fixed δ, take $\xi^S = \xi^T = 1.96$ to obtain 95% Wald-type intervals. For *uniform* bands over $\delta \in G$, use the multiplier-bootstrap critical values ξ^S and ξ^T as above. Since the uncentered EIF, viewed as a function of δ, is Lipschitz continuous in δ under our construction, the scaled estimation error process converges weakly in $\ell^\infty(G)$ to a tight Gaussian process, validating the uniform bands (Chernozhukov, Chetverikov, and Kato 2013; Kennedy 2018).

Simulations

We conducted an evaluation task for the proposed estimators using repeated simulations with finite synthetic data. The employed DGP is adapted from Kang and Schafer (2007) and Kennedy (2018), with bespoke modifications to introduce a three-leveled categorical exposure and to increase the noise in the system, thereby aligning the signal-to-noise ratio with amounts commonly observed in social science and observational clinical data. The DGP is given by:

$$
\begin{aligned}
W &\overset{iid}{\sim} N(\vec{0}, I_4), \\
\eta_1(W) &= \exp(-2W_1 + W_2 - 0.5W_3 - 0.25W_4), \\
\eta_2(W) &= \exp(-W_1 + 0.25W_2 + 2W_3 + 0.5W_4), \\
\pi(\alpha_k \,|\, W) &= \eta_k(W) / \left[\eta_1(W) + \eta_2(W) + 1\right], \; j \in \{1, 2\}, \\
\pi(\alpha_3 \,|\, W) &= 1 - \pi(\alpha_1 \,|\, W) - \pi(\alpha_2 \,|\, W), \\
A \,|\, W &\overset{iid}{\sim} \mathrm{Cat}_3\left(\pi(\alpha_1 \,|\, W), \pi(\alpha_2 \,|\, W), \pi(\alpha_3 \,|\, W)\right), \\
q(W) &= 2W_1 + W_2 + W_3 + W_4, \\
Q(W, A) &= \begin{cases} 10 - 8.7\, q(W) & \text{if } A = a_1 \\ 40 + 17.4\, q(W) & \text{if } A = a_2 \\ 50 + 26.1\, q(W) & \text{if } A = a_3 \end{cases}, \\
Y &= Q(W, A) + \varepsilon, \text{ where } \varepsilon \overset{iid}{\sim} N(0, 50).
\end{aligned}
$$

For the exposure model class, we use multinomial logistic regression with a linear predictor, and for the outcome we employ multivariate adaptive regression splines (MARS) with extra linear predictors W. These model classes are correctly specified in the sense that they contain the true propensity score π and outcome regression function Q, respectively. Following the approach of Kang and Schafer (2007), we introduce ad hoc misspecification in π and Q by using the same model classes but replacing the original covariates $W \in \mathbb{R}^4$ with a nonlinear transformation $X(W) \in \mathbb{R}^3$, which also constitutes a valid adjustment set, defined as:

$$X_1 = 10 + W_2/(1 + \exp(W_1)), \tag{22}$$

$$X_2 = (0.6 + W_1 W_3/25)^3, \tag{23}$$

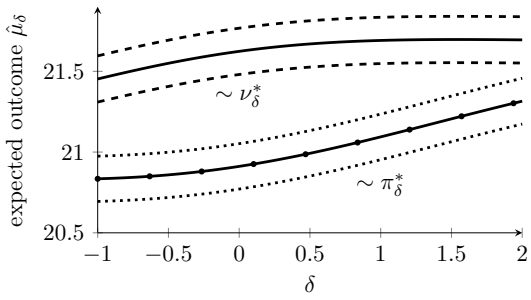
$$X_3 = (W_2 + W_4 + 20)^2. \tag{24}$$

We compare the performance of two estimators for the expected outcome after stochastic interventions: the plug-in estimator and the one-step bias-corrected estimator introduced in the previous section. This comparison is carried out for both stochastic intervention strategies $\pi^\star_\delta$ and $\nu^\star_\delta$ across various model misspecification scenarios, cost structures, and target distributions. Estimator performance is evaluated using integrated bias (iBias) and integrated root mean squared error (iRMSE), standard metrics in functional estimation. These measures are computed by averaging the absolute value of δ-specific bias and RMSE results over 100 equally spaced values of δ in the interval $[-2, 2]$.

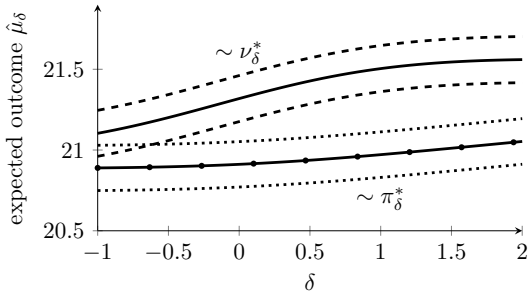
Setup 1: $c = (2.0, 1.0, 1.0)$, $\nu = (0.4, 0.4, 0.2)$					
		$\widehat{\mu}^S_\delta$ (under $\widehat{\pi}^\star_\delta$)		$\widehat{\mu}^T_\delta$ (under $\widehat{\nu}^\star_\delta$)	
Estim.	Misspec.	iBias	iRMSE	iBias	iRMSE
plug-in	–	0.40	2.36	3.03	5.69
plug-in	Q	1.99	3.21	15.22	15.43
plug-in	π	14.00	14.57	7.92	9.48
one-step	–	**0.02**	2.20	**0.57**	8.44
one-step	Q	**0.03**	**2.20**	**0.35**	**13.89**
one-step	π	**0.39**	**2.32**	**3.98**	**5.86**
Setup 2: $c = (1.0, 0.5, 2.0)$, $\nu = (0.5, 0.3, 0.2)$					
		$\widehat{\mu}^S_\delta$ (under $\widehat{\pi}^\star_\delta$)		$\widehat{\mu}^T_\delta$ (under $\widehat{\nu}^\star_\delta$)	
Estim.	Misspec.	iBias	iRMSE	iBias	iRMSE
plug-in	–	0.21	2.20	3.43	5.66
plug-in	Q	0.54	2.25	17.61	17.80
plug-in	π	14.67	15.14	3.08	5.59
one-step	–	**0.02**	2.14	**0.38**	7.38
one-step	Q	**0.02**	2.14	**0.50**	**11.24**
one-step	π	**0.36**	**2.23**	**2.66**	**4.82**
Setup 3: $c = (1.0, 1.0, 2.0)$, $\nu = (0.0, 0.2, 0.8)$					
		$\widehat{\mu}^S_\delta$ (under $\widehat{\pi}^\star_\delta$)		$\widehat{\mu}^T_\delta$ (under $\widehat{\nu}^\star_\delta$)	
Estim.	Misspec.	iBias	iRMSE	iBias	iRMSE
plug-in	–	1.20	3.04	9.46	11.82
plug-in	Q	3.82	4.77	27.41	27.76
plug-in	π	13.32	14.12	10.28	12.68
one-step	–	**0.04**	**2.40**	**0.75**	**8.86**
one-step	Q	**0.06**	**2.39**	**0.49**	**16.07**
one-step	π	**1.15**	**2.83**	**5.17**	**7.50**

Table 1: Integrated bias (iBias) and root mean squared error (iRMSE) for the estimated expected outcome after stochastic interventions $\pi^\star_\delta$ and $\nu^\star_\delta$. Results are averaged over 200 simulations and shown for two estimators (plug-in and one-step), across three model specifications: (i) correctly specified, (ii) misspecified outcome regression Q, and (iii) misspecified propensity score π; and under three combinations of cost functions c and target distributions ν.

Table 1 presents the results for a sample size of $n = 1000$ and averaged over 200 repetitions. Results across all three setups demonstrate the superior performance of the proposed one-step estimators over the plug-in estimators. Under correct specification, the one-step estimator achieves near-zero bias and lower or comparable RMSE, particularly under policy $\pi^\star_\delta$. Under policy $\nu^\star_\delta$, lower bias is traded for a moderately higher RMSE when the target distribution ν is not concentrated but spread over the sample space instead. The robustness of the proposed one-step estimators is especially notable under model misspecification: when the outcome regression Q is misspecified, the one-step estimator maintains low bias and RMSE in all setups, whereas the plug-in estimator exhibits increases in both metrics, most dramatically under $\nu^\star_\delta$. Similarly, with misspecification of the propensity



(a) $c = (0.0, 1.0, 2.0)$ and $\nu = (0.00, 0.75, 0.25)$



(b) $c = (0.5, 0.5, 2.0)$ and $\nu = (0.00, 0.50, 0.50)$

Figure 2: Estimated expected outcome after stochastic interventions $\pi^\star_\delta$ and $\nu^\star_\delta$ in the application case with 3-levels exposure: (1) no treatment, (2) low dose, and (3) high dose treatment. Broken lines represent uniform confidence bands.

score π, the one-step estimator substantially reduces bias and RMSE relative to the plug-in approach, though performance degrades more than with Q-misspecification. Across setups, $\widehat{\mu}^T_\delta$ tends to be more sensitive to misspecification than $\widehat{\mu}^S_\delta$, but the one-step correction consistently mitigates this sensitivity, highlighting its robustness and practical advantage in finite-sample, imperfect-model settings.

Application case

We estimate the expected outcomes after various stochastic intervention strategies involving pharmacological treatment with stimulants for ADHD on children's academic achievement. We focus on numeracy test scores in grade 8th among Norwegian children diagnosed with ADHD. Drawing on linked data from national registries, we compile comprehensive records on medication histories and academic outcomes for all children born between 2000 and 2007 in Norway who were diagnosed with ADHD and eligible for the national tests up to 2021, excluding those with severe conditions, resulting in a sample of 8,609 kids. Information at the student, family, and school levels is integrated from prescription databases, patient registries, and official health, education and demographic statistics, including medical consultations and diagnostic histories. To estimate *defined daily doses* (DDD) from prescription records, we apply the PRE2DUP algorithm (Tanskanen et al. 2015) to classify stimulant treatment into three categories: untreated (47%), *low dose* (21%) and *high dose* (32%).

Although stimulant medication is well-documented to reduce core ADHD symptoms (Cortese et al. 2018), its effects on educational outcomes are more modest (Faraone et al. 2021; Storebø et al. 2015). Evidence points to only small improvements in standardized test performance (Pelham et al. 2022; Jangmo et al. 2019) and limited long-term gains in academic achievement among Norwegian children (Varnet-Pérez et al. 2025).

Given that methylphenidate is the most widely prescribed stimulant, and that adverse effects such as sleep disturbances and weight loss are commonly reported (Graham and Coghill 2008), the overall cost of treatment may encompass both direct healthcare expenditures and costs associated with managing adverse effects. To account for this, we evaluate two cost/target configurations, both assigning a higher cost to high-dose treatment: (*a*) a target in which 75% of individuals receive low-dose treatment and 25% receive high-dose treatment, and (*b*) a balanced target with a 50–50% allocation between low and high doses.

Figure 2 displays the estimated expected outcomes after source-tilted ($\pi^\star_\delta$) and target-tilted ($\nu^\star_\delta$) stochastic interventions across the range $\delta \in [-1, 2]$, with 95% uniform confidence bands. In both scenarios, expected outcomes increase gradually with δ, indicating a small benefit, of less than one test score point, as the interventions shifts the propensity score toward the PoE distribution. Across most of the range of δ, the target-tilted policy $\nu^\star_\delta$ achieves consistently higher expected outcomes than the source-tilted policy $\pi^\star_\delta$, indicating that, for this case, shifting the target distribution dominates shifting the source. Results also indicate that target distributions placing greater emphasis on low-dose treatment are associated with better academic achievement outcomes.

Conclusions

We propose a cost-aware family of stochastic interventions for discrete treatments that generalizes incremental propensity score interventions and connects causal modeling to cost-sensitive decisions. Modeling a cost-penalized I-projection (CPIP) of the independent product of organic and target distributions yields closed-form Boltzmann–Gibbs couplings whose marginals, via a single tilt δ, interpolate from the propensity scores or from the target policy toward a product-of-experts limit when destination costs are strictly positive. We derive efficient influence functions under a nonparametric model and construct one-step estimators with uniform bands over δ, improving stability and misspecification robustness relative to plug-in baselines.

These policies enable graded scientific hypotheses under realistic constraints. Because δ is continuous and costs c and targets ν are modular, analysts can prototype and evaluate policies for prospective studies, turning observational registries into pre-experimental testbeds. Explicit costs clarify prioritization, aligning interventions with budgets and logistics while quantifying trade-offs. Clinician-informed targets integrate naturally, ensuring policies reflect expert priors and empirical regularities. Overall, this framework links identification and estimation from observational data to resource-aware experimental design when hard interventions are infeasible.

Acknowledgments This work was supported by the Research Council of Norway via grants with project numbers 302899 (*Effects of pharmacological treatment and special education on school performance in children with ADHD*) and 332645 (*Integreat – Norwegian centre for knowledge-driven machine learning*). Data curation for the application case was conducted by Dr. Guido Biele at the Norwegian Institute of Public Health. The study was approved by the Regional Committees for Medical and Healthcare Research Ethics (REK approval no. 96604).

References

Bickel, P.; Klaassen, C.; Ritov, Y.; and Wellner, J. 1993. *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer New York. ISBN 9780387984735.

Bonvini, M.; McClean, A.; Branson, Z.; and Kennedy, E. 2023. Incremental Causal Effects: An Introduction and Review. In Elliott, M.; Stuart, E.; and Westreich, D., eds., *Handbook of Matching and Weighting Adjustments for Causal Inference*, 1–24. Chapman and Hall/CRC, 1st edition. ISBN 9781003102670.

Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; and Robins, J. 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1): C1–c68.

Chernozhukov, V.; Chetverikov, D.; and Kato, K. 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6): 2786–2819.

Chizat, L.; Peyré, G.; Schmitzer, B.; and Vialard, F.-X. 2018a. Scaling Algorithms for Unbalanced Optimal Transport Problems. *Mathematics of Computation*, 87(314): 2563–2609. Published electronically on February 6, 2018.

Chizat, L.; Peyré, G.; Schmitzer, B.; and Vialard, F.-X. 2018b. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274(11): 3090–3123.

Cinelli, C.; Forney, A.; and Pearl, J. 2024. A Crash Course in Good and Bad Controls. *Sociological Methods and Research*, 53(3): 1071–1104.

Correa, J.; and Bareinboim, E. 2020a. A Calculus for Stochastic Interventions: Causal Effect Identification and Surrogate Experiments. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06): 10093–10100.

Correa, J.; and Bareinboim, E. 2020b. General transportability of soft interventions: completeness results. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713829546.

Cortese, S.; Adamo, N.; Del Giovane, C.; Mohr-Jensen, C.; Hayes, A.; Carucci, S.; Atkinson, L.; Tessari, L.; Banaschewski, T.; Coghill, D.; Hollis, C.; Simonoff, E.; Zuddas, A.; Barbui, C.; Purgato, M.; Steinhausen, H.-C.; Shokraneh, F.; Xia, J.; and Cipriani, A. 2018. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *The lancet. Psychiatry*, 5(9): 727–738.

Daniel, R. 2018. *Double Robustness*, 1–14. John Wiley and Sons, Ltd. ISBN 9781118445112.

Díaz, I. 2019. Machine learning in the estimation of causal effects: targeted minimum loss-based estimation and double/debiased machine learning. *Biostatistics*, 21(2): 353–358.

Duong, T.-D.; Li, Q.; and Xu, G. 2021. Stochastic intervention for causal effect estimation. In *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.

Díaz, I.; and Hejazi, N. 2020. Causal mediation analysis for stochastic interventions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3): 661–683. Published online: 05 February 2020.

Díaz, I.; and van der Laan, M. 2012. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2): 541–549.

Díaz, I.; and van der Laan, M. 2013. Assessing the causal effect of policies: an example using stochastic interventions. *The International Journal of Biostatistics*, 9(2): 149–165.

Díaz, I.; Williams, N.; Hoffman, K.; and Schenck, E. 2023. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*, 118(542): 846–857.

Faraone, S.; Banaschewski, T.; Coghill, D.; Zheng, Y.; Biederman, J.; Bellgrove, M.; Newcorn, J.; Gignac, M.; Al Saud, N.; Manor, I.; et al. 2021. The world federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. *Neuroscience & Biobehavioral Reviews*, 128: 789–818.

Frogner, C.; Zhang, C.; Mobahi, H.; Araya, M.; and Poggio, T. 2015. Learning with a Wasserstein Loss. In Cortes, C.; Lawrence, N.; Lee, D.; Sugiyama, M.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Graham, J.; and Coghill, D. 2008. Adverse effects of pharmacotherapies for attention-deficit hyperactivity disorder: epidemiology, prevention and management. *CNS drugs*, 22: 213–237.

Haneuse, S.; and Rotnitzky, A. 2013. Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine*, 32(30): 5260–5277.

Hejazi, N.; Rudolph, K.; van der Laan, M.; and Díaz, I. 2023. Nonparametric causal mediation analysis for stochastic interventional (in)direct effects. *Biostatistics*, 24(3): 686–707. Published online: 01 February 2022.

Hines, O.; Dukes, O.; Díaz-Ordaz, K.; and Vansteelandt, S. 2022. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 76(3): 292–304.

Hinton, G. 1999. Products of Experts. In *9th International Conference on Artificial Neural Networks (ICANN'99)*, 1–6. IEE. ISBN 978-0-85296-721-8.

Imbens, G.; and Rubin, D. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. USA: Cambridge University Press. ISBN 0521885884.

Jangmo, A.; Stålhandske, A.; Chang, Z.; Chen, Q.; Almqvist, C.; Feldman, I.; Bulik, C. M.; Lichtenstein, P.; D'Onofrio, B.; Kuja-Halkola, R.; and Larsson, H. 2019. Attention-Deficit/Hyperactivity Disorder, School Performance, and Effect of Medication. *Journal of the American Academy of Child and Adolescent Psychiatry*, 58(4): 423–432.

Kang, J.; and Schafer, J. 2007. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4): 523–539.

Kennedy, E. 2018. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 113(522): 645–656. Published online: 07 Aug 2018.

Kim, K.; Kennedy, E.; and Naimi, A. 2021. Incremental intervention effects in studies with dropout and many timepoints. *Journal of Causal Inference*, 9(1): 302–344.

Léonard, C. 2014. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4): 1533–1574. Mathematics Subject Classification: 46N10, 60J25, 60F10.

Levis, A.; Kennedy, E.; McClean, A.; Balakrishnan, S.; and Wasserman, L. 2024. Stochastic interventions, sensitivity analysis, and optimal transport. *arXiv preprint arXiv:2411.14285*.

McClean, A.; Li, Y.; Bae, S.; McAdams-DeMarco, M.; Díaz, I.; and Wu, W. 2024. Fair comparisons of causal parameters with many treatments and positivity violations. *arXiv preprint arXiv:2410.13522*.

Naimi, A. I.; Rudolph, J. E.; Kennedy, E. H.; Cartus, A.; Kirkpatrick, S. I.; Haas, D. M.; Simhan, H.; and Bodnar, L. M. 2021. Incremental propensity score effects for time-fixed exposures. *Epidemiology*, 32(2): 202–208.

Opacic, A.; Wei, L.; and Zhou, X. 2025. Disparity analysis: a tale of two approaches. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. Published: 25 February 2025.

Pearl, J. 2009. *Causality*. Cambridge, UK: Cambridge University Press, 2 edition. ISBN 978-0-521-89560-6.

Pelham, W.; Altszuler, A.; Merrill, B.; Raiker, J.; Macphee, F.; Ramos, M.; Gnagy, E.; Greiner, A.; Coles, E.; Connor, C.; et al. 2022. The effect of stimulant medication on the learning of academic curricula in children with ADHD: A randomized crossover study. *Journal of consulting and clinical psychology*, 90(5): 367.

Peyré, G.; and Cuturi, M. 2019. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning*, 11(5-6): 355–607.

Pfanzagl, J. 1982. *Contributions to a General Asymptotic Statistical Theory*, volume 13 of *Lecture Notes in Statistics*. New York: Springer-Verlag. ISBN 0387907769. With the assistance of W. Wefelmeyer.

Robins, J.; and Rotnitzky, A. 1997. Comment on the Bickel and Kwon article, 'Inference for semiparametric models: Some questions and an answer'. *Statistica Sinica*, 7(1): 167–199.

Robins, J.; Rotnitzky, A.; and Zhao, L. 1995. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429): 106–121.

Rudolph, J.; Kim, K.; Kennedy, E.; and Naimi, A. 2022. Estimation of the time-varying incremental effect of low dose aspirin on incidence of pregnancy. *Epidemiology*, 34(1): 38–44. Author manuscript available in PMC: Jan 1, 2024.

Sani, N.; Lee, J.; and Shpitser, I. 2020. Identification and Estimation of Causal Effects Defined by Shift Interventions. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, 949–958. PMLR. Author manuscript available in PMC: Dec 4, 2020.

Sarvet, A.; Wanis, K.; Young, J.; Hernandez-Alejandro, R.; and Stensrud, M. 2023. Longitudinal incremental propensity score interventions for limited resource settings. *Biometrics*, 79(4): 3418–3430.

Storebø, O.; Krogh, H.; Ramstad, E.; Moreira-Maia, C.; Holmskov, M.; Skoog, M.; Nilausen, T.; Magnusson, F.; Zwi, M.; Gillies, D.; Rosendal, S.; Groth, C.; Rasmussen, K.; Gauci, D.; Kirubakaran, R.; Forsbøl, B.; Simonsen, E.; and Gluud, C. 2015. Methylphenidate for attention-deficit/hyperactivity disorder in children and adolescents: Cochrane systematic review with meta-analyses and trial sequential analyses of randomised clinical trials. *British Medical Journal*, 351: h5203.

Tanskanen, A.; Taipale, H.; Koponen, M.; Tolppanen, A.-M.; Hartikainen, S.; Ahonen, R.; and Tiihonen, J. 2015. From prescription drug purchases to drug use periods–a second generation method (PRE2DUP). *BMC medical informatics and decision making*, 15(1): 21.

van der Laan, M.; and Rose, S. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Series in Statistics. Springer New York. ISBN 9781441997821.

van der Laan, M.; and Rose, S. 2018. *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Publishing Company, Incorporated, 1st edition. ISBN 3319653032.

Varnet-Pérez, T.; Øvergaard, K.; Frigessi, A.; and Biele, G. 2025. Long-term effect of pharmacological treatment on academic achievement of Norwegian children diagnosed with ADHD: a target trial emulation. *International Journal of Epidemiology*, 54(2): dyaf010.

Technical Appendix

A1 – CPIP and tilted marginals

Consider cost-penalized I-projection (CPIP) of the independent product of input distributions π and ν over discrete sample space $\mathcal{A}$, formulated as:

$$\inf_{\gamma \in \mathcal{M}_+^1(\mathcal{A}^2)} \sum_{a',a'' \in \mathcal{A}} c(a',a'')\,\gamma(a',a'') + \frac{1}{\delta}\mathbb{D}_{\mathrm{KL}}(\gamma \,|\, \pi \otimes \nu),$$

where $\delta > 0$, $\mathcal{M}_+^1(\mathcal{A}^2)$ denotes the set of probability measures over sample space $\mathcal{A}^2$, and $c : \mathcal{A}^2 \to \mathbb{R}$ is a nonnegative and integrable cost function. Owing to the strict convexity of the objective, the unique minimizer satisfies the KKT conditions obtained from the Lagrangian:

$$\mathcal{L}(\gamma,\lambda) = \sum_{a',a'' \in \mathcal{A}} \left[c(a',a'') + \frac{1}{\delta}\log\left(\frac{\gamma(a',a'')}{\pi(a')\,\nu(a'')}\right)\right]\gamma(a',a'') - \lambda\left[\sum_{a',a'' \in \mathcal{A}} \gamma(a',a'') - 1\right],$$

which admits a unique closed-form solution. The first-order condition is given by:

$$\begin{aligned}
&\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\gamma(a',a'')} = c(a',a'') + \frac{1}{\delta}\left[\log\left(\frac{\gamma(a',a'')}{\pi(a')\,\nu(a'')}\right) + 1\right] - \lambda = 0,\\
&\Rightarrow \log\left(\frac{\gamma(a',a'')}{\pi(a')\,\nu(a'')}\right) = \delta[\lambda - c(a',a'')] - 1,\\
&\Rightarrow \gamma(a',a'') = \pi(a')\,\nu(a'')\,e^{-\delta c(a',a'')}\,e^{\delta\lambda - 1} \propto \pi(a')\,\nu(a'')\,e^{-\delta c(a',a'')}.
\end{aligned}$$

Now suppose $\pi \equiv \pi(\cdot \,|\, w)$ is a conditional distribution, and the cost function c does not depend on the covariate profile $W = w$. Then, for each profile w, the solution plan becomes:

$$\gamma_\delta^\star(a',a'' \,|\, w) = \frac{\pi(a' \,|\, w)\,\nu(a'')\,e^{-\delta c(a',a'')}}{\sum_{a',a'' \in \mathcal{A}} \pi(a' \,|\, w)\,\nu(a'')\,e^{-\delta c(a',a'')}}.$$

The marginals of $\gamma^\star(a',a'' \,|\, w)$ are then straightforward to compute:

$$\pi_\delta^\star(a \,|\, w) = \frac{\pi(a \,|\, w)\sum_{a'' \in \mathcal{A}} \nu(a'')\,e^{-\delta c(a,a'')}}{\sum_{a',a'' \in \mathcal{A}} \pi(a' \,|\, w)\,\nu(a'')\,e^{-\delta c(a',a'')}} \quad \text{and} \quad \nu_\delta^\star(a \,|\, w) = \frac{\nu(a)\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w)\,e^{-\delta c(a',a)}}{\sum_{a',a'' \in \mathcal{A}} \pi(a' \,|\, w)\,\nu(a'')\,e^{-\delta c(a',a'')}}.$$

A2 – Remark 1: IPIs as special case

Let $A \in \{0,1\}$ be a binary point-exposure. Let the target marginal ν be the degenerate distribution that always assigns treatment, $\nu(a) = \mathbb{I}(a = 1)$, and let $c(a',a'') = \mathbb{I}(a' \neq a'')$ be the Hamming cost. Then, the tilted source marginal of the CPIP solution with regularization parameter δ is:

$$\begin{aligned}
\pi_\delta^\star(a \,|\, w) &= \frac{\pi(a \,|\, w)\sum_{a'' \in \mathcal{A}} \mathbb{I}(a'' = 1)\,e^{-\delta\mathbb{I}(a \neq a'')}}{\sum_{a',a'' \in \mathcal{A}} \pi(a' \,|\, w)\,\mathbb{I}(a'' = 1)\,e^{-\delta\mathbb{I}(a' \neq a'')}} = \frac{\pi(a \,|\, w)\,e^{-\delta\mathbb{I}(a \neq 1)}}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w)\,e^{-\delta\mathbb{I}(a' \neq 1)}},\\
\Rightarrow \pi_\delta^\star(1 \,|\, w) &= \frac{\pi(1 \,|\, w)\,e^{-\delta\cdot 0}}{\pi(1 \,|\, w)\,e^{-\delta\cdot 0} + \pi(0 \,|\, w)\,e^{-\delta\cdot 1}} = \frac{\pi(1 \,|\, w)\,e^{\delta}}{\pi(1 \,|\, w)\,e^{\delta} + \pi(0 \,|\, w)},
\end{aligned}$$

which coincides with an IPI with tilt parameter δ, and thus $\pi_\delta^\star(1 \,|\, w) = \widetilde{\pi}_\delta(1 \,|\, w)$ for all $w \in \mathcal{W}$.

A3 – Remark 2: Tilted marginals under treatment-specific costs

Let:

- $A \in \mathcal{A} = \{\alpha_1, \ldots, \alpha_K\}$ be a categorical point-exposure variable with K treatment options,
- The target marginal ν be any valid probability distribution over $\mathcal{A}$,
- The reallocation cost from $A = \alpha_j$ to $A = \alpha_k \neq \alpha_j$ be a value that is specific for the received treatment α_k and constant over profiles $W = w$, i.e., $c(\alpha_j, \alpha_k) = c(\alpha_k)\,\mathbb{I}(\alpha_j \neq \alpha_k)$, with $0 \leq c(a) < \infty$ for all $a \in \mathcal{A}$.

Then, the tilted source marginal of the CPIP solution with parameter δ corresponds to:

$$\pi^\star_\delta(a \,|\, w) = \frac{\pi(a \,|\, w) \sum_{a'' \in \mathcal{A}} \nu(a'')\, e^{-\delta c(a'')\, \mathbb{I}(a \neq a'')}}{\sum_{a', a'' \in \mathcal{A}} \pi(a' \,|\, w)\, \nu(a'')\, e^{-\delta c(a'')\, \mathbb{I}(a' \neq a'')}} = \frac{\pi(a \,|\, w) \left[\nu(a) + \sum_{a'' \neq a} \nu(a'')\, e^{-\delta c(a'')}\right]}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w) \left[\nu(a') + \sum_{a'' \neq a'} \nu(a'')\, e^{-\delta c(a'')}\right]},$$

$$= \frac{\pi(a \,|\, w) \left[\nu(a) + \sum_{a'' \in \mathcal{A}} \nu(a'')\, e^{-\delta c(a'')} - \nu(a)\, e^{-\delta c(a)}\right]}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w) \left[\nu(a') + \sum_{a'' \in \mathcal{A}} \nu(a'')\, e^{-\delta c(a'')} - \nu(a')\, e^{-\delta c(a')}\right]} = \frac{\pi(a \,|\, w)\, (\zeta_\delta + \xi_\delta(a))}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w)\, (\zeta_\delta + \xi_\delta(a'))},$$

where $\xi_\delta(a) := \nu(a) \left(1 - e^{-\delta c(a)}\right)$ and $\zeta_\delta := \sum_{a' \in \mathcal{A}} \nu(a')\, e^{-\delta c(a')}$.

Similarly, the tilted target marginal of the CPIP solution with parameter δ corresponds to:

$$\nu^\star_\delta(a \,|\, w) = \frac{\nu(a) \sum_{a' \in \mathcal{A}} \pi(a' \,|\, w)\, e^{-\delta c(a)\, \mathbb{I}(a' \neq a)}}{\sum_{a', a'' \in \mathcal{A}} \pi(a' \,|\, w)\, \nu(a'')\, e^{-\delta c(a'')\, \mathbb{I}(a' \neq a'')}} = \frac{\nu(a) \left[\pi(a \,|\, w) + \sum_{a' \neq a} \pi(a' \,|\, w)\, e^{-\delta c(a)}\right]}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w) \left[\nu(a') + \sum_{a'' \neq a'} \nu(a'')\, e^{-\delta c(a'')}\right]},$$

$$= \frac{\nu(a) \left[\pi(a \,|\, w) + e^{-\delta c(a)} (1 - \pi(a \,|\, w))\right]}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w) \left[\nu(a') + \sum_{a'' \in \mathcal{A}} \nu(a'')\, e^{-\delta c(a'')} - \nu(a')\, e^{-\delta c(a')}\right]},$$

$$= \frac{\nu(a) \left[1 - (1 - e^{-\delta c(a)})(1 - \pi(a \,|\, w))\right]}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w)\, (\zeta_\delta + \xi_\delta(a'))} = \frac{\nu(a) - \xi_\delta(a)(1 - \pi(a \,|\, w))}{\sum_{a' \in \mathcal{A}} \pi(a' \,|\, w)\, (\zeta_\delta + \xi_\delta(a'))}.$$

A4 – Remark 3: Efficient influence functions

Let ν and c be given, W to be a backdoor admissible set of pre-exposure covariates, $Q(w, a) = \mathbb{E}[Y \,|\, W = w, A = a]$, and:

$$\mathcal{S}_\delta[P] \equiv \mu^S_\delta = \sum_{a \in \mathcal{A}} \mathbb{E}_W \left\{\pi^\star_\delta(a \,|\, W)\, Q(W, a)\right\},$$
$$\mathcal{T}_\delta[P] \equiv \mu^s_\delta = \sum_{a \in \mathcal{A}} \mathbb{E}_W \left\{\nu^\star_\delta(a \,|\, W)\, Q(W, a)\right\},$$

Consider parametric submodel $P_\epsilon \in \mathfrak{P}$ indexed by a small fluctuation parameter $\epsilon \in \mathbb{R}$, and a point-mass contamination $O_i = (W_i, A_i, Y_i) \sim P^*$, such that, $P_\epsilon(O) = \epsilon\, \mathbb{I}(O = O_i) + (1 - \epsilon)\, P^*(O)$, where $P^* \in \mathfrak{P}$ is the true DGP distribution. Under some technical conditions involving *(i)* fully nonparametric or saturated model $\mathfrak{P}$, *(ii)* smoothness for the paths within the model, and *(iii)* boundedness of the outcome mean, the Gâteaux derivative and their variances, one has that $\mathcal{S}_\delta[P]$ and $\mathcal{T}_\delta[P]$ are pathwise differentiable at P^*.

The *uncentered* efficient influence function (EIF) of $\mathcal{S}_\delta[P]$ at P^* evaluated at O_i is given by $D^S_\delta(O_i) := \frac{\mathrm{d}\mathcal{S}_\delta[P_\epsilon]}{\mathrm{d}\epsilon}\big|_{\epsilon=0} + \mathcal{S}_\delta[P^*]$, and can be computed using the using the chain rule and gradient algebra for the Gâteaux derivative, as follows:

$$D^S_\delta(O_i) = \sum_{a \in \mathcal{A}} \frac{1}{H_\delta(W_i)^2} \left[H_\delta(W_i) \left(s_\delta(W_i, a)\, Q'(O_i, a) + s'_\delta(W_i, a)\, Q(W_i, a)\right)\right.$$
$$\left. - s_\delta(W_i, a)\, Q(W_i, a)\, H'_\delta(O_i)\right],$$
$$= \sum_{a \in \mathcal{A}} \left\{\frac{s_\delta(W_i, a)\, Q'(O_i, a)}{H_\delta(W_i)} + \frac{s'_\delta(W_i, a)\, Q(W_i, a)}{H_\delta(W_i)} - \frac{s_\delta(W_i, a)\, Q(W_i, a)\, H'_\delta(O_i)}{H_\delta(W_i)^2}\right\},$$

with

$$Q'(O_i, a) = \frac{\mathbb{I}(a = A_i)}{\pi(a \,|\, W_i)} \left[Y_i - Q(W_i, a)\right] + Q(W_i, a),$$
$$s_\delta(W_i, a) = (\zeta_\delta + \xi_\delta(a))\, \pi(a \,|\, W_i),$$
$$s'_\delta(W_i, a) = (\zeta_\delta + \xi_\delta(a))\, \mathbb{I}(a = A_i) - s_\delta(W_i, a),$$
$$H_\delta(W_i) = \sum_{a' \in \mathcal{A}} (\zeta_\delta + \xi_\delta(a'))\ \pi(a' \,|\, W_i),$$
$$H'_\delta(O_i) = (\zeta_\delta + \xi_\delta(A_i)) - H_\delta(W_i).$$

These expressions satisfy the following equivalences:

$$s_\delta(W_i, a')/H_\delta(W_i) = \pi^\star_\delta(a' \,|\, W_i) \quad \text{and} \quad (\zeta_\delta + \xi_\delta(a'))/H_\delta(W_i) = \pi^\star_\delta(a' \,|\, W_i)/\pi(a' \,|\, W_i).$$

Therefore,

$$\begin{aligned}
D^S_\delta(O_i) &= \sum_{a\in\mathcal{A}} \Bigg\{ \pi^\star_\delta(a \,|\, W_i) \left[\frac{\mathbb{I}(a = A_i)}{\pi(a \,|\, W_i)} \left[Y_i - Q(W_i, a) \right] + Q(W_i, a) \right] \\
&\quad + \left[\pi^\star_\delta(a \,|\, W_i) \frac{\mathbb{I}(a = A_i)}{\pi(a \,|\, W_i)} - \pi^\star_\delta(a \,|\, W_i) \right] Q(W_i, a) - \pi^\star_\delta(a \,|\, W_i)\, Q(W_i, a) \left[\frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} - 1 \right] \Bigg\}, \\
&= \frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \left[Y_i - Q(W_i, A_i) \right] + \sum_{a\in\mathcal{A}} \pi^\star_\delta(a \,|\, W_i)\, Q(W_i, a) + \frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} Q(W_i, A_i) \\
&\quad - \frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \sum_{a\in\mathcal{A}} \pi^\star_\delta(a \,|\, W_i)\, Q(W_i, a), \\
&= \underbrace{\frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \left[Y_i - \sum_{a\in\mathcal{A}} \pi^\star_\delta(a \,|\, W_i)\, Q(W_i, a) \right]}_{D^{S,1}_\delta(O_i)} + \underbrace{\sum_{a\in\mathcal{A}} \pi^\star_\delta(a \,|\, W_i)\, Q(W_i, a)}_{D^{S,2}_\delta(O_i)}.
\end{aligned}$$

Analogously, the *uncentered* EIF of $\mathcal{T}_\delta[P]$ at P^* evaluated at point O_i is given by $D^T_\delta(O_i) := \frac{\mathrm{d}\mathcal{T}_\delta[P_\epsilon]}{\mathrm{d}\epsilon}\big|_{\epsilon=0} + \mathcal{T}_\delta[P^*]$, and can be computed as:

$$D^S_\delta(O_i) = \sum_{a\in\mathcal{A}} \left\{ \frac{t_\delta(W_i, a)\, Q'(O_i, a)}{H_\delta(W_i)} + \frac{t'_\delta(W_i, a)\, Q(W_i, a)}{H_\delta(W_i)} - \frac{t_\delta(W_i, a)\, Q(W_i, a)\, H'_\delta(O_i)}{H_\delta(W_i)^2} \right\},$$

where $t_\delta(W_i, a) = \nu(a) - \xi_\delta(a)(1 - \pi(a \,|\, W_i))$ and $t'_\delta(W_i, a) = \xi_\delta(a)\ [\mathbb{I}(a = A_i) - \pi(a \,|\, W_i)]$.

These expressions satisfy the following equivalences:

$$\begin{aligned}
t_\delta(W_i, a')/H_\delta(W_i) &= \nu^\star_\delta(a' \,|\, W_i), \\
\xi_\delta(a')/H_\delta(W_i) &= \varrho_\delta(a')\, \pi^\star_\delta(a' \,|\, W_i)/\pi(a' \,|\, W_i), \text{ with} \\
\varrho_\delta(a') &= \xi_\delta(a')/(\zeta_\delta + \xi_\delta(a')).
\end{aligned}$$

Therefore,

$$\begin{aligned}
D^T_\delta(O_i) &= \sum_{a\in\mathcal{A}} \Bigg\{ \nu^\star_\delta(a \,|\, W_i) \left[\frac{\mathbb{I}(a = A_i)}{\pi(a \,|\, W_i)} \left[Y_i - Q(W_i, a) \right] + Q(W_i, a) \right] \\
&\quad + \left[\varrho_\delta(a)\, \pi^\star_\delta(a \,|\, W_i) \frac{\mathbb{I}(a = A_i)}{\pi(a \,|\, W_i)} - \varrho_\delta(a)\, \pi^\star_\delta(a \,|\, W_i) \right] Q(W_i, a) - \nu^\star_\delta(a \,|\, W_i)\, Q(W_i, a) \left[\frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} - 1 \right] \Bigg\}, \\
&= \frac{\nu^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \left[Y_i - Q(W_i, A_i) \right] + 2 \sum_{a\in\mathcal{A}} \nu^\star_\delta(a \,|\, W_i)\, Q(W_i, a) + \frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \varrho_\delta(A_i)\, Q(W_i, A_i) \\
&\quad - \sum_{a\in\mathcal{A}} \pi^\star_\delta(a \,|\, W_i)\, \varrho_\delta(a)\, Q(W_i, a) - \frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \sum_{a\in\mathcal{A}} \nu^\star_\delta(a \,|\, W_i)\, Q(W_i, a), \\
&= \underbrace{\frac{\nu^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \left[Y_i - Q(W_i, A_i) \right]}_{D^{T,1}_\delta(O_i)} + \underbrace{\left[2 - \frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \right] \sum_{a\in\mathcal{A}} \nu^\star_\delta(a \,|\, W_i)\, Q(W_i, a)}_{D^{T,2}_\delta(O_i)} \\
&\quad + \underbrace{\frac{\pi^\star_\delta(A_i \,|\, W_i)}{\pi(A_i \,|\, W_i)} \varrho_\delta(A_i)\, Q(W_i, A_i) - \sum_{a\in\mathcal{A}} \pi^\star_\delta(a \,|\, W_i)\, \varrho_\delta(a)\, Q(W_i, a)}_{D^{T,3}_\delta(O_i)}.
\end{aligned}$$